Developing an Ecologically Valid Measure of Creativity for Children with Autism Spectrum Disorders

Polly Cocker

D.Clin.Psy Thesis (Volume 1), 2018

University College London

UCL Doctorate in Clinical Psychology

Thesis declaration form

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

Signature: 

Name: Polly Cocker

Date: 19th June 2018

Overview

The subject of this thesis is creativity in Autism Spectrum Disorders (ASD). Part 1 provides a literature review of studies measuring creative thinking and the quality of creative ideas in children and adults with ASD. Both meta-analytic and narrative techniques are used to synthesise a profile of creativity in ASD. Recommendations are made to address the methodological limitations of the studies and more comprehensively and validly study creative performance in individuals with ASD.

Part 2 presents an empirical paper describing the development and piloting of a new ecologically valid measure of creativity in children with ASD. Three tasks are investigated in relation to their psychometric properties: interrater and test-retest reliability; criterion and construct validity; and measure acceptability. Preliminary between-group comparisons are made to explore creative performance in children with and without ASD and observe how task conditions moderate these effects.

A critical appraisal of the research project is put forward in Part 3. It offers a number of reflections on the process of developing the creativity tasks and scoring criteria as well as expanding upon limitations of the study. Further, it considers broader conceptual themes relating to research in the fields of creativity and ASD and the parallels with engaging in a creative research process. Finally, recommendations for future development of the task battery are made.

Impact Statement

This major research project concerns the study of creativity and ASD and is situated within a wider research effort to develop new measures comprising the Ecologically Valid Tests of Executive Dysfunction (Eco-TED).

Creativity holds value on an individual, organisational and societal level; across the lifespan, disciplines and cultures. The effective study and measurement of creativity betters understanding of the complex interplay of factors that can enhance or diminish creativity and in turn facilitates the development of strategies to maximise creative potential. The empirical study of creativity and ASD has received little attention thus far and this thesis aims to address this issue.

The Literature Review (Part 1) seeks to add to the knowledge base by synthesising the available studies into a more coherent account of creativity in ASD. Previous reviews have focused only on one aspect of creativity: the number of ideas produced. This does not allow for the comprehensive assessment of creativity in ASD.

To the author's knowledge, this is the first review that attempts to bring the other main components of creativity in ASD into focus. What emerges is not a creativity deficit, but a varied profile of creativity in ASD. This helps shift the narrative from a deficit focus to one of valuing difference, which can have a positive impact not only upon research efforts, but also the ASD communities they serve.

Furthermore, Part 1 highlights several methodological issues with the most widespread type of creativity measure, divergent thinking (DT) tasks, especially for use with an ASD population. Consideration of these issues can further research efforts in this field in the production of standardised, ecologically valid assessment tools.

The development of ecologically valid measures is a promising avenue of research in ASD as these tools prioritise representativeness and generalisability and

are therefore able to guide clinically useful adaptations and interventions. These advantages appear particularly pertinent in the field of creativity and ASD due to the recognised criterion problem in psychometric creativity research, whereby performance on DT tasks does not correspond with creative activities and achievements in real life.

Criterion and construct validity were relative areas of strength for the newly developed creativity tasks described in the Empirical Paper (Part 2), especially in correspondence with creativity ratings given by external experts in the applied domain. This finding holds promise for the tasks as an ecologically valid tool that can be used by researchers, educators and employers alike to measure real-life creative potential, particularly in children with neurodiversity. This may also go some way to addressing stereotypes that can exist around ASD in these settings.

In addition, exploratory analysis suggested that under certain task conditions, creative output was enhanced for both children with and without ASD and group differences diminished. This is clinically useful, as it can inform adaptations and interventions in the school and home environment that allow children to achieve their creative potential. Furthermore, clinicians working with children with ASD can provide appropriate support to foster creative development and think creatively within their own clinical contexts.

Table of Contents

Appendices

List of Tables

Part 1: Literature Review

Part 2: Empirical Paper

List of Figures

Part 1: Literature Review

Acknowledgements

First and foremost, I would like to thank the children and families that contributed their time and energy to take part in this study. Your positivity towards the research was enlivening and it was motivating to hear of the impact it can make. I am grateful for your willingness to be involved and kindness throughout the process.

Thank you to my supervisors, Dr Will Mandy and Professor Paul Burgess, for sharing your expertise, wisdom and enthusiasm for the project. I have learnt a great deal with your guidance and am immensely grateful for your balance of insight and approachability. Thank you also to the former and fellow trainees, students and research assistants in the wider research team who led the way with the project. A huge thank you to Artemis and Lizzie in particular for your valuable contributions and hard work along the way. This research has been a team effort so thank you all for keeping things on track.

To my friends and course squad, thank you for your encouragement, humour and perspective. To Josh, thank you for helping me to make my ideas better and sharing your creative drive and energy. To my parents and Martha, thank you for your faith in me and being the best guinea pigs, sounding board and support team I could ask for. And finally, for my Gran, thank you for your unwavering belief in learning and never failing to say how proud you were. I'm sorry you can't read this. I could not have done it without you all.

Part 1: Literature Review

A Meta-Analytic and Narrative Review of Creative Thinking in Autism Spectrum Disorders

1 Abstract

Aims: Recently published reviews have compared the performance of individuals with and without ASD on fluency tasks, a measure of generativity. These tasks are also used in the measurement of creativity, with performance assessed on various outcomes. Although there is a notion of impoverished creativity and imaginative ability in ASD, conflicting accounts link creative genius and ASD traits. The creative profiles of those with and without ASD are yet to be compared in a systematic way. The present review therefore aimed to compare performance of these two groups on fluency tasks on the outcomes that map onto creativity; namely flexibility, originality and usefulness.

Method: A systematic search of the literature was carried out using PsychINFO, MEDLINE, Embase and ERIC. A total of 15 studies met inclusion criteria. These studies used verbal and non-verbal fluency or divergent thinking tasks to measure creative ideation and products and reported on at least one creativity variable other than fluency. Nine studies reported on an index relating to flexibility: between category switching or within category clustering. A weighted, average effect size was calculated for each variable using a random-effects model. Ten studies reported on either variable of originality or usefulness and were synthesised in a narrative review.

Results: A moderate impairment in flexibly switching between categories during fluency tasks was found in the ASD group compared with matched controls (d = -0.75), although there was significant variability between studies reporting on this outcome (I^2 = 82%). However, the ability to cluster responses within-category was not impaired relative to controls (d = -0.07). The tendency was for the ASD group to be the same or better than controls in originality of responses, although the method of measuring originality varied between studies. However, overall usefulness of responses was lower in the ASD group.

Conclusion: The idea of a creativity deficit in ASD was not supported. Rather, the results indicate a unique profile of creativity in ASD. This is discussed in relation to theories of the creative process, in particular the use of exploitative or persistence strategies in generating creative ideas, which could be a viable route to creative production in ASD. Several methodological and reporting limitations of existing creativity measures are discussed, and recommendations made for future research to comprehensively assess creativity in ASD.

2 Introduction

2.1 Autism Spectrum Disorder (ASD)

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition characterized by deficits in social communication and interaction across contexts, as well as restricted, repetitive patterns of behaviour, interests and activities (American Psychiatric Association, 2013). Around 1% of the population has a diagnosis of ASD, although prevalence rates are increasing, and a higher proportion of males are diagnosed with ASD than females (Baird et al., 2006).

Co-occurring difficulties and comorbid psychiatric problems are frequent; for example, 70% of a community sample with ASD also had at least one other mental disorder according to *DSM-IV* criteria (Simonoff et al., 2008). There is high heterogeneity within the ASD population; previous classification systems attempted to capture this by distinguishing between autism sub-types but now a dimensional approach has been adopted (American Psychiatric Association, 2013). There are no known biomarkers for ASD and so diagnosis is based upon observed behavioural features.

The combination of the aforementioned issues leads to difficulty in providing a coherent, global and universal account of the distinguishing features of ASD.

2.2 Executive Dysfunction Hypothesis

One major cognitive theory that seeks to establish a link between brain and behaviours in ASD is the executive dysfunction theory (Ozonoff, Pennington, & Rogers, 1991). Executive functions (EF) are a set of cognitive capacities involved in the orchestration of goal-directed behaviour (Duncan, 1986). These executive functions have been linked to the frontal lobe and prefrontal cortex in particular, implicated in the regulation and coordination of other brain functions (Miyake & Friedman, 2012).

The executive dysfunction theory suggests that individuals with ASD have deficits in these higher-order cognitive skills such as planning, working memory, set monitoring and shifting, inhibition and generativity; and that these executive functioning impairments may be key for understanding the behavioural phenotype of ASD.

The presence of EF problems in ASD has received support from influential narrative reviews of the literature (Hill, 2004) and more recently published systematic, quantitative analyses (Lai et al., 2017) and there is growing evidence that these difficulties underpin certain behaviours and symptomatology observed in ASD (Pellicano, 2010, 2013; Turner, 1997).

2.3 Generativity: Definition and Measurement

A subtype of EF is generativity, defined as the ability to spontaneously generate appropriate novel responses (Turner, 1999). Restricted, repetitive patterns of behaviour, interests, or activities is specified as one of two core components in the Diagnostic and Statistical Manual of Mental Disorders (5th ed.; DSM-V; American Psychiatric Association, 2013) and these traits are thought to be closely related to a lack of imaginative activity (Wing & Gould, 1979). Both of these features may be underpinned by an impaired capacity for generativity (Turner, 1999).

Traditionally, generativity has been measured by fluency tasks that require the individual to spontaneously generate responses to a single cue or instruction. In verbal fluency tasks, individuals are provided with either phonemic (letter e.g. 'F', 'A' or 'S') or semantic (category e.g. 'animals' or 'clothing') cues and asked to generate as many words as possible within a given time period (usually 60 seconds) (e.g. Benton, 1968; Lezak, Howieson & Loring, 2004). These are essentially semantic or lexical retrieval tasks and require the participant to access their stored knowledge as opposed to producing new and inventive responses (Turner, 1999).

Design fluency tasks are the non-verbal equivalent of the above. An example is when participants are asked to draw as many different designs as possible within four minutes using a set number of lines (Jones-Gotman & Milner, 1977). Unlike verbal tasks, design fluency tasks do not rely on stored knowledge as they instruct specifically for the production of original designs and disallow designs which resemble common or well-known objects and symbols (Turner, 1999).

A further variant of fluency task is known as ideational fluency or divergent thinking (DT) tasks. An often-used example is the Uses of Common Objects task (Guilford, 1967), where participants are required to generate as many novel uses for an object (e.g. a brick or piece of elastic), as possible in a set time frame. Another is the Pattern Meanings Task (Wallach & Kogan, 1965). In this task individuals are shown a range of meaningless line drawings and asked to generate ideas of what the drawings could represent.

2.4 Creativity: Definition and Measurement

This latter type of fluency task is used in the measurement of both generativity and creativity. Generativity is the production of novel appropriate responses and, similarly, creativity is defined as the ability to produce an original and useful idea (Mayer, 1999; Runco & Jaeger, 2012). However, whereas generativity is primarily measured by *quantity* of output, the assessment of creativity is also concerned with the *quality* of these ideas and whether participants can produce something new and interesting as opposed to relying only on stored knowledge (Turner, 1999).

This invariably introduces complexities and varying degrees of subjectivity into the measurement of creativity (Amabile, 1982; Runco, 1988). Generativity may underpin creative thinking but ideational fluency, or number of ideas, appears to be an insufficient criterion alone against which to measure creative ideas or products.

Creativity can be studied from numerous angles and is often conceptualised within the overarching framework of the 4Ps: Person; Place; Process; and Product (Rhodes, 1961). There are numerous methods and tools for measuring creativity depending on the domain of interest; for example, the use of biographical inventories and self-rating attitudinal scales for assessing the creative person (Cropley, 2000). However, the term divergent thinking (DT) has become synonymous with research into the creative process and DT tasks are the most widespread psychometric tool in the measurement of the creativity (Hocevar, 1981).

Divergent thinking is the generation of multiple ideas in response to an open-ended problem and is thought to be involved in the ideation phase of the creative process. Guilford (1956) suggested that various intellectual abilities are incorporated by divergent thinking, and these abilities are possessed by creative individuals. DT tasks are designed to measure these abilities and find individual differences in: fluency (the number of ideas generated); flexibility (the number of different categories implied by the ideas); and originality (the number of unusual or unique ideas) (Hocevar, 1981; Runco, 2014).

The Guilford tradition inspired a legacy of creativity measurement using divergent thinking tasks and the most widely used creativity battery is the Torrance Tests of Creative Thinking (TTCT; Torrance, 1966), consisting of both verbal and figural subtests of divergent thinking. A description of the most widely used fluency and divergent thinking tasks in the measurement of creativity is presented in Table 1.1. The different variables measured by DT tasks and how they relate to creativity are described in more detail below; these are also outlined in Figure 1.1.

Table 1.1 Traditional measures of fluency and divergent thinking used to assess creativity

Task / Measure	Brief Description
TTCT Picture / Figure Completion Tasks (Torrance, 1974)	Participants are asked to make pictures from figures and give titles to their drawings. There are two conditions: (1) Repeated Figures; with a sheet containing ten pairs of parallel lines or (2) Incomplete Figures; a sheet containing ten incomplete figures and meaningless squiggles. Participants are asked to draw anything they want as long as it incorporates the lines or squiggles. No time limit is imposed.
Creativity Assessment Packet Test of Divergent Thinking (CAP; Williams, 1980)	The CAP is a test packet consisting of group administered tests for children. In the Test of Divergent Thinking (Forms A and B), participants are presented with 12 incomplete figures and asked to make them into original drawings and give a title to each. A 20-minute time limit is given.
Figural Synthesis Task (Finke & Slayton, 1988)	Participants are presented with a variety of geometric and alphanumeric shapes and asked to incorporate them into different recognizable designs, using a given number of shapes at once. In the original task participants were instructed to perform a mental combination of the shapes with their eyes closed, then write down the name and draw the resulting image; in other variants participants are given quasi two-dimensional stimuli and asked to arrange them. Participants are told to produce responses that 'look like real things' and to name each response. A variable time restriction of two to five minutes is imposed.
Use of Objects Task / Use of Common Objects Task	Participants are required to generate as many uses as possible for six objects: a brick; a pencil; a mug; 50cm

(Turner, 1999); also similar to TTCT Unusual / Alternative Uses (Torrance, 1974)	length of dowel rod; a 110 x 40cm rectangle of cloth; and a 1m piece of elastic. The first three are considered conventional objects whereas the latter three unconventional. Examples are provided for each object. For each conventional item, the experimenter provides both an established and imaginative function as an example. After examples are provided participants are instructed to “tell me all the other ways in which you think a [object] could be useful’. They are given 2.5 minutes to respond.
Pattern Meanings Task (Wallach & Kogan, 1965)	Participants are presented with six meaningless line drawings presented on individual cards. One is used as a practice stimulus. They are shown the card and asked, “tell me what this could be?”. Participants are given 2.5 minutes to generate as many ideas as possible.
TTCT Toy Improvement Task (Torrance, 1974); also as used in Craig & Baron-Cohen (1999)	Participants are asked to list the most interesting and unusual ways of changing a stuffed toy elephant to make it more fun to play with. The original task uses a picture of a toy elephant, whereas in later adaptations a soft toy elephant is used.

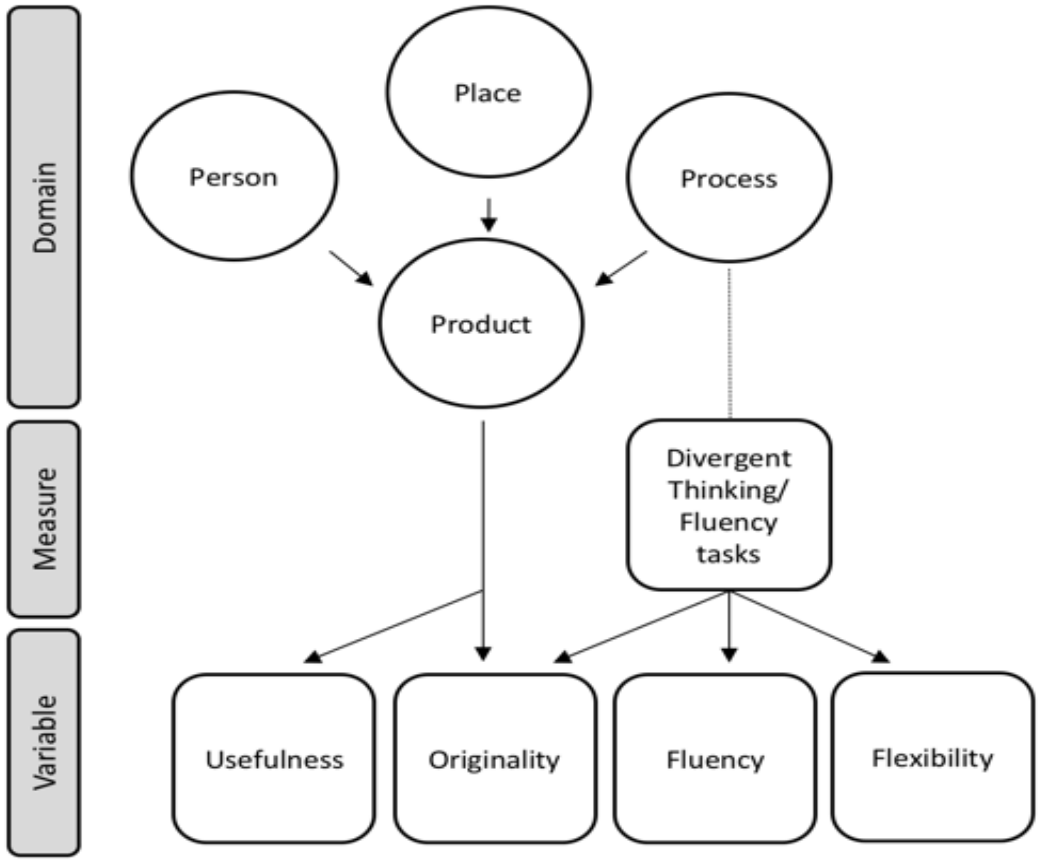


Figure 1.1 Simplified diagram to show the 4Ps creativity framework and relationship between different methods and variables of creativity measurement. Divergent thinking (DT) tasks are designed to map onto the ideation phase of the creative process, requiring the generation of multiple ideas in response to an open-ended task cue.

2.4.1 Fluency

Fluency is measured as the total number of ideas or responses given overall. Fluency is thought to relate to creativity based on the "associative theory" (Mednick, 1962), which suggested that creative ideas are those that are the furthest removed in association from the original thought.

Subsequent research has demonstrated that creative people more readily make remote associations and thus generate a higher number of ideas over time (Benedek & Neubauer, 2013). Furthermore, original ideas tend to come later in a response set once the more obvious ideas have been exhausted (Milgram & Rabkin, 1980).

On the other hand, it has been argued that creativity is a quasi-random process of recombining and connecting ideas such that the number of creative *as well as useless* ideas produced will increase in line with total output, suggesting creative achievement is best determined by level of productivity (Simonton, 1997, 2003).

2.4.2 Flexibility

Flexibility refers to the number of different categories or elements encompassed in the response set. In design and ideational fluency tasks, it is scored by counting the number of shifts from the first picture category (e.g. the Creativity Assessment Packet Test of Divergent Thinking; Williams, 1980) or by the number of different semantic or content categories the responses cover (e.g. the TTCT Toy Improvement Task; Torrance, 1974). In both methods, higher scores indicate a higher number of categories exhibited overall by the responses.

In verbal fluency tasks, flexibility is scored in relation to phonemic and semantic clusters. A semantic cluster is a group of successfully generated words that belong to the same subcategory whereas a phonemic cluster is defined as a group of successively generated words beginning with the same two phonemes (Troyer, Moscovitch & Winocur, 1997).

Runco (1985) suggests that ideational flexibility is an important metric of creativity as flexibly switching between response sets enables the generation of more remote, and therefore original, associations. In support of this, Nijstad, De Dreu, Rietzschel, & Baas (2010) found that participants who use many categories tend to use unordinary categories and generate extraordinary ideas.

Other models of creative ideation suggest that the creative process involves building on existing knowledge through repeated searches in associative memory; and therefore successively generated ideas often share commonalities (Nijstad & Stroebe, 2006). Efficient performance on verbal fluency tasks (i.e. generating as

many words as possible in relation to a phonemic or sematic cue) involves the clustering of responses to exhaust a phonetically or semantically related category before switching to a new subcategory (Troyer et al.,1997). Furthermore, it has been argued that creativity can be achieved through a systematic and persistent search process (Dietrich, 2004) and there is evidence that this yields original ideas (Nijstad et al., 2010).

This implies that different strategies may be used in the generation of creative ideas: both flexibly switching between categories to broaden perspective and generate remote associations; and clustering responses to narrow the focus and incrementally progress towards an original idea. Nijstad, Stroebe & Lodewijkx (2002) propose that both the number of categories and the number of ideas generated within each category should be measured in fluency and divergent thinking tasks to more fully examine these aspects of the creative process.

2.4.3 Originality

Originality refers to the extent that an idea or response can be considered novel or unusual (Mayer, 1999; Runco, Illies & Eisenman, 2005). The standard definition of creativity in the literature adopts the dual components of originality and usefulness (Flaherty, 2005; Runco & Jaeger, 2012) and it is therefore a necessary criterion for measuring creative performance. These criteria can be applied to the creative product.

In the case of DT tasks, originality is frequently scored as uniqueness and measured in relation to statistical norms, with points awarded for infrequently observed responses (Wallach & Kogan, 1965; Torrance, 1974). However, other researchers have reported the merits of using subjective scoring scales with DT tasks, with high reported reliability using only two to three raters (Silvia et al., 2008).

Other tools employ criterion-referenced methods, such as in the Creativity Assessment Packet Test of Divergent Thinking, where high originality scores are

awarded based on where the participant has drawn in relation to the frame (Williams, 1980), the rationale being that creative people will be less restricted by the closed stimuli.

2.4.4 Usefulness

As aforementioned, usefulness is widely accepted as the complementary criterion of creativity. Originality is often considered the primary or more desirable attribute of creative ideas (Nijstad et al., 2010); however, without also being useful or feasible these ideas may be bizarre rather than creative. Cropley, Kaufman & Cropley (2008) insist upon effectiveness and relevance in their definition of functional creativity, stating the product must satisfy the need for which it is created.

Despite the bipartite definition, originality and feasibility ratings of ideas in ideational fluency tasks are found to be negatively correlated; that is, more original ideas tend to be viewed as less feasible (Nijstad et al., 2010; Runco et al., 2005). This paradox is also observed beyond experimental settings, as highly original ideas that diverge from the established paradigm are often devalued or dismissed, although in some instances are later recognised as highly creative.

The apparent discrepancy between originality and usefulness scores raises issues for the measurement of creative products and the difficulty with operationalizing and integrating these criteria has meant that usefulness is often not measured or reported in DT tasks at all, such as the TTCT.

2.4.5 Summary

In summary, the distinct yet interrelated nature of these creativity variables emphasises the necessity of measuring the components of fluency, flexibility and originality to accurately interpret divergent thinking potential (Torrance, 1974). I would further argue that usefulness is an essential criterion and should also be assessed in DT tasks to accurately assess the creativity of products (ideas).

2.5 Creativity and ASD

The literature examining creativity in ASD is inconsistent. Looking at fluency in particular, recent meta-analyses have demonstrated that individuals with ASD are impaired on traditional measures of generativity relative to typically developing controls and a moderate overall effect size for reduced fluency has been illustrated (Demetriou et al., 2017; Lai et al, 2017). At the task level, a medium impairment in fluency was found for phonemic, semantic and ideational fluency tasks and a smaller but still significant impairment was found for design fluency tasks (Pullinger, 2017).

These analyses reveal differences in productivity at a group level during DT tasks. However, Pullinger (2017) highlighted that aspects of performance, such as the way individuals with ASD cluster and switch their responses (flexibility) and the prototypicality of responses (originality), remain to be explored in a systematic way. Qualitative analysis of fluency task performance may therefore yield useful findings relating to the strategic processes employed in such tasks and further elucidate the executive and creative profiles of ASD.

DT tasks have been used to measure creativity in children with ASD and have found poorer ideational fluency relative to controls (Craig and Baron-Cohen, 1999). However, other studies have found that whilst individuals with ASD or ASD traits produce fewer responses overall in DT tasks, a higher proportion of these ideas are original relative to controls (Best, Arora, Porter & Doherty, 2015; Liu, Shih & Ma, 2011). It would appear that the most we can conclude from any substantial literature so far is that individuals with ASD produce less ideas overall than controls in fluency tasks; however, the ability to engage in divergent thinking and produce creative ideas is far from determined.

Furthermore, there exists a disparity in the literature between performance on psychometric process measures and studies of the creative person. Fitzgerald (2004) notes the correlation between supposed high autistic traits and creative

achievements in certain eminent male figures, such as the writer Lewis Carroll and philosopher Ludwig Wittgenstein. The general population associate ASD with traits such as intelligence and creativity (Jensen et al., 2015) and creative savant abilities (Treffert, 2014). Creativity and intellectual strengths are also frequently endorsed personal character values within the ASD population (Kirchner, Ruch & Dziobek, 2016). There appear to be commonly held views about creativity and ASD in the general population, but these beliefs appear to be based on eminent individuals and stereotypes as opposed to empirical evidence.

2.6 Limitations of Divergent Thinking Tasks

A possible reason for the discrepancy in findings could be due to limitations in methodology. Originality is traditionally scored by awarding points for infrequent responses and summing the points (Silvia et al., 2008). However, this perhaps oversimplifies what it is to be original: not just uncommon, but novel, surprising or innovative (Boden, 2004).

Despite the consensus definition in the creativity literature, the criterion of usefulness is often neglected in the most widely used creativity batteries such as the TTCT, possibly due to the abstract nature of the tasks (Zeng, Proctor & Salvendy, 2011). This failure to operationalize the standard definition of creativity reduces the validity of DT tasks in measuring creative products.

The psychometric properties of DT tests also pose further limitations. Several studies have cited high correlation between subscale scores on the TTCT (Kim, 2006) suggesting originality is conflated with fluency scores (Dixon, 1979), reducing construct validity. Factor analytic results have failed to delineate the subscales of the TTCT and instead suggest taking a composite score as the primary outcome variable (Heausler & Thompson, 1988). However, this neglects important theoretical distinctions between the various components of creativity and Torrance specifically discourages the use of a composite score (Torrance, 1974). The result in practice is

that the scores on the TTCT are misrepresented or difficult to interpret, limiting the research validity of the tool.

DT tests have been shown to have inconsistent convergent validity with other measures of creativity (Hocevar, 1981) and poor ecological and discriminant validities (Zeng et al, 2011). In a large scale longitudinal study the TTCT was a moderate predictor of later life self-reported creative achievements (Cramond, Matthews-Morgan, Bandalos & Zuo, 2005; Torrance, 1981), although this finding has not been replicated to the same extent in other studies (Stenberg & Lubart, 1996).

Earlier research questioned the administration and testing conditions of DT tests, suggesting creativity may be suppressed in traditional testing environments (Wallach & Kogan, 1965) and emphasising the importance of predicting real world creative potential rather than tested ability (Wallach, 1976; Wallach & Wing, 1969). Experience and expertise within a particular domain is closely tied to creativity (Weisberg, 2006) and as DT tasks lack domain specificity they are unlikely to be accessing an individual's full creative potential. This could be particularly prominent in an ASD population where narrow and specialized interests are often a defining characteristic and where it might be expected to find domain expertise.

2.7 Rationale for Present Review

To our knowledge, there are no existing systematic reviews of creativity and ASD. Creativity holds intrinsic value on an individual, societal and organisational level (Plucker, Beghetto & Dow, 2004) and yet is a neglected research topic in the psychology literature (Sternberg & Lubart, 1996). The study of creativity and ASD has received even less attention and yet there exist conflicting accounts depending on the method and approach. The study of creativity in ASD thus far has failed to demonstrate a specific or universal deficit and the pattern of strengths and weaknesses in the autistic profile is unclear.

There is an overlap in the literature between creativity and generativity and considerable similarity between fluency and divergent thinking type tasks, which assess the creative ideation process. However, recent reviews have focused exclusively on the domain of fluency and other variables of creativity have yet to be systematically reviewed and synthesised. Analysis of process outcomes (fluency and flexibility) lends itself to meta-analytic techniques due to the widespread use and similarity of fluency type tasks. Assessment of the quality of creative products (originality and usefulness) is appropriate to more qualitative methods of review due to the disparity in operationalised definitions of originality and utility.

Therefore, this review aims to provide a comprehensive quantitative and narrative synthesis of performance of individuals with ASD on fluency and divergent thinking tasks of creativity. Specifically, it aims to investigate:

- How do individuals with ASD cluster and switch their response sets on fluency tasks, relative to controls?
- Do individuals with ASD generate original and useful responses on fluency tasks, relative to controls?
- How can performance on these fluency tasks inform our understanding of the creative profiles of individuals with ASD?

3 Method

3.1 Search Strategy and Selection Criteria

A systematic literature search was carried out on the 6th, 8th and 13th September 2017 using four electronic databases: PsychINFO, MEDLINE, Embase and ERIC. Key concepts were defined as creativity, divergent thinking, generativity and ASD (see Figure 1 for details). Search terms and synonyms were identified under these headings. The search strategy was broadened by including keywords and subject headings tagged within articles found after a preliminary search.

Duplicates were removed and then abstracts screened per the inclusion and exclusion criteria below.

3.2 Inclusion Criteria

Studies were included that met the following criteria: (1) empirical paper measuring creative thinking or products; (2) including at least one fluency or divergent thinking task; (3) ASD participants satisfied formal diagnostic criteria according to the DSM (3rd, 4th or 5th edition; APA, 1987, 2000, 2013) or International Statistical Classification of Diseases and Related Health Problems, 10th revision (ICD-10; World Health Organisation, WHO, 1992); (4) included a matched control group; (5) published in a peer reviewed journal.

3.3 Exclusion Criteria

Studies were excluded if: (1) the only dependent variable reported was fluency or number of correct responses; (2) they were not published in English; (3) they used single case-study or case report design; (4) they were brain imaging studies that modified task conditions; (5) they specifically measured speech/reading fluency or narrative production; (6) they were observational studies of play behaviour; (7) they exclusively measured imaginative ability or use of pretence; (8) they exclusively measured other components of EF.

3.4 Studies Included in Review

Only articles that appeared to meet the above criteria were retrieved for more detailed evaluation. Studies were sought that analysed creativity task performance on variables other than fluency, in particular flexibility, originality and usefulness. This strategy aimed to extend findings of recent reviews focusing exclusively on fluency and address the limitations of assessing only one dimension of creativity.

Where it was not clear (e.g. dependent variables not described), the article was retrieved to minimise omissions. A handsearch of references also provided additional articles that were then screened following the same procedure. If articles met the criteria they were included regardless of whether assessment of creativity was the primary aim of the study.

The studies were grouped based on the reported dependent variables. Those studies that reported on flexibility were suitable for meta-analysis (process described in next section, Statistical Procedures). Due to variation in measurement and reporting of originality and usefulness scores, a review of performance in these domains is not suited for a quantitative review. A narrative synthesis was therefore adopted for these studies.

3.5 Study Quality

The papers included in the final analyses were assessed for quality in relation to validity threats and attempts to minimise sources of bias. Numerous critical appraisal tools exist and there is no 'gold standard' tool for assessing health and medical research (Katrak, Bialocerkowski, Massy-Westropp, Kumar & Grimmer, 2004).

Therefore, the Standard Quality Assessment Criteria for Evaluating Primary Research Papers (Kmet, Lee, & Cook, 2004) was chosen and adapted for the present review (for example, by adding a question relating to ASD diagnosis and screening procedures). The final scale consisted of 13 questions rated as Yes (2), Partial (1) or No (0) relating to the quality of methodology and reporting in each paper, yielding a total possible score of 26. A copy of this is included in Appendix A.

4 Statistical Procedures

4.1 Choice of Measures

Most studies used a variant of verbal fluency tasks. These included both semantic and letter versions and followed the same format whereby participants were asked to generate as many category/letter exemplars as possible within 60 seconds and were variations of the original 'FAS' verbal fluency task (Benton, 1968).

Design fluency tasks were largely the same across the literature. These mainly resembled the figure completion tasks in the TTCT (see Table 1.1) although not all required participants to name their creations. As designs and titles are scored separately the titles were excluded from analysis and only data relating to the designs themselves was extracted.

Only one study reporting a flexibility score used a measure of ideational fluency, the Toy Improvement task.

4.2 Dependent Variables of Measures

There was some variation in how the dependent variable of flexibility was scored in the fluency tasks. Therefore, to reduce bias, the dependent variables were selected before the data were extracted.

4.2.1 Design and Ideational Fluency Tasks

In design and ideational fluency tasks, flexibility was reported as the number of shifts from the first category in the response set, or as the number of different content categories the responses cover. In both methods, higher scores indicated a higher number of categories exhibited overall by the responses. This was taken as the primary outcome variable in these tasks.

4.2.2 Verbal Fluency Tasks

For verbal fluency tasks, flexibility was scored in relation to phonemic and semantic clusters. All studies defined a semantic cluster as a group of successfully

generated words that belong to the same subcategory (e.g. farm animals) and a phonemic cluster as a group of successively generated words beginning with the same two phonemes (e.g. dolphin and dog).

The majority of studies that included a measure of verbal fluency reported on the number of switches between sets of clusters or unclustered words and this was therefore taken as the primary outcome measure for verbal fluency tasks. When this was not available, the total number of clusters overall or the proportion of responses belonging to a cluster was taken as the most closely related variable.

Most studies also reported the cluster size, and this was entered into a second analysis. Only one study (Turner, 1999) did not report cluster size and therefore proportion of responses belonging to a cluster was taken as the dependent variable.

Typically, studies reported phonemic clusters for letter fluency tasks and semantic clusters for semantic fluency tasks (Carmo et al., 2015). However, some studies reported both types of clusters for each type of task. In cases where both types of cluster were reported, the estimated effect sizes were averaged to provide a single summary statistic for each study. This avoided selective data extraction based on a priori theoretical distinctions of typical or appropriate cluster generation. Furthermore, there is a basis for this method of combining across outcomes that measure the same underlying cognitive concept (in this case, switching or clustering) within the neuropsychological literature (Wykes, Huddy, Cellard, McGurk & Czobar, 2011).

4.3 Effect Size Calculation

All effect sizes were calculated using means and standard deviations (SD). Where studies split the sample into groups based on ASD diagnosis, the groups were combined to form one clinical and one control group.

To calculate the mean, the following formula was used:

$$M_{pooled} = \frac{N_1 M_1 + N_2 M_2}{N_1 + N_2}$$

The N_1 and N_2 denote the sample size of the groups and M_1 and M_2 denote the group means. The combined SD was calculated using the following formula:

$$SD_{pooled} = \sqrt{\frac{SD_1^2 + SD_2^2}{2}}$$

The SD_1 and SD_2 denote the SDs of each group. This follows the Cochrane guidelines for combining groups when performing data extraction for continuous outcome variables (Higgins & Green, 2011).

Effect size was calculated using Cohen's *d*. The following formula was used:

$$Cohen's\ d = \frac{M_1 - M_2}{SD_{pooled}}$$

A negative Cohen's d statistic represented lower creativity scores in the ASD group in comparison to the control group. The standard convention was taken that 0.2 is a small effect size, 0.5 is a moderate effect size and 0.8 is a large effect size (Cohen, 1988). The effect size estimate is slightly biased and gives a slightly larger than true estimate of the population value and was therefore corrected using a factor provided by Hedges and Olkin (1985, p.80).

4.4 Heterogeneity of Effect Sizes

In order to determine the generalisability of the findings, the consistency of effects across studies was determined using estimates of the heterogeneity of effect sizes. A measure of the statistical heterogeneity between studies was calculated using Cochran's chi-squared (Q) test, where a P value of <0.1 indicates significant heterogeneity between studies (a higher significance value is taken due to the low power of the test).

However, as Q has low power especially when there are a small number of studies, the I^2 statistic was also determined. This describes the percentage of total

variation across studies that is due to heterogeneity rather than chance and is not dependent on the number of studies included (Higgins & Thompson, 2002). The classification for I^2 is that 25% is low heterogeneity, 50% is moderate heterogeneity and 75% is high heterogeneity between studies (Higgins, Thompson, Deeks & Altman, 2003).

When high heterogeneity between studies was found, a sensitivity analysis was performed by re-running the meta-analysis after removing low quality studies suspected of skewing the results from the sample to determine the robustness of the results (Bown & Sutton, 2010).

4.5 Publication Bias

Publication bias refers to the tendency for authors to submit, and journals to publish, positive results over negative or non-significant findings. As the analyses included studies published in peer-reviewed journals only, publication bias was assessed by visual inspection of a funnel plot. Asymmetry indicates 'missing' studies in the literature; however, an asymmetrical plot should not be equated with publication bias because there are a number of possible causes (Higgins & Green, 2011; Sterne et al., 2011).

Therefore, contour enhanced funnel plots were also constructed using Stata 14 software (StataCorp, 2015), allowing the statistical significance of study estimates to be considered by adding contour lines onto the graph depicting the range for different significance values (Peters, Sutton, Jones, Abrams & Rushton, 2008). This overlay aids interpretation of the potential causes of asymmetry as publication bias will likely result in a lack of studies in areas of non-significance. Other sources of asymmetry may be poor methodological quality (leading to over-inflated effects in smaller studies), true heterogeneity or chance (Egger, Smith, Schneider & Minder, 1997).

However, due to the small number of studies included in the analyses caution should be taken when interpreting the plots (Brown & Sutton, 2010).

4.6 Statistical Procedures

Meta-analytic procedures were run using RevMan 5.3 (The Cochrane Collaboration, 2014). A random effects model was used allowing for an estimate of heterogeneity in the weighting of the studies and variation between study outcomes along a normal distribution. This was used rather than a fixed effects model, which makes the assumption that the studies examined as a whole were performed under similar task conditions, using similarly defined outcomes with similar samples. These assumptions were not met by the studies included in the current review due to differences in sample characteristics (both child and adult samples) and non-standardisation of task administration and conditions, making a random effects model more appropriate.

5 Results

A flow diagram depicting the number of papers retrieved and included in the final analyses is shown in Figure 1.2. A description of the studies included in the final analyses is displayed in Table 1.2 along with study quality scores.

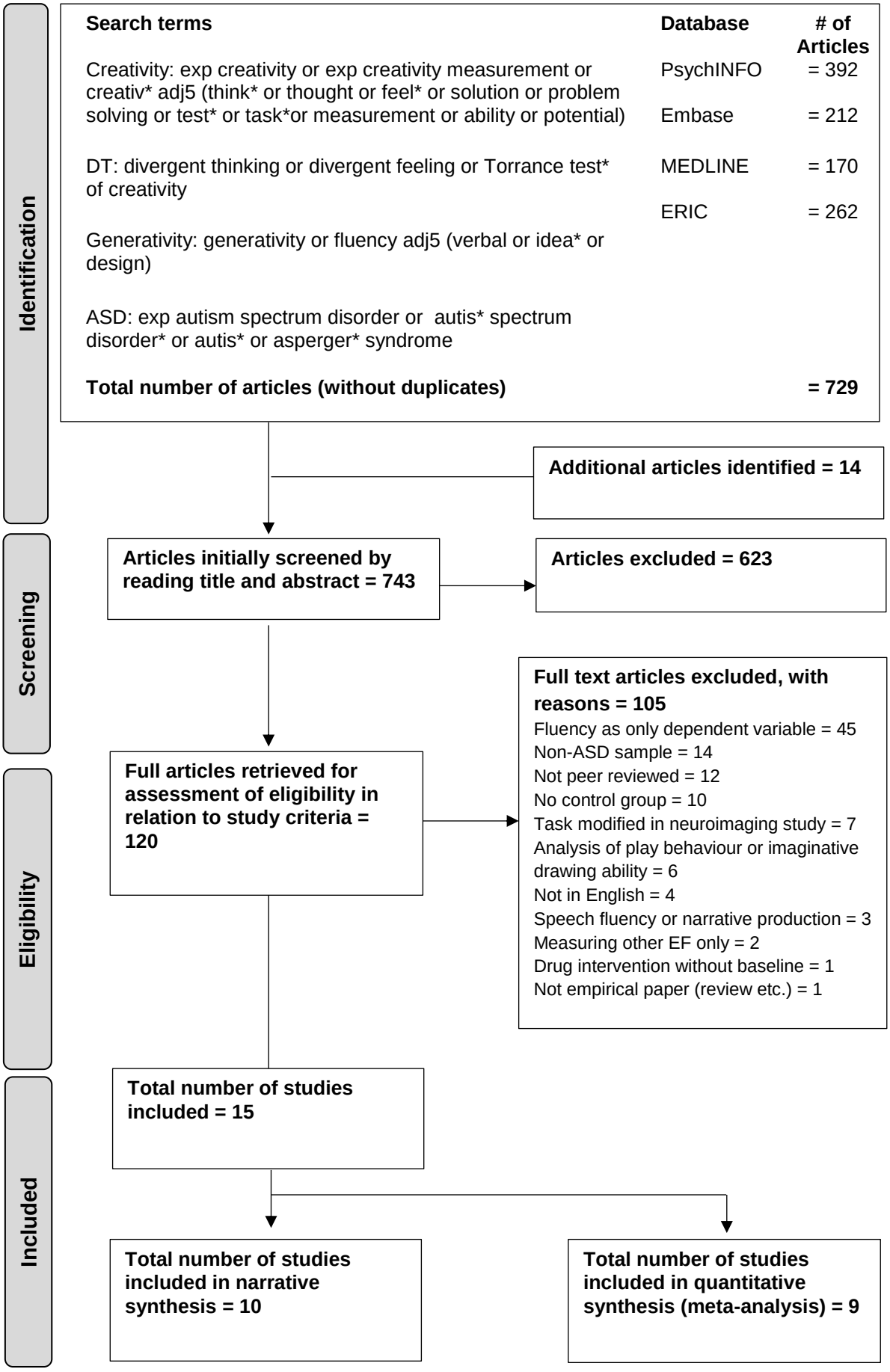


Figure 1.2 Flow diagram depicting search strategy according to PRISMA criteria

Table 1.2 Summary of studies comparing creativity on fluency tasks in children and adults with ASD and matched controls

Study authors, year	**Sample size**	**Gender M:F**	**Mean age, years (SD)**	**Mean IQ (SD)**	**Diagnostic method**	**IQ assessment**	**Matching**	**Type of task (*name of assessment/ battery*)**	**Dependent variable**	**Study quality score**
Pastor-Cerezuela et al., 2016	47 ASD 53 TD	40:7 43:10	6.67 (1.14) 6.74 (1.08)	98.89 (19.52) 99.64 (16.76)	DSM-IV-TR, ADOS, GARS-2	Raven	A, G, NVIQ	CFT (*ITPA*)	Number of switches Mean cluster size	22
Pring et al., 2012	9 ASD 9 MLD	7:2 7:2	32.22 (6.59) 33.56 (5.49)	82.33 (16.59) 83.55 (19.19)	MLD screened with measure adapted from ADOS	Raven[a], PPVT	A, G, VIQ	DFT (*TTCT, FST*)	Flexibility	21
Spek et al., 2009	62 HFA/Asp 30 Control	57:5 28:2	39.67 (11.41) 39.89 (11.45)	113.33 (14.57) 116.77 (11.33)	ADI-R, DSM-IV-TR, ICD-10	WAIS-III VCI	A, E, G, VIQ	LFT CFT (*GIT*)	Relative number of switches Mean cluster size	21

Study authors, year	**Sample size**	**Gender M:F**	**Mean age, years (SD)**	**Mean IQ (SD)**	**Diagnostic method**	**IQ assessment**	**Matching**	**Type of task (*name of assessment/ battery*)**	**Dependent variable**	**Study quality score**
Dichter et al., 2009	39 ASD	38:1	9.72 (2.66)	101.69 (17.5)	DSM-IV, ADI-R, SRS	Leiter-R	A	IFT (*UCO*)	Number of unusual responses	21
	39 TD	38:1	10.57 (3.35)	111.67 (16.11)						
Channon et al., 2001	15 Asp	13:2	13.89 (2.19)	13.60 (2.25)	DSM-IV	Raven	A, G, NVIQ, CELF-R	IFT	Problem appreciation Social appropriateness Effectiveness	20
	15 TD	13:2	14.38 (2.00)	12.37 (2.37)						
Bishop & Norbury, 2005	14 HFA	14:0	8.30 (0.99)	107.21 (15.62)	SCQ, ADOS-G	Raven	A, NVIQ	IFT (*UCO*)	Number of not useful responses	20
	18 TD	15:3	8.56 (1.00)	110.83 (10.83)						

Study authors, year	**Sample size**	**Gender M:F**	**Mean age, years (SD)**	**Mean IQ (SD)**	**Diagnostic method**	**IQ assessment**	**Matching**	**Type of task (*name of assessment/ battery*)**	**Dependent variable**	**Study quality score**
Carmo et al., 2015	20 HFA	19:1	25.25 (6.71)	103.80 (9.11)	DSM-IV, ASDS, ADOS	WAIS-III	A, E, G, VIQ	LFT CFT	Relative number of switches Proportional cluster size	19
	20 TD	19:1	25.05 (7.63)	109.95 (13.82)						
Kasirer & Mashal, 2016	34 ASD	29:5	12.59 (1.92)	33.35 (6.85)	DSM-IV-TR, SCQ	TONI-3[a], WISC-IV vocabulary subtest	A, G, NVIQ, Hebrew naming test, vocabulary	IFT	% of original responses	19
	39 TD	28:11	12.26 (1.58)	31.85 (6.57)						
Begeer et al., 2014	26 ASD	23:3	13.67 (6.08)	109 (12.2)	DSM-IV, SRS	PPVT	A, G, PPVT	CFT	Relative number of switches Proportional cluster size	18
	26 TD	22:4	11.67 (5.08)	109 (9.5)						

Study authors, year	**Sample size**	**Gender M:F**	**Mean age, years (SD)**	**Mean IQ (SD)**	**Diagnostic method**	**IQ assessment**	**Matching**	**Type of task (*name of assessment/ battery*)**	**Dependent variable**	**Study quality score**
Dunn, Gomes & Sebastian, 1996	10 Autistic 10 Normal	NR NR	6.79 (1.90) 4.93 (1.51)	102.40 (10.06) 106.40 (12.10)	DSM-III-R	SB-4	NVIQ, PPVT-R	CFT (*MSCA, CELF-R*)	Prototypicality ratings	18
Kasirer & Mashal, 2014	17 ASD 17 TD	14:3 8:9	21.06 (3.44) 22.71 (2.02)	38.8 (4.13) 40.71 (1.89)	DSM-IV, AQ	TONI-3[a], WAIS vocabulary subtest	A, NVIQ, Hebrew naming test	IFT	% of original responses	18
Inokuchi & Kamio, 2013	30 HFASD 18 Control	25:5 15:3	19.20 (2.60) 20.10 (2.00)	99.6 (12.8) 101.9 (13.9)	DSM-IV-TR	WAIS-R	A, G, IQ	LFT CFT	Number of clusters Mean cluster size	18

Study authors, year	**Sample size**	**Gender M:F**	**Mean age, years (SD)**	**Mean IQ (SD)**	**Diagnostic method**	**IQ assessment**	**Matching**	**Type of task (*name of assessment/ battery*)**	**Dependent variable**	**Study quality score**
Turner, 1999	22 HFA	19:3	12.00 (5.33)	100 (22.3)	DSM-III-R	WAIS/WISC VIQ[a], Raven	A, IQ	LFT CFT IFT (*UCO, PM*)	% of responses forming cluster Proportion of highly imaginative responses	17
	21 HFC	18:3	11.92 (4.42)	101 (17.8)						
Lui, Shih & Ma, 2011	16 Asp	16:0	10.60 (NR)	99.3 (15.4)	DSM-IV	TONI-3	G, SES	DFT (*CAP*)	Flexibility Originality	16
	42 Control	42:0	10.40 (NR)	98.7 (13.1)						
Craig & Baron-Cohen, 1999	30 Autism/Asp	NR	12.75 (2.81)	8.29 (2.77)	DSM-IV, ICD-10	TROG	VMA (Autism and MLD only)	DFT (*TTCT*) IFT (*TTCT*)	Flexibility Originality	15
	15 MLD	NR	12.33 (2.33)	6.75 (1.66)						

[a] Reported IQ score where multiple assessments used

Sample characteristics are reported using the original author's language: ASD = Autism Spectrum Disorder; Asp = Asperger Syndrome; HFA = High functioning autism; HFC = High functioning controls; MLD = Moderate learning disabilities; TD = Typically developing

General abbreviations: A = Age; E = Education; G = Gender; NVIQ = Non-verbal IQ; NR = Not reported or insufficient information available; SD = Standard deviation; SES = Socio-economic status; VCI = Verbal Comprehension Index; VIQ = verbal IQ; VMA = verbal mental age

Abbreviations for tasks/measures: ADI-R = Autism Diagnostic Interview-Revised; ADOS = Autism Diagnostic Observation Schedule; ASDS = Asperger's Syndrome Diagnostic Scale; AQ = Autism-spectrum Quotient; CFT = Category Fluency Task; CAP = Creativity Assessment Packet; CELF-R = Clinical Evaluation of Language Fundamentals-Revised; DFT = Design Fluency Task; DSM = Diagnostic and Statistical Manual; FST = Figural Synthesis Task; GARS-2 = Gilliam Autism Rating Scale, Second Edition; GIT = Groninger Intelligentie Test; IFT = Ideational Fluency Task; ITPA = Illinois Test of Psycholinguistic Aptitudes; Leiter-R = Leiter International Performance Scale-Revised; LFT = Letter Fluency Task; MSCA = McCarthy Scales of Children's Abilities; PM = Pattern Meanings; PPVT-R = Peabody Picture Vocabulary Test-R; Raven = Raven's Coloured or Standard Progressive Matrices; SCQ = Social Communication Questionnaire; SB-4 = Stanford Binet Intelligence Scale, 4th Edition; SRS = Social Responsiveness Scale; TONI-3 = Test of Nonverbal Intelligence; TROG = Test of Reception of Grammar; TTCT= Torrance Tests of Creative Thinking; UCO = Uses of Common Objects; WAIS = Weschler Adult Intelligence Scale; WISC = Weschler Intelligence Scale for Children

5.1 Flexibility

5.1.1 Switching

Nine studies were entered into the meta-analysis of flexibility as measured by switching and category generation on the fluency tasks. In total, 213 ASD participants were compared to 218 control participants. The results of this analysis are presented in Figure 1.3. The overall effect of ASD on flexibility as assessed by fluency tasks was estimated to be d = -0.75 (95% CI = -1.25 to -0.25). This is a medium effect size according to Cohen (1992). There was a significant group difference overall (p = 0.003).

Heterogeneity between studies was significantly greater than chance overall (Q = 45.08, df = 8, p < 0.001) and the degree of heterogeneity was considered to be high (I^2 = 82%).

Publication bias was assessed by means of funnel plots. Inspection of the funnel plot shows an asymmetrical scatter, with two outlying studies on the left side (Figure 1.4). However, as the smaller studies fall mainly within the areas of statistical non-significance this would suggest publication bias is not the cause of the asymmetry (Figure 1.5). The two outlying studies (Craig & Baron-Cohen, 1999; and Liu, Shih & Ma, 2011) appear to have contributed to the significant heterogeneity found between studies, with perhaps inflated effect sizes due to their lower quality (both studies received the lowest quality ratings out of the included studies; see Table 1.2).

Due to the significant heterogeneity, a sensitivity analysis was run removing these two low quality outlying studies (Craig & Baron-Cohen, 1999; Liu, Shih & Ma, 2011), leaving a total of seven studies in the meta-analysis of switching and category generation. This compared 167 ASD participants to 161 control participants. The direction of the result was replicated, d = -0.41 (95% CI = -0.64 to -0.19), with the ASD group scoring significantly lower on this index of flexibility than

the control group ($p < 0.001$). However, the magnitude of this difference was reduced to a small effect size according to Cohen (1992). Heterogeneity between studies became non-significant ($Q = 5.90$, $df = 6$, $p = 0.43$, $I^2 = 0\%$).

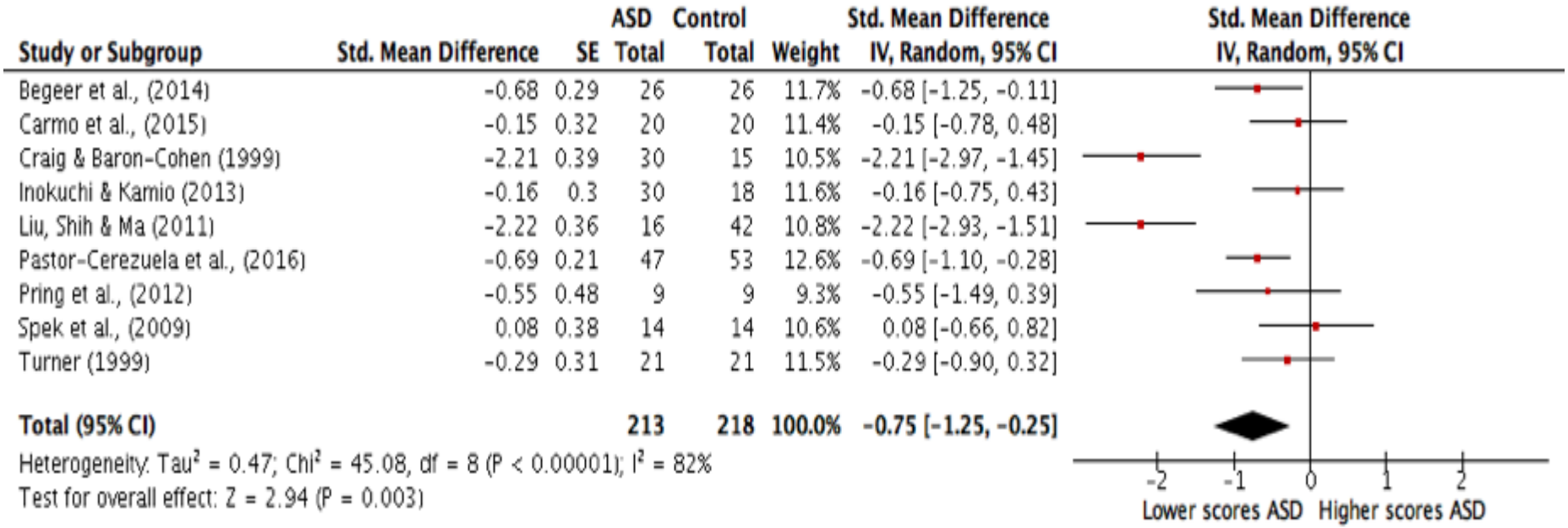


Figure 1.3 Forest plot for studies comparing switching on fluency tasks for ASD and control participants

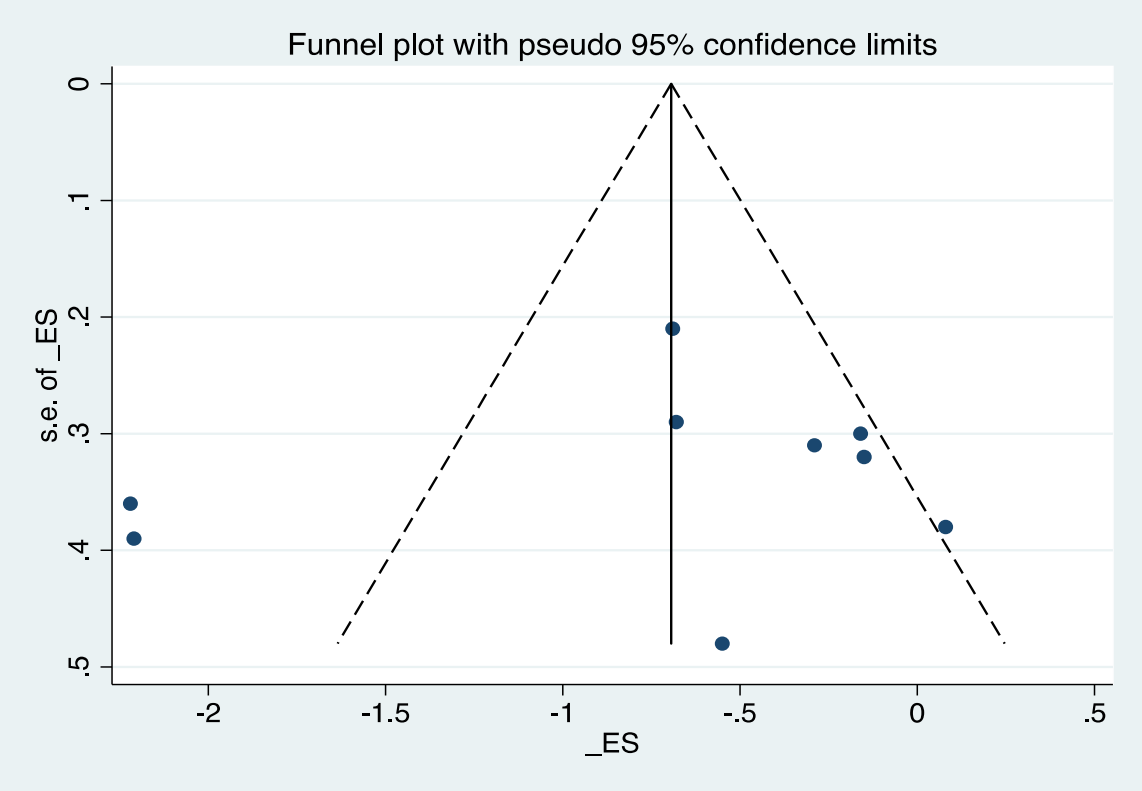


Figure 1.4 Funnel plot for studies assessing switching on fluency tasks for ASD and control participants. The effect size is plotted on the x axis and the standard error along the y axis.

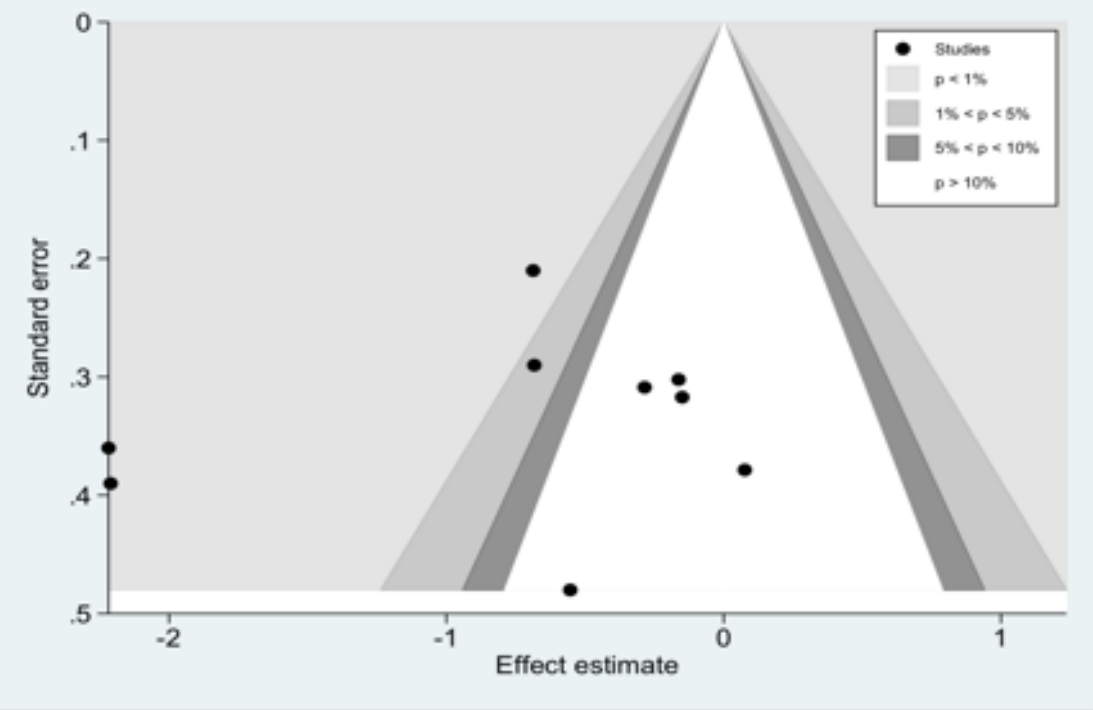


Figure 1.5 Contour enhanced funnel plot for studies assessing switching on fluency tasks for ASD and control participants. Contour lines correspond to milestones of statistical significance.

5.1.2 Cluster size

Six studies were entered into the analysis of cluster size generated on verbal fluency tasks. Overall 159 ASD participants and 152 control participants formed the sample groups. The results of this analysis are presented in Figure 1.6. The overall effect of ASD on cluster size produced in verbal fluency tasks was *d* = -0.07 (95% CI = -0.36 to 0.21). This is a very small observed group difference and did not reach significance (*p* = 0.62). Heterogeneity between studies was not significant overall (Q = 7.69, df = 5, *p* = 0.17, I^2 = 35%).

Inspection of the funnel plot revealed some spread, with the largest studies scattering more widely towards the limits of the 95% confidence interval, shown by the dotted lines in Figure 1.7. However, the studies fall within the triangular region enclosed by the limits, which would indicate the absence of biases and heterogeneity (Higgins & Green, 2011). Furthermore, the plot appears largely symmetrical and there does not appear to be evidence of publication bias due to the distribution of the studies within zones of non-significance (Figure 1.8).

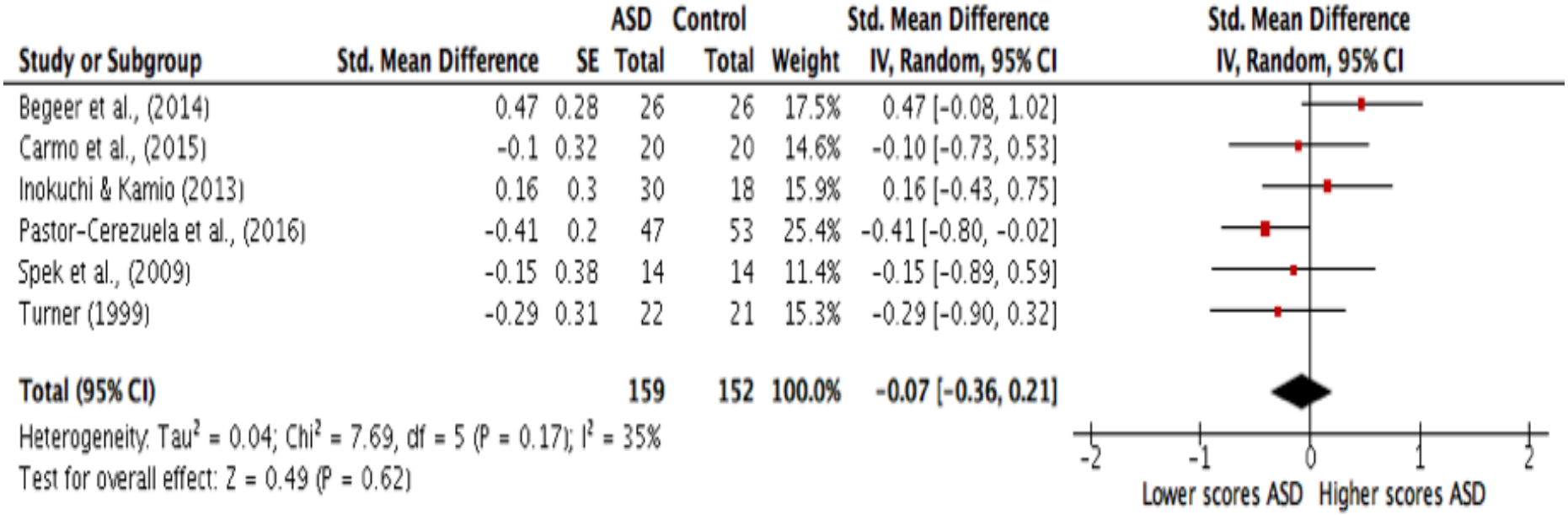


Figure 1.6 Forest plot for studies comparing cluster size in fluency tasks for ASD and control participant

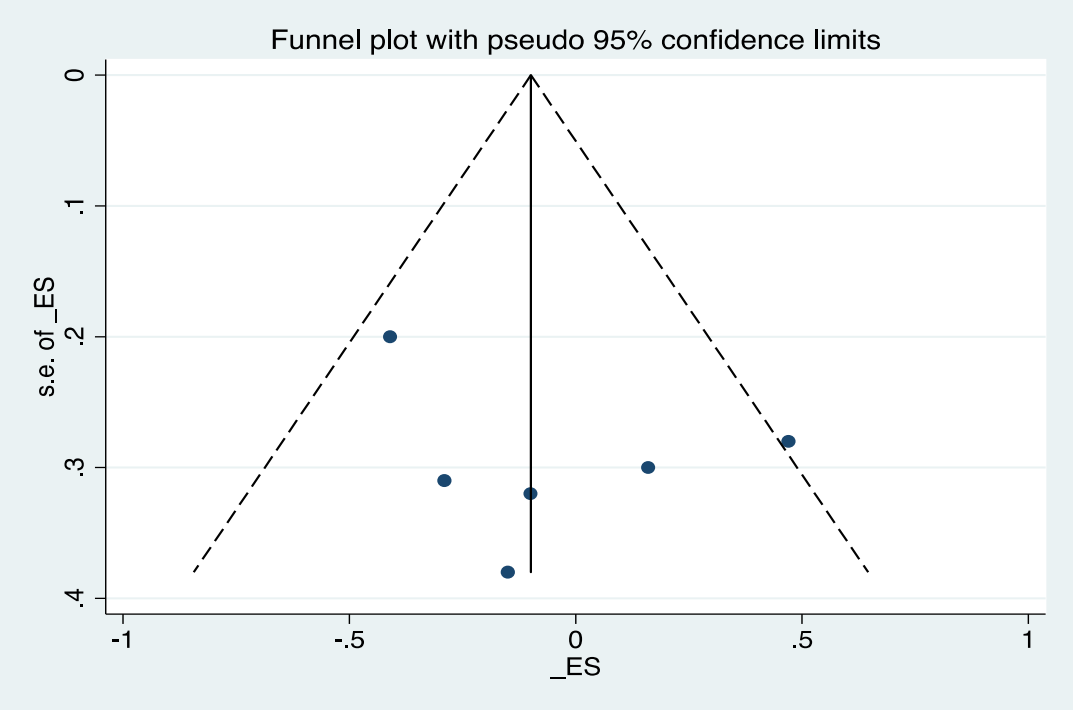


Figure 1.7 Funnel plot for studies comparing cluster size in fluency tasks for ASD and control participants. The effect size is plotted on the x axis and the standard error along the y axis.

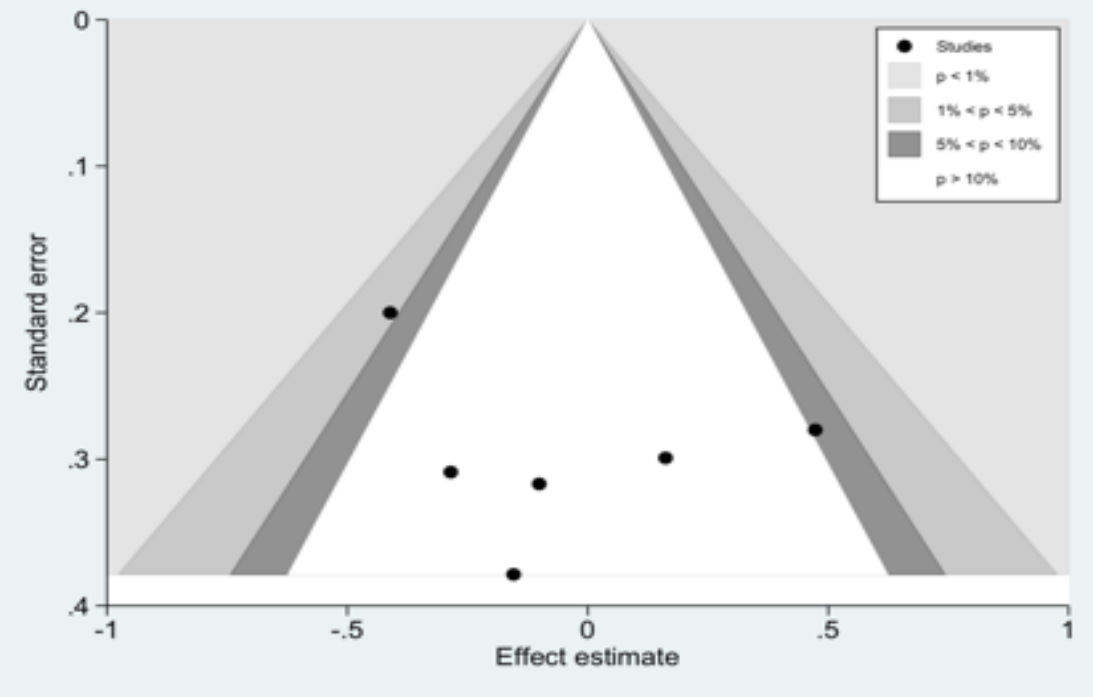


Figure 1.8 Contour enhanced funnel plot for studies assessing cluster size on fluency tasks for ASD and control participants. Contour lines correspond to milestones of statistical significance.

5.2 Originality

A total of eight studies reported on the originality of responses given in fluency tasks. Studies either defined originality as a unique or statistically infrequent response, measured against standardised norms; or allocated a score based on certain criteria. These two types of study are reviewed separately in the following sections. The overall group comparisons are summarised in Table 1.3.

5.2.1 Norm-Referenced

There were three studies using standardised norms, which provide a somewhat contradictory picture of whether ASD is associated with lower originality on standardised measures.

The largest study by Craig & Baron-Cohen (1999) found that children with ASD produced significantly fewer original responses on the TTCT Figure Completion Tasks (Repeated Figures condition: *M* = 20.95, *SD* = 22.2 for ASD; *M* = 52.3, *SD* = 14.0 for control; Incomplete Figures condition: *M* = 25.75, *SD* = 9.01 for ASD; *M* = 33.3, *SD* = 2.5 for control). The reported score was a composite of originality, fluency and flexibility scores, thereby conflating these variables. The same pattern was also observed in the TTCT Toy Improvement Task (*M* = 2.6, *SD* = 2.25 for ASD; *M* = 8.4, *SD* = 3.13 for control). However, the total originality score was calculated by summing each original response, therefore confounding originality with fluency of responding. Therefore, whilst the results may indicate lower originality in the ASD group, this may also reflect sub-optimal outcome measures.

The two other studies report fluency independent originality scores. Unlike Craig & Baron-Cohen (1999), Pring, Ryder, Craig and Hermelin (2012) found no group differences on the TTCT Figure Completion Tasks in an adult sample. However, the very small number (N=9) of participants in each group make it difficult to interpret this null finding as the study is insufficiently powered to detect anything

less than a large effect. Dunn, Gomes and Sebastian (2006) assigned prototypicality ratings based on standardised norms to words produced in a verbal fluency task and found that children with ASD produced significantly more original responses on average than typically developing controls (*M* = 4.07, *SD* = 0.56 for ASD; *M* = 3.09, *SD* = 0.81 for control).

In conclusion, taking fluency out of the equation appears to hint at an advantage for ASD on normative originality scores. However, this should be taken with caution due to small samples and inconsistencies in outcome measures, and moreover a lack of studies overall.

5.2.2 Criterion-Referenced

Six studies were found that assessed originality using a set of criteria, again painting a variable picture for ASD performance with as many diverse sets of criteria as there are studies. One study reported an impairment in ASD, whereas two other studies reported no difference between groups and three studies reported superior performance of individuals with ASD relative to controls.

Turner (1999) found that adults with ASD produced a lower proportion of highly imaginative responses than controls in two ideational fluency tasks (Pattern Meanings Task: *M* = 13.5, *SD* = 4.3 for ASD; *M* = 31.2, *SD* = 6.9 for control; Uses of Objects Task: *M* = 16.5, *SD* = 9.9 for ASD; *M* = 38.6, *SD* = 10.4 for control). This was defined as “a highly interpretative response that takes account of all the characteristics of the stimulus in an imaginative, but plausible, fashion” (p. 193) and so is a rather complex criterion incorporating more than originality.

Dichter, Lam, Turner-Brown, Holtzclaw and Bodfish (2009) used the same Uses of Objects task as Turner (1999) but found that children with ASD produced as many unusual responses as controls (*M* = 3.05, *SD* = 3.15 for ASD; *M* = 3.51, *SD* = 3.17 for control). Pring et al., (2012) also observed no group differences between the ASD and MLD group using judges’ ratings of originality in a Figural Synthesis

Task. However, again, care must be taken with interpreting this result due to the very small sample size.

Lui et al., (2011) administered the CAP and found that the ASD group scored significantly higher on this originality scale than the control group (*M* = 15.56, *SD* = 4.65; *M* = 11.97, *SD* = 3.96) and were therefore significantly less likely to be bound or blocked by the stimuli closures and created more synthesis in their designs. Finally, Kasirer and Mashal (2014, 2016) found that both adults and children with ASD in fact produced a higher proportion of novel responses on a metaphor generation task (Adults: *M* = 37.06, *SD* = 31.38 for ASD; *M* = 19.4, *SD* = 11.97 for control; Children: *M* = 34.71, *SD* = 21.21 for ASD; *M* = 18.97, *SD* = 18.18 for control).

Taken overall, these studies would convey somewhat advantageous performance for ASD on criterion-referenced originality scores, on both verbal and non-verbal measures. However, the variable criterions of originality make it difficult to synthesise results more fully and raise concerns of the construct validity of the measures.

5.2.3 Summary

Looking at the results from studies using both norm and criterion-referenced definitions of originality, it appears that the reported deficits of ASD relative to controls stem from two papers only: Craig & Baron-Cohen (1999) and Turner (1999). All other studies conclude that ASD are equivalent to or exceed performance of their non-ASD peers, in verbal, design and ideational fluency tasks. These results are summarised in Table 1.3. It may be therefore prudent not to place undue weighting on these two papers when determining the composite creative profiles of ASD, especially in light of the highlighted methodological issues and lower study quality rating (see Table 1.2).

Table 1.3 Summary of study group comparisons for originality

Study authors, year	Type of fluency task	Sample	Group comparison	
			Norm referenced	Criterion referenced
Dunn, Gomes & Sebastian, 2006	Verbal	Child	ASD > Controls	
Pring et al., 2012	Design	Adult	ASD = Controls	ASD = Controls
Lui, Shih & Ma, 2011	Design	Child		ASD > Controls
Craig & Baron-Cohen, 1999	Design	Child	ASD < Controls	
	Ideational	Child	ASD < Controls	
Turner, 1999	Ideational	Adult		ASD < Controls
Dichter et al., 2009	Ideational	Child		ASD = Controls
Kasirer & Mashal, 2014	Ideational	Adult		ASD > Controls
Kasirer & Mashal, 2016	Ideational	Child		ASD > Controls

5.3 Usefulness

The second component of assessing creative products is usefulness, without which ideas may be original but useless. Only two studies were found that used a usefulness criterion to score responses in a fluency or divergent thinking task.

Channon, Charman, Heap, Crawford and Rios (2001) rated the quality of solutions generated in a social predicaments task and found that the ASD group scored significantly lower than the control group overall on the variables of problem appreciation, social appropriateness, and effectiveness. However, when only the optimal chosen solutions were considered, these group differences became non-significant for total problem appreciation (*M* = 11.67, *SD* = 2.16 for ASD; *M* = 12.67, *SD* = 2.06 for control) and total effectiveness (*M* = 7.67, *SD* = 2.61 for ASD; *M* = 9.33, *SD* = 2.50 for control). Social appropriateness of solutions remained significantly lower in the ASD group (*M* = 6.87, *SD* = 2.07; *M* = 9.13, *SD* = 2.20).

Bishop & Norbury (2005) reported that children with ASD produced a higher frequency of 'not useful' responses (*M* = 6.36, *SD* = 4.27) than control participants (*M* = 3.89, *SD* = 3.25) on a Uses of Objects Task. However, they do not report whether this observation reaches significance.

Taken together these results suggest that young people with ASD may produce more irrelevant responses overall in fluency tasks. However, when asked to select their best idea they choose something effective.

6 Discussion

This review sought to provide a comprehensive quantitative and narrative account of performance of individuals with ASD on fluency and divergent thinking tasks of creativity, relative to matched controls. More specifically, task performance was analysed in relation to: (1) flexibility, measured as how individuals cluster and switch their response sets; (2) originality, defined as either a statistically infrequent response or against specified criteria; (3) and usefulness. Performance on these indices is discussed below in relation to how it can inform our understanding of the creative profile of individuals with ASD.

6.1 Flexibility

The results of the first meta-analysis showed a small to medium impairment in ASD in switching between categories on fluency tasks and generating responses belonging to several different category types. However, the second meta-analysis found no difference between groups in cluster size, that is, the number of responses generated within a certain category. This could account for the impaired fluency performance reported in ASD: the problem does not lie in the generation of ideas within a category, but rather with a difficulty moving on or switching between categories, which could result in fewer ideas produced overall.

The question is then whether impaired switching and intact clustering abilities enables the generation of creative ideas. Divergent thinking posits that the more flexibly ideas are generated, the more creative they become, as they digress from the established paradigm and explore unique avenues. However, a different process is that of exploitative creativity, involving incremental shifts within an existing paradigm to refine ideas until a new concept emerges. Therefore, it may not be necessary to be flexible in order to be creative.

Nijstad et al., (2010) put forward the dual pathway model of creative ideation, suggesting that to produce something original and appropriate requires cognitive flexibility or persistence, or both. These two pathways appear to be independent of each other in DT tasks. Importantly, the authors incorporate the creative domains of person and place, illustrating that different dispositional and contextual factors will influence the processing route taken. They reviewed evidence demonstrating that original ideas were not only the product of cognitive flexibility, operationalised as the number of content categories engaged with, but that originality was also predicted by the exploration of a few categories in great depth (albeit to a lesser extent).

They also suggested that a global processing mode is associated with the cognitive flexibility pathway, which is interesting to consider in light of characteristic features of ASD such as attention to detail and a preference for localised processing style (Happe & Frith, 2006; Koldewyn, Jiang, Weigelt & Kanwisher 2013). According to this dual pathway model, it would appear that although the flexibility pathway is impaired in ASD, the persistence pathway is not, and this would provide a viable route to creative ideas.

There is some evidence that the persistence pathway is used in creativity tasks by individuals with ASD. Lewis & Boucher (1991) analysed the thematic content of drawings produced by children with ASD and children with learning disabilities and found that autistic children drew more pictures than controls which were strongly related to previous pictures. A study by Hampton, Thiébaut, Wu, White and Burgess (in press) also supports this notion of exploitative creativity in ASD. They found that adults with ASD were more likely to systematically cluster their responses on design fluency tasks by producing sets of responses that varied only slightly from each other and by the same method. The authors suggest that the capacity for a detailed-focussed style may positively enhance creativity and be a

distinct process from flexible thought. However, data comparing group performance on other creativity variables was not available to support this conclusion.

There is a potential shortcoming in the explanatory power of the dual pathway account for the results of the current review. The findings of the meta-analysis indicate less flexibility in the ASD group. Therefore, we might have expected to find larger clusters in the ASD group if the persistence pathway was preferentially activated. However, there was no difference found between groups in cluster size.

A possible reason for this null finding is that only a small number of studies were found that fulfilled the criteria of measuring cluster or category size, and all of these studies reported only on measures of verbal (phonemic and semantic) fluency. As design and ideational fluency tasks are arguably more creative tasks than those of phonemic or semantic fluency (Turner, 1999) it is possible that the creative process was not being validly assessed in the examined studies.

Furthermore, Nijstad et al., (2010) noted that original ideas will only be produced via the persistence pathway when participants are given sufficient time to exhaust the conceptual categories or are somehow induced to stay in these categories for longer. This was not the case in the studies included in this review as participants were given one minute per task to generate ideas. These limitations in methodology and reporting potentially bias results and do not allow for a full assessment of the creative process in ASD.

6.2 Originality

Using both norm and criterion-referenced definitions of originality indicated comparable or advantageous performance of ASD in producing original words, designs and ideas in fluency and divergent thinking tasks. However, the results were not unanimous.

Double the number of studies used criterion-referenced methods for measuring creativity compared to norm-referenced methods. Whilst this perhaps speaks to the merits of more subjective rating scales in capturing the richness of the definition of originality (Silvia et al., 2008), it also highlights the inconsistencies in measuring this complex concept (Simonton, 2017). The lack of a consensus definition is a well-known problem in creativity research, limiting the generalisability of the conclusions that can be drawn from the literature and the potential impact and applications of the findings (Plucker et al., 2004).

The studies included in the current review used variable criteria to assess originality. Whilst more standard terms included novelty of response (e.g. Kasirer & Mashal, 2014, 2016) or unusual responses (Dichter et al., 2009), Turner (1999) incorporated both imagination and plausibility in her definition. There is a conceptual overlap between creativity and imagination; however, important distinctions can be made (Runco, 2014). Imagination is the ability to think of unreal or impossible things whereas creativity is the ability to think of original or unusual ideas. In addition, imagination involves pretence whereas creativity is purposeful.

Perhaps a useful analogy is that whilst creativity is the ability to think outside the box, imagination is the ability to think the box a circle. Craig and Baron-Cohen (1999) distinguish between reality-based and imaginative creativity and conclude that children with ASD in fact generate a significantly higher proportion of reality-based or practical ideas than verbal-age matched controls on tests of divergent thinking. Hampton et al., (in press) also found that adults with ASD use significantly less pretence in their responses on an ideational fluency task.

In light of this evidence and the characteristic difficulties with imaginative ability in ASD (Wing & Gould, 1979), it would appear that placing value on imaginative ideas alone could disadvantage individuals with ASD in the measurement of creativity. The criteria established by Turner (1999) to credit responses that showed an “imaginative” element may therefore have biased results

against participants with ASD. Whilst a review of the literature regarding imaginative ability and ASD is beyond the scope of the current review, it is useful here to distinguish between the concepts so as to aid in a clear definition of creativity and the tools that seek to measure it.

6.3 Usefulness

Only two studies were found that measured the usefulness or effectiveness of responses on fluency tasks and therefore there is limited evidence to draw any firm conclusions. It would appear that individuals with ASD produce fewer useful ideas overall on these tasks. However, the results indicate that when given the opportunity to weigh up ideas and select their optimal answer, individuals with ASD are at least as good as TD controls at taking account of pertinent aspects of the scenario and generating a helpful solution.

Why is the usefulness criterion so often neglected in creativity research? Although Bishop & Norbury (2005) demonstrated 78.6% inter-rater agreement for scoring of 'not useful' responses on their ideational fluency task, they note that "disagreements often arose in the Use of Objects task when a response indicated a playful or other creative use, where raters might disagree between coding 'correct' or 'not useful'" (p. 18). This highlights the difficulty in operationalising the concept of utility, and how it is sometimes felt to be juxtaposed to the very notion of creativity, rather than an essential component of it.

An issue may lie with the task instructions, whereby participants are required to 'tell me all the ways in which you think a [object] can be useful'. Without providing further elaboration it may be initially unclear how the participant's response is useful. Furthermore, the participant may be left in doubt as to the purpose of the task: useful for *what* or to *whom*? Without this contextual information it is a difficult task to produce something creative, and perhaps an even greater challenge for participants with ASD who are perhaps less able to infer this information from their

environments without explicit instruction in open-ended test situations (White, Burgess & Hill, 2004).

Zeng, Proctor & Salvendy (2011) highlight that the abstract nature of widely used ideational fluency tasks such as the TTCT do not lend themselves to the usefulness criterion. For example, an item on the Just Suppose task on the TTCT asks children "What if a great fog were to fall over the earth and all we could see of people were their feet?" (as cited in Cooper, 1991; p. 196). Whilst this question may be interesting, it is difficult to see how it relates to everyday creativity or holds ecological validity. Is an individual that cannot imagine an impossible idea incapable of generating a creative solution to an everyday problem?

In the design and use of creativity measures the focus must be on their purpose and relevance, otherwise they potentially disadvantage the participants using them and invalidate the research question being asked of them. Designing tasks that portray everyday problems and allow for the generation of both original and useful solutions, would better enable the measurement of everyday creativity.

6.4 Implications

The heterogeneity of clinical presentations within ASD combined with the heterogeneity of research methods used to measure creativity makes it difficult to draw firm conclusions about the nature of creativity in ASD. However, several methodological issues have been highlighted by this review which should be addressed to further the study.

Firstly, tasks should be designed which resemble everyday examples of creative problems to merit reality-based creativity. Secondly, these tasks should not be timed and should prime for a narrow in-depth focus to tap into the persistence pathway. Thirdly, the different components of creativity should be measured comprehensively, and more outcome measures reported: fluency, flexibility,

clustering, originality and usefulness. Finally, the development of consensual and standardised measures of originality and usefulness is a priority for future research.

These modifications would allow for the comprehensive assessment of the creative process in ASD and thereby inform clinical adaptations and interventions designed to enhance the creative potential of children with ASD in applied settings. For example, an individual with ASD may indicate a preference for the persistence pathway in a creative problem-solving task by demonstrating low rates of switching but large within-category clusters. These conditions could then be mirrored in a classroom setting during creative activities by category priming prior to idea generation; encouraging elaboration upon previous ideas; and allowing time for categories to be fully explored.

This would provide a more structured and person-centred creative environment as opposed to more traditional methods of ‘brainstorming’, which prize rapid generation of random ideas over a more systematic and gradual search process. In addition, the individual could be guided to develop strategies that increase cognitive flexibility, such as practicing remote association type tasks, and thereby tap into the alternative creative pathway, which may allow more effective routes to original and useful ideas.

6.5 Conclusions

The findings of this meta-analytic and narrative review indicate a unique profile of creativity in ASD, demonstrating less flexible and exploratory strategies but unimpaired generation of ideas within-category. Further research into the creative process is required to determine whether individuals with ASD benefit from more exploitative and persistence strategies in creativity tasks. However, this lack of flexibility does not appear to translate into less creative ideas. Individuals with ASD produce original responses on fluency tasks in both verbal and visuo-spatial domains at a level matching or exceeding controls.

However, the limited numbers of studies available on which to base these conclusions, as well as the methodological and reporting issues highlighted, limit the generalisability of these findings. Furthermore, the different components of creativity are yet to be measured comprehensively in ASD and development of a standardised creativity instrument would allow synthesis and collaboration across studies. These areas should be prioritised in future research.

7 References

Amabile, T. M. (1982). A consensual assessment technique. *Journal of personality and social psychology*, *43*, 997-1013.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., revised). Washington, DC: The Association.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). Arlington, VA: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Association.

Baird, G., Simonoff, E., Pickles, A., Chandler, S., Loucas, T., Meldrum, D., & Charman, T. (2006). Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the Special Needs and Autism Project (SNAP). *The Lancet*, *368*(9531), 210-215.

Begeer, S., Wierda, M., Scheeren, A. M., Teunisse, J. P., Koot, H. M., & Geurts, H. M. (2014). Verbal fluency in children with autism spectrum disorders: Clustering and switching strategies. *Autism*, *18*(8), 1014-1018.

Benton, A. L. (1968). Differential behavioral effects in frontal lobe disease. *Neuropsychologia*, *6*(1), 53-60.

Best, C., Arora, S., Porter, F., & Doherty, M. (2015). The relationship between subthreshold autistic traits, ambiguous figure perception and divergent thinking. *Journal of autism and developmental disorders*, *45*(12), 4064-4073.

Bishop, D. V., & Norbury, C. F. (2005). Executive functions in children with communication impairments, in relation to autistic symptomatology: I: Generativity. *Autism*, *9*(1), 7-27.

Boden, M. A. (2004). *The creative mind: Myths and mechanisms* (2nd Ed.). London: Routledge.

Bown, M. J., & Sutton, A. J. (2010). Quality control in systematic reviews and meta-analyses. *European Journal of Vascular and Endovascular Surgery*, *40*(5), 669-677.

Carmo, J. C., Duarte, E., Pinho, S., Marques, J. F., & Filipe, C. N. (2015). Verbal fluency as a function of time in autism spectrum disorder: An impairment of initiation processes?. *Journal of clinical and experimental neuropsychology*, *37*(7), 710-721.

Channon, S., Charman, T., Heap, J., Crawford, S., & Rios, P. (2001). Real-life-type problem-solving in Asperger's syndrome. *Journal of Autism and Developmental Disorders*, *31*(5), 461-469.

Cohen J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* New York, NY: Routledge Academic.

Cooper, E. (1991). A critique of six measures for assessing creativity. *The Journal of Creative Behavior, 25*(3), 194-204.

Craig, J., & Baron-Cohen, S. (1999). Creativity and imagination in autism and Asperger syndrome. *Journal of Autism and Developmental Disorders*, *29*(4), 319-326.

Cramond, B., Matthews-Morgan, J., Bandalos, D., & Zuo, L. (2005). A report on the 40-year follow-up of the Torrance Tests of Creative Thinking: Alive and well in the new millennium. *Gifted Child Quarterly*, *49*(4), 283-291.

Cropley, A. J. (2000). Defining and measuring creativity: Are creativity tests worth using?. *Roeper review*, *23*(2), 72-79.

Cropley, D. H., Kaufman, J. C., & Cropley, A. J. (2008). Malevolent creativity: A functional model of creativity in terrorism and crime. Creativity Research Journal, 20, 105–115.

Demetriou, E. A., Lampit, A., Quintana, D. S., Naismith, S. L., Song, Y. J. C., Pye, J. E., ... & Guastella, A. J. (2017). Autism spectrum disorders: a meta-analysis of executive function. *Molecular psychiatry*.

Dichter, G. S., Lam, K. S., Turner-Brown, L. M., Holtzclaw, T. N., & Bodfish, J. W. (2009). Generativity abilities predict communication deficits but not repetitive behaviors in autism spectrum disorders. *Journal of autism and developmental disorders*, *39*(9), 1298.

Dietrich, A. (2004). The cognitive neuroscience of creativity. *Psychonomic bulletin & review*, *11*(6), 1011-1026.

Dixon, J. (1979). Quality versus quantity: The need to control for the fluency factor in originality scores from the Torrance Tests. *Journal for the Education of the Gifted, 2*, 70–79.

Duncan, J. (1986). Disorganisation of behaviour after frontal lobe damage. *Cognitive Neuropsychology*, 3, 271–290.

Dunn, M., Gomes, H., & Sebastian, M. J. (1996). Prototypicality of responses of autistic, language disordered, and normal children in a word fluency task. *Child Neuropsychology*, *2*(2), 99-108.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629-634.

Finke, R. A., & Slayton, K. (1988). Explorations of creative visual synthesis in mental imagery. *Memory & Cognition*, *16*(3), 252-257.

Fitzgerald, M. (2004). *Autism and creativity: Is there a link between autism in men and exceptional ability?* Hove, UK: Routledge.

Flaherty, A. W. (2005). Frontotemporal and dopaminergic control of idea generation and creative drive. *Journal of Comparative Neurology, 493*, 147–153.

Guilford, J. P. (1956). The structure of intellect. *Psychological Bulletin, 53*(4), 267-293.

Guilford, J. P. (1967). *The nature of human intelligence.* New York: McGraw-Hill.

Hampton, S. C., Thiébaut, F. I., Wu, H. C., White, S. J. and Burgess, P. W. (in press). Unusual strengths in creativity in autism.

Happé, F., & Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *36*(1), 5-25.

Heausler, N. L., & Thompson, B. (1988). Structure of the Torrance Tests of Creative Thinking. *Educational and Psychological Measurement, 48*, 463-468.

Hedges L. V., Olkin I. (1985). *Statistical methods for meta-analysis.* San Diego, CA: Academic Press.

Higgins, J. P., & Green, S. (Eds.) (2011). *Cochrane Handbook for Systematic Reviews of Interventions* (Version 5.1.0). Retrieved from http://cochrane-handbook.org.

Higgins, J., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in medicine*, *21*(11), 1539-1558.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal*, *327*(7414), 557.

Hill, E. L. (2004). Evaluating the theory of executive dysfunction in autism. *Developmental Review, 24*(2), 189-233.

Hocevar, D. (1981). Measurement of creativity: Review and critique*. Journal of Personality assessment, 45(*5), 450-464.

Inokuchi, E., & Kamio, Y. (2013). Qualitative analyses of verbal fluency in adolescents and young adults with high-functioning autism spectrum disorder. *Research in Autism Spectrum Disorders*, *7*(11), 1403-1410.

Jensen, C. M., Martens, C. S., Nikolajsen, N. D., Gregersen, T. S., Marx, N. H., Frederiksen, M. G., & Hansen, M. S. (2015). What do the general population know, believe and feel about individuals with autism and schizophrenia: Results from a comparative survey in Denmark. *Autism, 20*(4), 496-508.

Jones-Gotman, M., & Milner, B. (1977). Design fluency: The invention of nonsense drawings after focal cortical lesions. *Neuropsychologia*, *15*(4-5), 653-674.

Kasirer, A., & Mashal, N. (2014). Verbal creativity in autism: Comprehension and generation of metaphoric language in high-functioning autism spectrum disorder and typical development. *Frontiers in human neuroscience*, *8*, 615.

Kasirer, A., & Mashal, N. (2016). Comprehension and generation of metaphors by children with autism spectrum disorder. *Research in Autism Spectrum Disorders*, *32*, 53-63.

Katrak, P., Bialocerkowski, A. E., Massy-Westropp, N., Kumar, V. S., & Grimmer, K. A. (2004). A systematic review of the content of critical appraisal tools. *BMC medical research methodology, 4*(1), 22.

Kim, K. H. (2006). Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking (TTCT). *Creativity research journal*, *18*(1), 3-14.

Kirchner, J., Ruch, W., & Dziobek, I. (2016). Brief Report: Character Strengths in Adults with Autism Spectrum Disorder Without Intellectual Impairment. *Journal of Autism and Developmental Disorders, 46*(10), 3330-3337.

Kmet, L. M., Lee, R. C., & Cook, L. S. (February, 2004). *Standard Quality Assessment Criteria for Evaluating Primary Research Papers from a Varity of Fields.* Retrieved from https://www.ihe.ca/advanced-search/standard-quality-assessment-criteria-for-evaluating-primary-research-papers-from-a-variety-of-fields

Koldewyn, K., Jiang, Y. V., Weigelt, S., & Kanwisher, N. (2013). Global/local processing in autism: Not a disability, but a disinclination. *Journal of Autism and Developmental Disorders*, *43*(10), 2329-2340.

Lai, C. L. E., Lau, Z., Lui, S. S., Lok, E., Tam, V., Chan, Q., ... & Cheung, E. F. (2017). Meta-analysis of neuropsychological measures of executive functioning in children and adolescents with high-functioning autism spectrum disorder. *Autism Research*, *10*(5), 911-939.

Lewis, V., & Boucher, J. (1991). Skill, content and generative strategies in autistic children's drawings. *British Journal of Developmental Psychology*, *9*(3), 393-416.

Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment* (4th ed.). Oxford: Oxford University Press.

Liu, M. J., Shih, W. L., & Ma, L. Y. (2011). Are children with Asperger syndrome creative in divergent thinking and feeling? A brief report. *Research in Autism Spectrum Disorders*, *5*(1), 294-298.

Mayer, R. E. (1999). Fifty Years of Creativity Research. In R. J. Sternberg (Ed.), *Handbook of Creativity* [Online] (pp. 449-460). Cambridge: Cambridge University Press. Retrieved from: Cambridge Books Online http://dx.doi.org/10.1017/CBO9780511807916.024

Mednick, S. (1962). The associative basis of the creative process. *Psychological review*, *69*(3), 220.

Milgram, R. M., & Rabkin, L. (1980). Developmental test of Mednick's associative hierarchies of original thinking. *Developmental Psychology, 16*(2), 157-158.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current directions in psychological science*, *21*(1), 8-14.

Nijstad, B. A., De Dreu, C. K., Rietzschel, E. F., & Baas, M. (2010). The dual pathway to creativity model: Creative ideation as a function of flexibility and persistence. *European Review of Social Psychology*, *21*(1), 34-77.

Nijstad, B. A., Stroebe, W., & Lodewijkx, H. F. (2002). Cognitive stimulation and interference in groups: Exposure effects in an idea generation task. *Journal of experimental social psychology*, *38*(6), 535-544.

Nijstad, B. A., & Stroebe, W. (2006). How the group affects the mind: A cognitive model of idea generation in groups. *Personality and social psychology review*, *10*(3), 186-213.

Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). Executive function deficits in high-functioning autistic individuals: relationship to theory of mind. *Journal of child Psychology and Psychiatry*, *32*(7), 1081-1105.

Pastor-Cerezuela, G., Fernández-Andrés, M. I., Feo-Álvarez, M., & González-Sala, F. (2016). Semantic verbal fluency in children with and without autism spectrum disorder: Relationship with chronological age and IQ. *Frontiers in psychology*, *7*, 921.

Pellicano, E. (2010). Individual differences in executive function and central coherence predict developmental changes in theory of mind in autism. *Developmental psychology*, *46*(2), 530.

Pellicano, E. (2013). Testing the predictive power of cognitive atypicalities in autistic children: Evidence from a 3-year follow-up study. *Autism Research*, *6*(4), 258-267.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of clinical epidemiology*, *61*(10), 991-996.

Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational psychologist*, *39*(2), 83-96.

Pring, L., Ryder, N., Crane, L., & Hermelin, B. (2012). Creativity in savant artists with autism. *Autism*, *16*(1), 45-57.

Pullinger, J. (2017). *A meta-analytic review of generativity in autistic spectrum disorders.* Doctoral thesis (D.Clin.Psy), University College London.Rhodes, M. (1961). An analysis of creativity. *The Phi Delta Kappan*, *42*(7), 305-310.

The Cochrane Collaboration (2014). *Review Manager (RevMan)* [Computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre.

Runco, M. A. (1985). Reliability and convergent validity of ideational flexibility as a function of academic achievement. *Perceptual and Motor Skills*, *61*(3_suppl), 1075-1081.

Runco, M. A. (1988). Creativity research: Originality, utility, and integration. *Creativity Research Journal, 1*, 1-7.

Runco, M. A. (2014). *Creativity: Theories and Themes: Research, Development, and Practice* (Second Edition). San Diego: Academic Press. Retrieved from https://doi.org/10.1016/B978-0-12-410512-6.00001-1.

Runco, M. A., Illies, J. J., & Eisenman, R. (2005). Creativity, originality, and appropriateness: What do explicit instructions tell us about their relationships?. *The Journal of Creative Behavior*, *39*(2), 137-148.

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal*, *24*(1), 92-96.

Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., … & Richard, C. A. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts*, *2*(2), 68.

Simonoff, E., Pickles, A., Charman, T., Chandler, S., Loucas, T., & Baird, G. (2008). Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*(8), 921-929.

Simonton, D. K. (1997). Creative productivity: A predictive and explanatory model of career trajectories and landmarks. *Psychological Review*, *104*(1), 66.

Simonton, D. K. (2003). Scientific creativity as constrained stochastic behavior: the integration of product, person, and process perspectives. *Psychological bulletin*, *129*(4), 475.

Simonton, D. K. (2017) Big-C versus little-c creativity: Definitions, implications, and inherent educational contradictions. In A. R. Beghetto, & B. Sriraman (Eds.),

Creative Contradictions in Education (pp. 3-19). Switzerland: Springer International Publishing.

Spek, A., Schatorjé, T., Scholte, E., & van Berckelaer-Onnes, I. (2009). Verbal fluency in adults with high functioning autism or Asperger syndrome. *Neuropsychologia*, *47*(3), 652-656.

StataCorp (2015). *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP.

Sternberg, R. J., & Lubart, T. I. (1996). Investing in creativity. *American psychologist*, *51*(7), 677.

Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., ... & Tetzlaff, J. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, *343*, d4002.

The Cochrane Collaboration (2014). *Review Manager (RevMan)* [Computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre.

Torrance, E. P. (1966). *The Torrance Tests of Creative Thinking-Norms-Technical Manual Research Edition-Verbal Tests, Forms A and B-Figural Tests, Forms A and B*. Princeton, NJ: Personnel Press.

Torrance, E. P. (1974). *The Torrance Tests of Creative Thinking-Norms-Technical Manual Research Edition-Verbal Tests, Forms A and B- Figural Tests, Forms A and B.* Princeton, NJ: Personnel Press.

Torrance, E. P. (1981). Predicting the Creativity of Elementary School Children (1958-80)—and the Teacher Who" Made a Difference". *Gifted Child Quarterly*, *25*(2), 55-62.

Treffert, D. A. (2014). Savant Syndrome: Realities, Myths and Misconceptions. *Journal of Autism and Developmental Disorders, 44*(3), 564-571.

Troyer, A.K., Moscovitch, M. and Winocur, G. (1997). Clustering and switching as two components of verbal fluency: evidence from younger and older healthy adults. *Neuropsychology, 11,* 138–146.

Turner, M. (1997). Towards an executive dysfunction account of repetitive behaviour in autism. In J. Russell (Ed.), *Autism as an executive disorder* (pp. 57-100). New York, NY, US: Oxford University Press.

Turner, M. A. (1999). Generating novel ideas: Fluency performance in high-functioning and learning disabled individuals with autism. *Journal of Child Psychology and Psychiatry, 40*(2), 189-201.

Wallach, M. A. (1976). Tests tell us little about talent: Although measures of academic skills are widely used to determine access to contested educational opportunities, especially in their upper ranges they lack utility for predicting professional achievement. *American Scientist*, *64*(1), 57-63.

Wallach, M. A., & Kogan, N. (1965). *Modes of thinking in young children: A study of the creativity-intelligence distinction.* New York: Holt, Rinehart, & Winston.

Wallach, M. A., & Wing, C. W., Jr. (1969). *The talented student: A validation of the creativity intelligence distinction.* New York: Holt, Rinehart & Winston.

Weisberg, R. W. (2006). Modes of expertise in creative thinking: Evidence from case studies. In K. A. Ericsson, N. Charness, P. J Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance*, (pp. 761-787). New York: Cambridge University Press.

White, S. J., Burgess, P. W., & Hill, E. L. (2009). Impairments on “open-ended” executive function tests in autism. *Autism Research*, *2*(3), 138-147.

Williams, F. E. (1980). *Creativity assessment packet: Manual.* Buffalo, NY: DOK Publication.

Wing, L., & Gould, J. (1979). Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification. *Journal of autism and developmental disorders*, *9*(1), 11-29.

World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines* (Vol. 1). World Health Organization.

Wykes, T., Huddy, V., Cellard, C., McGurk, S. R., & Czobor, P. (2011). A meta-analysis of cognitive remediation for schizophrenia: methodology and effect sizes. *American Journal of Psychiatry*, *168*(5), 472-485.

Zeng, L., Proctor, R. W., & Salvendy, G. (2011). Can traditional divergent thinking tests be trusted in measuring and predicting real-world creativity?. *Creativity Research Journal*, *23*(1), 24-37.

Part 2: Empirical Paper

Developing an Ecologically Valid Measure of Creativity for Children with Autism Spectrum Disorders

1 Abstract

Aims: The study of creativity in Autism Spectrum Disorders (ASD) is limited by a lack of appropriate tools and the methodological issues inherent in divergent thinking (DT) tasks. Not least of these is the criterion problem, evidenced by poor correlation between performance on DT tasks and everyday creative activities and achievements. This study aimed to develop and pilot ecologically valid measures of creativity for children with ASD.

Method: Three tasks were designed to resemble everyday creative activities and scenarios: Toy Improvement; Story Scenes; and Situations and Solutions tasks. In addition, each task contained several items which varied the contextual demands. The task battery was administered to 15 participants with ASD and 15 typically developing controls aged between 8 and 14 years with normal intellectual functioning (IQ > 70). The psychometric properties of the tasks were investigated alongside a preliminary investigation of group performance.

Results: Interrater reliability was in the fair to good range (ICC range from .57 to .69). Test-retest reliability was low (r = .056 to .738). However, correspondence with other measures of creativity showed more promise; in particular, ratings given by experts in the Story Scenes task (r = .941). Performance on several variables of the Story Scenes and Situations and Solutions task was significantly related to IQ. Group differences were found only in certain conditions of each task. These differences mainly related to the quality of the creative product, with the control group producing more highly original, useful and creative ideas. However, this was not a uniform picture. No group differences were found in other variables relating to the creative process and the tendency was for the ASD group to be better at self-selecting their most creative ideas for evaluation.

Conclusion: The creativity tasks had variable psychometric properties and there is scope for improvement. The Story Scenes and Situations and Solutions tasks

showed some promise in relation to criterion, convergent and ecological validities and were popular with participants. Performance on the creativity tasks varied by condition, suggesting that performance on traditional DT tasks may be related to design rather than a domain-general creativity deficit in ASD. However, further replication is required in a sample with matched IQ.

2 Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition characterized by deficits in social communication and interaction across contexts; as well as restricted, repetitive patterns of behaviour, interests and activities (Diagnostic and Statistical Manual 5th ed.; DSM-V; American Psychiatric Association, 2013). Prior to these dyadic diagnostic criteria, Wing (1981) defined a triad of impairments in ASD in social behaviour, communication and imagination, linking repetitive or stereotyped behaviours to a lack of imaginative activity (Wing & Gould, 1979). This link may be mediated by an impaired capacity for generativity (Turner, 1999).

The ability to generate novel ideas is seen as a prerequisite for creativity. Craig and Baron-Cohen (1999) refer to "a lack of normal creativity" (p. 5) in children with ASD; however, there is contradictory evidence that high autistic traits are associated with superior creative ability in certain eminent individuals (Fitzgerald, 2004). A possible explanation for this discrepancy is that they are considering distinct types of creativity: 'Big C' versus 'Little C' (Amabile, 2014). These are discrete areas of study in the creativity field and research efforts are divided along these lines.

'Big C' creativity refers to notable and distinguished individuals who revolutionize their field with their creative developments, such as those studied by Fitzgerald (2004). This research is based on biographical accounts and historiometric approaches, drawing on historical documentation.

'Little C' creativity, on the other hand, refers to everyday creative problem solving and ordinary creative activities, as investigated by Craig and Baron-Cohen (1999). This enquiry relies on the psychometric approach and is interested in measuring creative abilities in the present, understanding creative processes and the characteristics of creative products (Plucker & Renzulli, 1998). As 'Little C' measures are not based on indisputable accounts of creative eminence, they are

harder to quantify. Research examining the association between everyday 'Little C' creativity and autism is limited (see Part 1 of this Thesis) and the nature of creativity in ASD warrants further investigation.

2.1 Definition of Creativity

The study of creativity necessitates a definition of the concept, and criteria against which it can be measured. The consensus view is that for something to be creative, it must be both original and useful (e.g. Barron, 1955; Mayer, 1999; Stein, 1953). Runco and Jaeger (2012) describe this as the standard definition and attest that without the dual component, ideas and products may be novel and interesting but essentially useless.

The criteria of originality and utility can only be applied to the creative product, which can be seen as the outcome of a dynamic interaction between the creative components of the person, cognitive processes and the context. Although there are multiple facets of creativity, all factors culminate in a creative product, and therefore any tool attempting to measure creativity should be designed to assess creative output in relation to the criteria of originality and usefulness (Plucker & Renzulli, 1998).

2.2 Creativity Within a Social Context

Creativity can be seen as a form of social action in that creative value is determined by audience and context (Weisberg, 1986) and insofar as creativity can be defined as "appropriate novelty that is recognized as such by people knowledgeable in a domain" (Amabile, 2014; p. 9). This social valuation is what separates ideas that are simply unusual from those that are creative. Creative products can therefore not be measured devoid of their social context.

In the evaluation of creative products in everyday practice, the ratings of external judges or experts are frequently used. This method is replicated in research settings by using the Consensual Assessment Technique (CAT; Amabile, 1982) and

this has been called the “gold standard” of creativity assessment (Baer & McKool, 2009). These expert ratings do not rely on external objective standards but rather implicit theories of creativity, and methods such as CAT can therefore be used to evaluate the social validity of psychometric assessments (Runco & Behleda, 1986).

The social influences upon creativity cannot be considered independently of the socio-communicative impairments often characteristic of ASD. Mentalising is the ability to view oneself from the outside, and others from the inside, and has been found to be deficit in children with ASD (White, Hill, Happé, & Frith, 2009). To produce something creative, that is both original and useful in a given domain, requires an understanding of the social context and the ability of the creator to consider for *what purpose* or *to whom* a product may be valuable.

It is possible that susceptibility to the intrinsic social demands of creative contexts could differentiate typically developing children and children with ASD. For example, White, Burgess and Hill (2009) suggest that poor understanding of the implicit demands made by researchers in open-ended task situations can affect task performance in children with ASD, highlighting the importance of task instructions and the social context in which testing takes place. This is perhaps particularly salient in a creative context, where both priming and explicit instruction can influence the quantity and quality of creative ideas (Rietzschel, Nijstad & Stroebe, 2007; Silvia et al., 2008).

2.3 Creativity Measurement and ASD

Tests of divergent thinking (DT) are the most widely used psychometric tool in the measurement of creative potential. DT is the generation of multiple ideas in response to an open-ended problem and is thought to be involved in the ideation phase of the creative process. Individual differences in DT tests are found in the areas of fluency (the number of ideas); originality (the number of unusual or unique ideas); and flexibility (the number of different categories endorsed by the ideas).

These tasks have been used to measure creativity in children with ASD and have found poorer ideational fluency relative to controls (Craig and Baron-Cohen, 1999). However, other studies have found that whilst individuals with ASD produce fewer responses overall in DT tasks, a higher proportion of these ideas are original relative to controls (Best, Arora, Porter & Doherty, 2015; Liu, Shih & Ma, 2011). In a meta-analytic review, individuals with ASD showed a small to moderate impairment in switching between categories in DT type tasks (see Part 1 Literature Review). However, it remains unclear as to whether this corresponds to fewer creative ideas or, instead, points to a unique profile of creativity in ASD requiring further research.

2.4 Limitations of Divergent Thinking Tasks

The evidence does not appear to support the concept of impaired creativity in ASD. Rather, the literature illustrates that there has not yet been an adequate test of creativity developed for this population and points to directions for future research in this area.

One issue would appear to lie with the scoring of DT tasks, which conflate fluency and originality (scored as statistical rarity of responses). Silvia et al., (2008) instead highlight the merits of using subjective ratings of creativity and asking participants to self-select their top creative responses. The authors found that both Average scoring and Top 2 scoring methods produced dependability levels of .80 with two or three raters, whereas the Top 2 scores stood out in their relationship with associated creative personality variables, suggesting the construct validity of this method.

Furthermore, although individuals with ASD may demonstrate less flexible switching between ideational categories in DT tasks, they appear unimpaired in generating ideas within a category (see Part 1 Literature Review). The dual pathway model of creativity (Nijstad, De Dreu, Rietzschel, & Baas, 2010) proposes that different strategies may be used in the generation of creative ideas: flexibly

switching between categories to broaden perspective and generate remote associations (flexibility pathway); or clustering responses to narrow the focus and incrementally progress towards an original idea (persistence pathway). According to this dual pathway model, it would appear that although the flexibility pathway is impaired in ASD, the persistence pathway is not. This could provide a viable route to creative ideas. Nijstad, Stroebe & Lodewijkx (2002) recommend that within-category fluency should also be measured in DT tasks to capture this important component of the creative process.

Cropley (2000) highlighted an important advancement in creativity testing in the recognition that real-life creative production involves both generation of ideas (divergent thinking) *and* narrowing down focus to select the best ideas to carry forward (convergent thinking). Zeng, Proctor & Salvendy (2011) outline a general model of the creativity process: problem analysis, ideation, evaluation, and implementation. They highlight that existing measures focus almost exclusively on ideational productivity rather than the other three areas of creative thinking. As a result, they suggest that DT tasks be structured around situational problem contexts that require goal-oriented creative thinking and selection of original and valuable ideas (products).

The psychometric properties of DT tests also pose further limitations. DT tests lack construct validity as they fail to operationalize the standard definition of creativity. The criterion of usefulness is neglected, possibly due to the abstract, unrealistic nature of the tasks (Zeng et al., 2011). DT tests have been shown to have inconsistent convergent validity with other measures of creativity (Hocevar, 1981) and poor ecological and discriminant validities (Zeng et al, 2011). In a large scale longitudinal study the TTCT was a moderate predictor of later life self-reported creative achievements (Cramond et al., 2005; Torrance, 1981), although this finding has not been replicated to the same extent in other studies (Stenberg & Lubart, 1996).

Earlier research questioned the administration and testing conditions of DT tests, suggesting creativity may be suppressed in traditional testing environments (Wallach & Kogan, 1965) and emphasising the importance of predicting real world creative potential rather than tested ability (Wallach, 1976; Wallach & Wing, 1969). Experience and expertise within a particular domain is closely tied to creativity (Weisberg, 2006) and as DT tasks lack domain specificity they are unlikely to be accessing an individual's full creative potential. Overall, these limitations suggest several areas for improvement of DT tasks in design, administration and scoring.

2.5 Development of an Ecologically Valid Measure of Creativity

The aforementioned limitations of DT tests have led researchers to suggest that DT scores should not be equated with real-life creativity, and to increase ecological validity tasks should be designed that resemble real-world problems in a particular domain (Okuda, Runco & Berger, 1991). Mayer (1999) suggested that in the field of creativity research, "what is needed is a methodology that combines the scientific respectability of the psychometric and experimental approaches with the authenticity of the biographical approach" (p. 459). Focusing on the development of ecologically valid measures could be one solution.

The use of ecologically valid measurement tools is increasingly prioritised in ASD research because if a tool is better able to capture strengths and difficulties experienced in real life then it is more useful for guiding clinical interventions (Burgess et al, 2006; Kenworthy, Yerys, Anthony & Wallace, 2008). For example, Diener, Wright, Smith & Wright (2014) developed a scoring criterion to assess 3D design projects in young people with ASD enrolled on a design technology programme. This scoring method was validated by a team of Google experts. The authors suggest this assessment process can be used not only as a measure of visual-spatial creativity in youths with ASD but by potential employers looking to determine real-world creative potential.

Furthermore, designing a task for a specific population ensures the test is optimal for purpose by using appropriate methodology and giving particular consideration to participant characteristics and factors that could influence test performance and otherwise invalidate test results. Recent research efforts have focused on the development of an ecologically valid test battery to capture executive functioning deficits that present in children with ASD (Bristow, 2016; Ledger-Hardy, 2017; Pullinger, 2017). These studies took a function-led approach to task-development, starting from the bottom-up as opposed to adopting a more theoretical top-down perspective. The same strategy could be particularly helpful in designing an ecologically valid measure of creativity, especially in light of the inherent difficulties of operationalising such a complex construct (Zeng et al., 2011).

2.6 Aims of Current Study

In summary, the study of creativity in ASD thus far has failed to demonstrate a specific or universal deficit and the pattern of strengths and weaknesses in the creative ASD profile is unclear. Traditional tests of divergent thinking have clear limitations as measures of creativity, not least in their methodological issues and lack of construct and ecological validity.

Furthermore, DT tests do not account for the fact that creativity cannot be measured independently of the context in which it takes place and these social demands are particularly important to consider in light of the socio-communicative impairments characteristic of ASD. Overcoming the methodological limitations of existing creativity measures and reducing the implicit demands or mentalising component of the test situation by providing explicit instructions and contextual cues could facilitate the creative output of children with ASD.

The primary aims of the current study are therefore to develop and pilot a new measure of creativity that:

- is specifically developed for school-aged children with ASD;
- has ecologically validity, resembling real-life creative tasks and environments;
- varies the task demands to mirror real-life social contexts and situations;
- specifies a scoring guideline for originality and usefulness against which to measure the creative products;
- has test-retest reliability;
- corresponds to real-life creative activities and pursuits as a measure of criterion validity;
- demonstrates construct validity by corresponding with expert ratings of creative products and parental ratings of creative characteristics; and
- demonstrates discriminant validity against measures of executive functioning and IQ.

This study extends an ongoing research project, developing an ecologically valid test battery to assess executive functioning in children with ASD (Bristow, 2016; Ledger-Hardy 2017; Pullinger, 2017). Previous trainees have already developed and piloted several novel measures that comprise the Ecologically-Valid Test of Executive Dysfunction (Eco-TED). Initial analysis has shown promising results for a selection of tests and identified recommendations for the next phase of development. A secondary aim of the current investigation is to modify and carry forward several of the existing measures in light of these results and continue data collection on the Eco-TED, alongside the development of new creativity tasks to form part of the test battery.

3 Method

3.1 Participants

In total 30 participants were recruited to take part in this study. Of these, 22 had taken part in the Eco-TED research previously.

The ASD group was recruited from a specialist social and communication disorders clinic and comprised of 15 children between the ages of 8 and 12 years at the initial time of recruitment (M = 146.67 months; SD = 26.98). All had received a recent clinical consensus diagnosis of ASD given on the basis of information gathered using the Developmental, Dimensional and Diagnostic Interview (3di; Skuse et al., 2004), the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 1989) and children's school reports. There were 10 males and 5 females in the clinical group. Seven participants had comorbid diagnoses of either ADD/ADHD (n=4); dyspraxia and specific language impairment (n=1); OCD (n=1); or anxiety (n=1).

The control group consisted of 15 participants aged between 8 and 12 years old at the initial time of recruitment (M = 146.93 months; SD = 21.70). Eight were recruited through a mainstream school and the others through convenience and snowball sampling. There were 12 males and 3 females in the control group. Participants with a diagnosis of a neurodevelopmental disorder (such as ADHD) were excluded from the study. Three of the participants had a sibling with ASD.

For both the clinical and control group, only children with a full-scale IQ in the normal range (FSIQ > 70) were included in the study. As is shown in Table 2.1, the typically developing participants had, on average, higher FSIQ scores compared to participants with ASD. In addition, all children were fluent in English. See Table 2.1 for a summary of participant characteristics.

Table 2.1. Participant characteristics for the ASD and TD groups.

	ASD (n=15) Mean (SD)	**TD (n=15) Mean (SD)**	***t***	***p***
Age in months	146.67 (26.98)	146.93 (21.70)	0.03	.637
FSIQ	101.13 (11.48)	111.87 (10.79)	2.64	.013*
Gender (*m:f*)	10:5	12:3		
Ethnicity (*n)*				
White British	11	13		
White Other	2	2		
Mixed Ethnicity	2	0		
Clinical Diagnosis (*n)*				
ASD[a]	7			
HFA[b]	2			
Asperger Syndrome	6			

* $p < .05$
[a]Autism Spectrum Disorder
[b] High Functioning Autism

3.2 Task Development

The primary aim of this study was the development of new creativity tasks to add to the existing Eco-TED battery. This was an extended and iterative process between the researcher and supervising members of the research team, Dr Will Mandy and Professor Paul Burgess, both of whom have experience in developing neuropsychological measures.

Following initial group consultation and subsequent observation and generation of ideas, the rudimentary scripts and materials for several tasks were designed and informally piloted with five typically developing children. These tasks were based on either novel ideas or adaptation of traditional DT tasks to overcome existing limitations and tailor for use specifically with a young ASD population.

During design and development the primary consideration was to create tasks that resembled everyday situations and used familiar scenarios and materials. Several tasks were then abandoned or refined based on feasibility and clinical utility. This process spanned a period of ten months and culminated in three tasks, which are subsequently described in detail.

In addition, a task script was developed (see Appendix B). This was modelled on existing neuropsychological measures in the language and level of detail in instructions for administration and was in keeping of the style of the existing Eco-TED battery. The task script also underwent a series of revision (five in total) following review and feedback from the research team.

3.3 Measures

3.3.1 Creativity Tasks.

3.3.1.1 Toy Improvement Task.

3.3.1.1.1 Background.

This was based on the task developed by Torrance (1974) in the Torrance Tests of Creative Thinking (TTCT) battery. In the original task, participants are shown a picture of a toy elephant and are asked to generate ideas to make it more fun to play with. This was later adapted by Craig and Baron-Cohen (1999) who used a soft toy elephant in their version of the task to make it easier for children by enabling them to manipulate a 3D stimulus.

There are, however, several limitations of the task in its existing format for assessing creativity in a young ASD population. During informal piloting, several

children commented that they would not play with a toy elephant themselves and therefore the task instructions appeared futile and non-realistic. This also could hypothetically increase the cognitive and metalizing demands of the task for children with ASD by requiring them to infer for whom the elephant is being made more fun if it is not otherwise explicitly stated. Traditionally the task is not scored on the variable of Usefulness (the dual component of creativity); however, in the given format it is difficult to operationalize what a useful response may look like when the instructions are simply to 'make it more fun to play with'. Finally, the task requires participants to generate as many ideas as possible without the evaluative component required in the normal creative process, thereby emphasizing quantity over quality of ideas.

The task was therefore modified to include several real-life scenarios whereby the toy had to be developed for either a particular function (items 2 and 3) or context (items 4 and 5). The aim of these adaptations was to also allow the usefulness of responses to be scored, thereby enabling a creativity composite of originality and usefulness to be derived. The child was required to generate multiple ideas (DT phase) and then select their best ideas, thus encompassing multiple phases of the creative process: analysis, ideation, and evaluation.

3.3.1.1.2 Administration.

Participants were given a soft toy elephant and asked to improve it for various conditions. At the start of task, participants were presented with the soft toy elephant and allowed to handle the object as they like. An orientation trial was completed to ensure participants were familiar with the object and able to identify and describe salient features. The main task then proceeded, consisting of five items.

The first item (Condition 1) was a repetition of the traditional task instruction, simply to "tell me as many ways you can think of that would make this toy elephant more fun to play with. It can be anything you like, just tell me all the ways to make it

better". Condition 2 (Items 2 and 3) then required modification of the toy elephant for a specific function: to help babies to learn how to walk; and to help teachers in school. Condition 3 (Items 4 and 5) required adaptation for a broader context: for use in the garden; and for use in hospital.

No time limit was imposed, and participants were given plenty of encouragement for all ideas. The task was discontinued when the participant indicated they had no further ideas or had been silent for 30 seconds following a prompt. The researcher recorded responses to allow the child to focus on the task and play with the elephant. No further elaboration was sought by the researcher unless an item was misheard. Following three consecutive incorrect responses indicating the child had misunderstood the instructions, the researcher reminded the child of the instructions. At the end of each item, the researcher asked the participant to choose their two best ideas.

3.3.1.1.3 Scoring Criteria.

Please refer to Appendix C for a detailed description of the scoring criteria. An outline of the variables used in this task is provided in Table 2.2.

The following variables were calculated: Fluency (the total number of responses, including repetitions but excluding errors); Originality (defined as an unusual, surprising and interesting response); Usefulness (defined as an appropriate, feasible and functional response); Flexibility (the number of thematic categories encompassed by the response set); Within-Category Fluency (the mean number of ideas within each category).

A Creativity score was calculated by multiplying the Originality and Usefulness scores for each response. This was done rather than taking an average score to differentiate responses that were both original *and* useful as opposed to high scorers in one domain but not the other (Simonton, 2017). So that scores in other domains would not be conflated with Fluency scores, Fluency independent

scores for Originality, Usefulness and Creativity were generated (by dividing scores in each domain by item fluency). These scores are reported throughout.

Of the two responses selected by participants as their best ideas for each item, the highest creativity score was recorded as the Best Chosen Response. To determine whether participants actually chose their highest scoring response as their best response, the Best Chosen Response score was divided by the highest Creativity score for that item to create a Creativity Evaluation score, where a maximum score of 1 indicated selection of the most creative idea as the best idea.

For each criterion, scores were averaged across the items within each condition (item 1 was scored separately). A Total score was also calculated by averaging across all items, except for Total Fluency, which was summed.

Table 2.2. List of the variables used in both the Toy Improvement and Situations and Solutions tasks.

Variable	Description
Fluency	The total number of responses, including repetitions but excluding errors.
Originality	How unusual, surprising and interesting a response is. Scale of 1 – 7. Fluency independent scores were calculated.
Usefulness	How appropriate, feasible and functional a response is. Scale of 1 – 7. Fluency independent scores were calculated.
Creativity	The product of Originality and Usefulness scores. Scale of 1 – 49. Fluency independent scores were calculated.
Best Chosen Response	The highest Creativity score of the two responses nominated by each participant as their best ideas.
Creativity Evaluation	The correspondence between Best Chosen Response and the highest Creativity score. Calculated as a proportion with a maximum score of 1.
Flexibility	The number of thematic or object categories included in the response set.
Within-Category Fluency	The mean number of ideas within each thematic or object category.

3.3.1.2 Story Scenes Task.

3.3.1.2.1 Background.

This task used the premise of the Storytelling task in the Eco-TED (Pullinger, 2017), whereby participants were asked to listen to and then recall stories using picture prompt cards. This had shown promising results in differentiating between children with and without ASD. Storytelling is frequently used in clinical interventions with children with ASD, for example, Comic Strip Conversations (Gray, 1998) and

the ADOS (Lord et al., 1989). Furthermore, during piloting of the current task all children confirmed familiarity with either comic strips or telling stories in a school setting, suggesting ecological validity of the premise. Storytelling also allowed for the assessment of creativity in a literary domain, as opposed to the more domain general ability required by traditional DT tasks.

Therefore, the current task was designed to incorporate these elements. In the Story Scenes task, participants are shown picture scene cards in the form of a comic strip and asked to create a story using these pictures. The scenes were in cartoon format (see Appendix D). Two different Story Scenes packs were used, each with four pictures making up the comic strip.

In each pack, three pictures were thematically related whereas one picture was random and not obviously related to the set. The inclusion of the random scene was intended to make participants think creatively and generate unusual connections between story components. Story Pack 1 depicted human characters whereas Story Pack 2 depicted scenes without human characters, to encourage participants to incorporate novel elements into their stories.

The start of the story was read aloud by the researcher and the participant was asked to continue when they were ready. Providing the start of the story was intended to overcome the difficulties children with ASD may experience in getting started and to be able to assess the coherence and elaboration of the story scripts from the initial starting point.

3.3.1.2.2 Administration.

The story scenes were always presented in the same order and with the same instructions for each participant (please refer to Appendix B for full instructions). To begin, participants were presented with Story Pack 1 cards and asked to tell a story using the pictures in order, as in a comic strip. They were encouraged to be as creative as possible and allowed to add in other things that could not be observed in the pictures too. In the second condition, the Story Pack 1

cards were shuffled, and the participant was asked to tell a *different* story using the same pictures.

The cards were then collected, and Story Pack 2 cards presented. The child was again asked to tell a story using the pictures in order, as in the first condition with Story Pack 1. Three blank cards were then added to the comic strip, at the beginning, middle and end positions. In this third condition, the child was required to tell a *different* story by filling in the missing pictures to complete the comic strip.

Each child therefore produced four stories overall. These conditions are summarised in Table 2.3. Stories were recorded using an audio recorder.

Table 2.3. Story Scenes task conditions.

	Condition 1 (Items 1 and 3)	Condition 2 (Item 2)	Condition 3 (Item 4)
Story Pack 1	Tell a story using the pictures in order	Tell a different story once pictures are shuffled	
Story Pack 2	Tell a story using the pictures in order		Tell a different story once three blank pictures are added

3.3.1.2.3 Scoring Criteria.

Each story was divided into syntactic units following Norbury and Bishop's (2003) guidelines.

Originality was determined by calculating the number of original units. This was represented as a proportion of the total number of units to control for story length and verbal productivity. An original unit was defined as a novel addition to the story, beyond a basic description of the scene (e.g. the girl is wearing blue trousers) or a variation of the basic story outline (see Appendix D). These plot summaries

were written by the main researcher following piloting. For items 2 and 4, the unit had to be both original *and* unique to the trial to receive a point. Therefore, any repetitions of story features from the previous item were deemed not original.

Usefulness was defined in relation to the narrative coherence and narrative elaboration of the story, adapted from the scales described by Dillon and Underwood (2012; see Appendix E). Narrative coherence referred to the degree to which the story could be linked together to form a logical and appropriate action sequence and narrative elaboration referred to the coherent development of both story and character beyond the original starting point.

A Creativity score was calculated by multiplying the sum of the narrative coherence and elaboration scores by the proportion of original units (yielding a score between 0 - 6). Variables were calculated for each condition and for the task overall. The Total Proportion of Original Units and Total Creativity were calculated by taking an average score whereas Total Narrative Coherence and Total Narrative Elaboration were summed across items.

3.3.1.3 Situations and Solutions Task.

3.3.1.3.1 Background.

A previous trainee project contributing to the development of the Eco-TED (Ledger-Hardy, 2017) used qualitative analysis of interview data collected from parents of children with ASD to design a questionnaire capturing the everyday behavioural difficulties experienced by children with ASD. These difficulties were conceptualized as sitting within a broader framework of executive functioning.

Several of these items were taken as the basis for the design for this task (e.g. Item 6, 'My child has difficulty organising him/herself to get ready in the morning') thereby enhancing ecological validity by reflecting situations commonly encountered by children with ASD and increasing congruity with other measures included within the Eco-TED battery. The task also aimed to demonstrate practical

clinical utility by requiring participants to generate ideas to solve these everyday problems.

However, as creative problems differ from more routine problems in their degree of novelty, complexity and structure (Mumford, Baughman, Threlfall, Supinski & Costanza, 1996), the items were designed to vary in their degree of familiarity. The final task consisted of six items or problem situations, three which reflected problems relating to characteristic ASD and childhood difficulties (Condition 1) and three which were more unusual problems not likely to be encountered routinely and therefore relying less on prior knowledge or experience (Condition 2). Importantly, these situations were designed to simulate novel but plausible events rather than imaginary propositions.

3.3.1.3.2 Administration.

Participants were presented with pictures of eight everyday objects and then read a description of the problem situation. They were asked to generate as many solutions as possible using the objects provided. Participants were allowed to use the objects in whichever way they like, thus allowing both creative and ordinary uses. They were instructed that they could use the objects in combination and multiple times but could only use the objects provided.

Prior to commencing the test items the participants completed a recall exercise to ensure understanding of the instructions and a naming and recall exercise of the eight objects. Participants were reminded briefly of the instructions after every item and were prompted following an error. The researcher wrote down the responses. Following each item, the child selected their two best ideas.

3.3.1.3.3 Scoring Criteria.

Please refer to Appendix C for a detailed description of the scoring criteria. The same variables and scoring method were used as in the Toy Improvement task (see Table 2.2) with the exception of:

- Flexibility, which in this task was counted as the number of *object* categories included in the response set; and
- Within-Category Fluency, which similarly took the mean frequency of ideas per *object* category, as opposed to using thematic categories.

For each criterion, scores were averaged across the items within each condition: Condition 1 with items resembling everyday problems (items 1, 2, and 5); and Condition 2 with items resembling unusual problems (items 3, 4 and 6). A Total score was also calculated by averaging across all items, except for Total Fluency, which was summed.

3.3.1.4 Measure Acceptability Scale.

Following each creativity task, participants were asked to rate how much they liked that activity using a 100mm visual analogue scale (see Appendix F). Any other comments or feedback about the tasks were also recorded.

3.3.2 Parent-Report Questionnaires

3.3.2.1 Social Communications Disorder Checklist (SCDC; Skuse et al., 1997).

The SCDC was used to measure autistic traits in the clinical group and screen the control group for any socio-communicative difficulties. This measure has high sensitivity (0.9) and specificity (0.69) when discriminating ASD from non-ASD cases. In addition, the SCDC is a very reliable instrument, with excellent internal consistency (α = .93) and high test-retest reliability (*r* =.81 over two years; Skuse, Mandy & Scourfield, 2005).

3.3.2.2 Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997).

The SDQ was used as a means of characterising the clinical sample to assess the generalizability of the findings. This measure has acceptable reliability and validity (Goodman, 2001) and has been shown to have good predictive validity across a range of clinical disorders (Goodman & Goodman, 2009).

3.3.2.3 Parent's Evaluation of Children's Creativity (PECC; Runco, Johnson & Bear, 1993; Appendix G).

The PECC was used to measure the parent-reported creative characteristics of the child participants as a measure of construct validity. This questionnaire was compiled using social validation methodology and is representative of parents' implicit views of creativity. It consists of 25 items (e.g. To what degree or how often is this child adventurous?) rated on a scale of 1 (rarely) to 7 (extremely). The average overall rating is taken as the main outcome. This measure has high inter-rater agreement between mothers and fathers (r = 0.72) and moderate convergent validity with children's self-assessments of creativity (r = 0.48 with mother ratings and r = 0.36 with father ratings; Runco, Johnson & Bear, 1993).

3.3.2.4 The Behaviour Rating Inventory of Executive Functions (BRIEF; Gioia, Isquith, Guy & Kenworthy, 2000) parent-report form.

This questionnaire was used in the current study to assess discriminant validity of the creativity tasks. It is an 86-item questionnaire designed to assess everyday behavioural manifestations of executive dysfunction in the home environment. An overall score (Global Executive Composite; GEC) is derived, with higher scores indicating a higher degree of behavioural problems. The BRIEF has satisfactory internal consistency and test retest reliability (r = 0.81) and detects large between-group differences when measured in children with and without ASD, providing criterion validity (Demetriou et al., 2017; Gioia et al., 2000).

3.3.3 Test Battery for Child Participants

3.3.3.1 Inventory of Creative Activities and Achievements (ICAA; Jauk, Benedek, Dunst & Neubauer, 2013; Appendix H).

This questionnaire is a self-report measure of the frequency of different activities and the level of achievement in eight creative domains (Literature, Music, Arts and Crafts, Cooking, Sports, Visual Arts, Performing Arts, Science) over the

last 10 years. Items are summed across subscales to derive a domain-general Activities and a domain-general Achievements scale, with high scores indicating a greater level of creative pursuit. This measure has demonstrated satisfactory internal consistency, (α = 0.78 for the Activities scale; α = 0.71 for the Achievements scale) and high convergent validity with the Creative Achievement Questionnaire (*r* = 0.68; Jauk, Benedek, & Neubauer, 2014). The ICAA was used to assess criterion validity. Both the Total and Literature scores were analysed, as the latter domain-specific subscale was thought to most closely map onto the creative ability measured in the Story Scenes task.

3.3.3.2 The Schoolbag task, Lego task and Alternating Sequence task from the Eco-TED battery (Appendix I).

These tasks are described in more detail in Bristow (2016) and Pullinger (2017). In all tasks, test-retest reliability was below the required level and veridicality was an area of weakness. However, these measures showed the most promising results in the pilot study for picking up everyday executive functioning difficulties experienced by children with ASD (Bristow, 2016). They were used in the current study to contribute to ongoing data collection with the Eco-TED battery and to assess discriminant validity of new creativity measures. The Schoolbag Total Time Paused, Lego Total Time and Alternating Sequence Total Errors were chosen as the outcome measures for each task.

3.3.3.3 The Six-Parts and the Zoo Map sub-tests of the Behavioural Assessment of Dysexecutive Syndrome in Children (BADS-C; Emslie, Wilson, Burden, Nimmo-Smith & Wilson, 2003).

These subtests have been chosen as the adult versions are known to differentiate individuals with ASD from typically developing controls (Hill & Bird, 2006). The BADS-C has excellent inter-rater reliability and good construct and discriminant validity (Baron, 2007; Engel-Yeger, Josman, & Rosenblum, 2009).

These subtests were used as a further measure of discriminant validity of the new creativity tasks.

3.3.3.4 Two-subtest form of the Wechsler Abbreviated Scale of Intelligence (WASI-II; Wechsler, 2011).

This was used to determine the IQ of the participants to allow for matching of clinical and control samples and to ensure all participants had an IQ > 70 to limit the chance of test performance being affected by general learning difficulties. The WASI-II has acceptable reliability and validity (Irby & Floyd, 2013), and its two subtest version correlates strongly with the full scale version (r = 0.83; Homack & Reynolds, 2007).

3.4 Procedure

3.4.1 Ethics

Ethical approval was granted by the Westminster NHS Research Ethics Committee (ref 15/LO/1332; Appendix J). All parents provided written consent and children provided written and verbal assent prior to commencing the study. Copies of the participant information sheets and consent forms can be found in Appendix K.

3.4.2 Participant Recruitment

The ASD participants were contacted by a member of the hospital clinical team if they had previously given permission to be contacted for research purposes. The control participants were initially sent the invitation letter and information sheets by the school or by the point of contact for those who were recruited by other means and asked to opt in to the study. The researcher then made contact and arranged a time for testing to take place. Copies of the invitation letters sent to parents are in Appendix L.

3.4.3 Testing Procedure

Testing took place in a quiet room either at school, home or in a clinic space at UCL or GOSH. The testing session lasted no longer than two hours and breaks

were offered throughout. The main researcher carried out all sessions except for three that were conducted by a trained research assistant at GOSH, one of which was observed as part of training. Participants completed all measures unless they had previously participated in the research, in which case only the new creativity tasks and questionnaires were administered. Parents were given the option to be sent the questionnaires beforehand or to complete them at the time of testing. Participants were given a £5 voucher as a thank you for taking part.

3.4.4 Retest Procedure

Ten participants (5 clinical and 5 control) were approached to take part in the study a second time, two months after their initial testing session. Of these, two declined and one failed to respond, resulting in a total of 3 clinical and 4 control participants for the retest analysis with a mean time interval of 74 days between testing sessions. These participants were offered an additional £5 for taking part.

3.4.5 Scoring

All data were scored by the main researcher. An independent second rater, blind to diagnosis, also scored each response on the Toy Improvement and Situations and Solutions tasks on the variables of Originality and Usefulness using the newly developed subjective rating method.

Training involved reviewing a sample dataset with scores already input before proceeding with independent practice scoring data, which was then discussed and revised with the main researcher. During test scoring progress was intermittently reviewed and any internal inconsistencies were highlighted by the main researcher and reviewed by the second rater.

For the test data, each response was typed into a database by the main researcher and sorted alphabetically. There were a total of 1328 responses; repetitions were removed to reduce potential bias in scoring (such as infrequent

responses automatically being awarded higher Originality scores) and this yielded a net total of 1149 responses to be scored.

The instructions were to read the entire set of responses for each item and then assign the least and most Original and Useful category exemplars respectively. This then served as anchor points for scoring the rest of the dataset. Raters were encouraged to score the entire set of responses for an item in one session to ensure consistency, but to take breaks between items to prevent scoring fatigue and saturation. The intraclass correlation coefficient (ICC) was calculated and the average rating between raters used.

For the Story Scenes Task, all stories were transcribed by the second rater and checked by the main researcher. Initial stories were scored by the main researcher and reviewed with the second rater to learn the scoring method. Following this the second rater practiced scoring a set of stories and received feedback. Scoring was then shared between raters and all stories checked by the main researcher to ensure consistency.

3.4.6 Consensual Assessment Technique

To assess construct validity of the new creativity task and developed scoring system, a small panel of 'experts' were asked to judge the creative products using the Consensual Assessment Technique (CAT; Amabile, 1982). CAT is based on the assumption that a product or response is creative insofar as a panel of appropriate independent raters, familiar with the product domain, agree that it is. The technique does not require any training by the researcher and as such the judgements are subjective and socially validated.

The chosen experts were tutors with experience of teaching English creative writing to children and were recruited through a personal contact of the main researcher. Teacher ratings are a frequently used criteria of creativity and also as a method for validating existing scales and creativity tests (Hocevar, 1981). Furthermore, limited instruction or training is required to validly derive these

measurements. For example, Drevdahl (1956) asked faculty members in arts and science department to rate their students on a 7-point scale of creativity using both their own subjective definition and the definition provided by researcher. No significant difference was found between undefined and defined ratings.

In this study, a panel of five tutors with a minimum of 3 years and 200 hours (range 200 – 350 hours) of teaching English creative writing to children aged between 8 and 14 years old were invited to take part. They were provided with a sample of creative stories generated by the clinical (n=4) and control (n=4) groups in the Story Scenes Task Item 3 and asked to independently rate the creativity of each story using their own subjective judgements on a scale of 1 (not creative) to 7 (extremely creative).

These stories were selected to represent the full range of scores on the Story Scenes Creativity scale and two stories (one ASD and one control) were selected from each quartile range. Raters were given a £5 amazon voucher for taking part. A copy of the instructions provided to the expert raters are provided in Appendix M. Inter-rater reliability was assessed and an average rating taken.

3.5 Data Analysis Procedures

The primary aim of this study was to conduct an initial exploration of the acceptability and psychometric properties of the three newly developed creativity tasks and rating scales. Firstly, to examine inter-rater reliability of the newly developed subjective Originality and Usefulness scales used in the Toy Improvement and Situations and Solutions tasks, correlations between raters' scores were assessed. Test-retest reliability was established by examining correlations between Total task scores at both time points.

Questionnaire measures of everyday creative behaviour (both domain-general and domain specific) were correlated with Total task scores as a measure of criterion validity. Construct validity was examined by measuring correlations

between: Total task scores and parental ratings of creativity; Creativity scores on the Story Scenes task and expert judgements of the same stories made using the CAT. If the operationalised definition of creativity is conceptually related to these socially validated definitions of creativity then correspondence between these scores would be expected. Discriminant validity was assessed by correlations with the measures of executive functioning (Eco-TED, BADS-C and BRIEF) and FSIQ.

Next, between-group comparisons were carried out to assess performance on the various measures of creativity as a function of varying task condition. This preliminary analysis was intended to explore the characteristics of the creative profile in children with ASD and to assess whether these task manipulations differentiated the clinical ASD group from the control group, or in fact ameliorated group differences. The ASD and control groups were compared on their performance on each outcome variable for every task condition, yielding a total of 52 comparisons. (Please see paragraph below for consideration of how Type I error risk was understood and managed).

Where data were not normally distributed, non-parametric tests were applied rather than transforming the data due to the inherent difficulties with assessing normality of small N datasets (Field, 2009) and the problems with interpreting transformed data (Grayson, 2004). These cases are indicated. Correlations were conducted using Kendall's tau statistic as this is arguably a better estimate of the effect in the population when using small samples with tied ranks (Field, 2009). Effect sizes are reported according to the conventions described by Cohen (1992).

As the analysis took an exploratory stance, multiple tests were run which increased the chance of Type I error. However, as the Bonferroni is conservative and increases the chance of Type II error, it was decided to report significant results prior to applying this correction and indicate if any results remained significant following the Bonferroni correction for multiple comparisons (calculated by dividing the α-level by the number of comparisons within a subtest). Any interpretations of

the data are therefore tentative. Furthermore, due to the moderate sample size, the study had limited statistical power to detect any small effects and therefore null findings must also be interpreted with caution due to the increased risk of Type II error.

4 Results

4.1 Sample Characteristics: SCDC and SDQ

Questionnaire data were missing for one ASD participant due to failure to collect the measures. As expected, the ASD group scored significantly higher on the SCDC (M = 16.71, SD = 4.97) than the control group (M = 1.73; SD = 2.58; U = 1.50, $p < .001$). The ASD group also scored significantly higher on the SDQ Overall Stress index (M = 19.29, SD = 2.05) than the control group (M = 4.80, SD = 1.27; $t(27) = -6.10$, $p < .001$, 95% CI = -19.36, -9.61).

4.2 Inter-rater Reliability

The inter-rater reliabilities were assessed using two-way random, consistency, average-measures ICC (2,2) to determine the degree that both coders provided consistency in their subjective ratings of Originality and Usefulness across the entire response set in both the Toy Improvement and Situations and Solutions tasks. The resulting ICCs were greater than 0.50 and were within the fair to good range of reliability (Cicchetti, 1994). These are presented in Table 2.4. Therefore, an average rating was taken for the scales and used in all subsequent analysis.

Table 2.4. Intraclass Correlation Coefficients for the Originality and Usefulness Scales

Scale	ICC	95% CI
Toy Improvement Originality	.57	.50, .64
Toy Improvement Usefulness	.60	.54, .66
Situations and Solutions Originality	.62	.55, .68
Situations and Solutions Usefulness	.69	.63, .74

4.3 Test-retest Reliability

The small number of participants (n = 7) and the time lag between test sessions (*M* = 74 days) pose potential threats to the validity of the test-retest analysis and limit the generalizability of the results.

Correlation coefficients are reported in Table 2.5. There were no significant correlations between measures of creativity at both time points although Toy Improvement Total Originality; Story Scenes Total Narrative Elaboration; and Situations and Solutions Total Flexibility measures approached significance. The former two measures were within the acceptable range of above .70 suggested for psychometric measures (Cicchetti & Sparrow, 1990) and therefore demonstrated the most promise in relation to retest reliability.

Table 2.5. Test-retest reliability coefficients for the Total task scores

Task	**Variable**	**Mean (SD) Time 1**	**Mean (SD) Time 2**	**Correlation coefficient**	***p***
Toy Improvement	Total Fluency	20.57 (3.78)	25.29 (9.53)	-.173[T]	.618
	Total Originality	3.02 (0.35)	3.24 (0.44)	.714	.071[+]
	Total Usefulness	3.70 (0.38)	3.82 (0.38)	-.163	.727
	Total Creativity	11.93 (1.96)	12.82 (1.52)	.056	.904
	Total Best Chosen Response	16.09 (3.34)	16.89 (2.45)	-.219	.637
	Total Creativity Evaluation	.90 (.09)	.85 (.10)	-.118	.802
	Total Flexibility	2.54 (0.41)	3.17 (0.88)	.291	.527
	Total Within-Category Fluency	1.66 (0.27)	1.63 (0.42)	-.280	.543
Story Scenes	Total Proportion Original Units	.72 (.08)	.81 (.08)	.023	.960
	Total Narrative Coherence	8.00 (1.15)	8.71 (1.70)	.508	.244

	Total Narrative Elaboration	7.86 (1.95)	8.14 (2.79)	.738	.058[+]
	Total Creativity	2.90 (0.84)	3.45 (1.09)	.615	.142
Situations and Solutions	Total Fluency	20.57 (5.80)	22.29 (8.42)	-.028	.953
	Total Originality	3.25 (0.36)	3.53 (0.46)	-.028	.952
	Total Usefulness	4.66 (0.41)	4.74 (0.32)	-.026	.956
	Total Creativity	15.57 (3.00)	17.09 (2.40)	-.105	.822
	Total Best Chosen Response	19.68 (4.03)	23.33 (4.29)	.164	.725
	Total Creativity Evaluation	.92 (.07)	.95 (.06)	-.085	.856
	Total Flexibility	4.36 (1.39)	4.69 (1.43)	.551[T]	.091[+]
	Total Within-Category Fluency	1.17 (0.13)	1.27 (0.22)	-.098[T]	.761

[+] $p < .10$
[T] Kendall's tau coefficient. All other statistics are Pearson's product-moment correlation coefficient.

4.4 Criterion Validity

Correlations with self-reported creative activities and achievements were assessed (see Table 2.6). The ICAA domain-general Total Creative Activities and Total Creative Achievements scores were analysed alongside the domain-specific Literary Activities and Literary Achievements subscales. ICAA data were missing for one control participant due to failure to return the measure.

The ICAA Literary Achievements scale was found to significantly correlate with three of the four Story Scenes measures (Total Proportion Original Units; Total Narrative Elaboration; and Total Creativity), as well as the Situations and Solutions Total Usefulness scales. ICAA Literary Activities did not correspond with any measures in the Story Scenes task but did relate to the Toy Improvement (Total Fluency and Total Flexibility) and Situations and Solutions tasks (Total Originality; Total Usefulness; Total Creativity; and Total Best Chosen Response). This latter correlation was the only to remain significant following Bonferroni correction.

There were significant correlations between ICAA Total Creative Achievements and the Toy Improvement (Total Fluency and Total Flexibility) and Story Scenes tasks (Total Narrative Elaboration). Also between ICAA Total Creative Activities and Toy Improvement (Total Flexibility) and Story Scenes (Total Proportion Original Units).

Table 2.6. Correlations between Total task scores and questionnaire measures of creativity

Task	Variable	ICAA Total Creative Activities	ICAA Total Creative Achievements	ICAA Literary Activities	ICAA Literary Achievements	PECC
Toy Improvement	Total Fluency	.249 T	.291 T*	.309 T*	.139 T	.269 T*
	Total Originality	-.041	.069 T	.044	.099 T	.089
	Total Usefulness	-.109	-.022 T	.067	.255 T	.066
	Total Creativity	-.092	.027 T	.012	.192 T	.081
	Total Best Chosen Response	.068	.104 T	.091	.232 T	.134
	Total Creativity Evaluation	-.179	-.045 T	-.123	-.019 T	-.177
	Total Flexibility	.500**	.267 T*	.519**	.233 T	.288
	Total Within-category fluency	.158	.126 T	.092	.013 T	.227
Story Scenes	Total Proportion Original Units	.342 T**	.160 T	.232 T	.319* T	.199 T
	Total Narrative Coherence	.165 T	.120 T	.164 T	.249 T	.187 T
	Total Narrative Elaboration	.301	.302 T*	.279	.379 T**	.132

Task	Variable	ICAA Total Creative Activities	ICAA Total Creative Achievements	ICAA Literary Activities	ICAA Literary Achievements	PECC
	Total Creativity	.314	.239[T]	.352	.293[T]*	.244
Situations and Solutions	Total Fluency	-.050	.032[T]	.147	-.013[T]	.272
	Total Originality	.192	-.022[T]	.433*	.048[T]	.114
	Total Usefulness	.099	.072[T]	.448*	.356[T]**	-.058
	Total Creativity	.183	.057[T]	.483**	.250[T]	.073
	Total Best Chosen Response	.342	.202[T]	.570**[B]	.248[T]	.260
	Total Creativity Evaluation	.235	.060[T]	-.033	.054[T]	-.149
	Total Flexibility	.066	.015[T]	.225	-.040[T]	.336
	Total Within-Category Fluency	.113[T]	.125[T]	.248[T]	.011[T]	.144[T]

* $p < .05$
** $p < .01$
[B] Significant following Bonferroni correction for multiple comparisons.
[T] Kendall's tau coefficient. All other statistics are Pearson's product-moment correlation coefficients.
PECC = Parent's Evaluation of Children's Creativity
ICAA = Inventory of Creative Activities and Achievements

4.5 Construct Validity

4.5.1 Convergent Validity

4.5.1.1 Parent ratings

The parental-report PECC was only significantly related to one measure, Toy Improvement Total Fluency (see Table 2.6). This did not remain significant following Bonferroni correction.

4.5.1.2 Expert ratings using CAT

The creativity of a sample of eight stories produced by participants in the Story Scenes Task Item 3 were evaluated by a panel of five independent 'experts' using the CAT. Inter-rater reliability was determined by a two-way random, consistency, average-measures ICC (2,5) and found to be within the excellent range (ICC = .95, 95% CI = .87, .99). An average of the five raters' scores was therefore taken for each story.

The correlation between this expert consensus rating and the Creativity measure for Item 3 was used to assess the validity of the developed scoring method for the Story Scenes task. The Pearson's correlation coefficient was found to be highly significant (r = .941, p < .001) demonstrating excellent construct validity of the newly developed scoring method in relation to the gold-standard CAT.

4.5.2 Discriminant Validity

Correlations with theoretically related but distinct constructs of IQ and EF were assessed (see Table 2.7). BRIEF and BADS-C measures were missing for one ASD participant due to time limitations during the testing session. Only two comparisons remained significant following Bonferroni correction (indicated within text below).

FSIQ was significantly correlated with all measures in the Story Scenes task; six of the eight measures in the Situations and Solutions task (Total Fluency; Total Originality; Total Usefulness; Total Creativity; Total Best Chosen Response; and

Total Flexibility); and Toy Improvement Total Best Chosen Response. The correlation between Story Scenes Total Creativity and FSIQ was particularly robust and remained significant following Bonferroni correction.

BRIEF GEC was significantly correlated with two measures in the Toy Improvement task (Total Usefulness and Total Creativity); three in the Situations and Solutions task (Total Usefulness; Total Creativity; and Total Best Chosen Response); and with all measures of the Story Scenes task. The correlation between Story Scenes Total Narrative Coherence and BRIEF GEC remained significant following Bonferroni correction.

The BADS-C Six Part subtest scaled score correlated significantly with Toy Improvement Total Fluency; and Situations and Solutions Total Usefulness. The BADS-C Zoo Map 2 subtest correlated significantly with the Toy Improvement Total Creativity Evaluation score.

Of the Eco-TED battery, the only significant correlations were between AS Total Errors and Situations and Solutions Total Best Chosen Response and Total Flexibility.

Table 2.7. Correlations between Total task scores and measures of IQ and EF

Task	Variable	FSIQ	BADS-C Zoo Map 1	BADS-C Zoo Map 2	BADS-C Six Part	BRIEF GEC	Eco-TED School Bag Total Time Paused	Eco-TED Lego Total Time	Eco-TED AS Total Errors
Toy Improvement	Total Fluency	.130	-.025	-.137	-.274*	.056	-.115	-.101	-.088
	Total Originality	.193 [r]	.128	.283	.152 [r]	-.185	-.113 [r]	.037	.039
	Total Usefulness	.307 [r]	-.056	.095	.253 [r]	-.291*	.154 [r]	.090	.027
	Total Creativity	.311 [r]	.045	.197	.227 [r]	-.296*	-.019 [r]	.048	.115
	Total Best Chosen Response	.412 [r]*	-.021	.175	.099	-.190	-.076 [r]	.083	.095
	Total Creativity Evaluation	-.016 [r]	-.109	.296*	.088	-.058	-.185 [r]	.174	.173
	Total Flexibility	.257 [r]	.025	-.089	-.274 [r]	-.005	-.024 [r]	-.116	-.023
	Total Within-Category Fluency	-.149	.000	-.128	-.181 [r]	.020	-.229 [r]	-.058	-.194

Task	Variable	FSIQ	BADS-C Zoo Map 1	BADS-C Zoo Map 2	BADS-C Six Part	BRIEF GEC	Eco-TED School Bag Total Time Paused	Eco-TED Lego Total Time	Eco-TED AS Total Errors
Story Scenes	Total Proportion Original Units	.372**	-.120	.038	.085	-.385**	.069	-.007	-.134
	Total Narrative Coherence	.339*	-.049	.086	.161	-.436**[B]	-.121	.002	.076
	Total Narrative Elaboration	.434[r]*	-.017	-.003	.148[r]	-.279*	-.206[r]	-.017	-.090
	Total Creativity	.532[r]**[B]	-.029	.076	.153[r]	-.360**	-.117[r]	-.044	-.090
Situations and Solutions	Total Fluency	.412[r]*	.024	-.113	-.094[r]	-.057	.062[r]	-.200	-.116
	Total Originality	.499[r]*	.066	.178	-.011[r]	-.257	.149[r]	-.057	-.227
	Total Usefulness	.396[r]*	-.003	-.038	.376[r]*	-.321*	-.089[r]	-.076	-.173
	Total Creativity	.498[r]**	.024	.076	.194[r]	-.316*	.045[r]	-.094	-.257

Task	Variable	FSIQ	BADS-C Zoo Map 1	BADS-C Zoo Map 2	BADS-C Six Part	BRIEF GEC	Eco-TED School Bag Total Time Paused	Eco-TED Lego Total Time	Eco-TED AS Total Errors
Situations and Solutions	Total Best Chosen Response	.551 [r]**	-.101	.016	.189 [r]	-.289*	.016 [r]	-.120	-.303*
	Total Creativity Evaluation	-.196 [r]	-.124	.019	.143 [r]	-.013	-.106 [r]	.088	-.062
	Total Flexibility	.398 [r]*	.097	.016	-.030 [r]	-.165	.015 [r]	-.219	-.304*
	Total Within-Category Fluency	.242	-.070	.168	-.089	-.228	-.136	-.124	-.204

* $p < .05$
** $p < .01$
[B] Significant following Bonferroni correction for multiple comparisons.
[r] Pearson's correlation coefficient. All other reported statistics are Kendall's tau coefficient.
AS: Alternating Sequence Task
GEC: Global Executive Composite

4.6 Measure Acceptability

The average of acceptability ratings given by participants for each task is presented in Table 2.8. Scores were made on a Visual Analogue Scale of 100mm, and a maximum score of 100 indicated maximum enjoyment of the task.

There were no group differences suggesting both the ASD and control participants rated each task comparably. The Situations and Solutions task was most liked by participants and the Toy Improvement task the least; although high average ratings for all tasks suggest favourable opinion of the battery as a whole.

Table 2.8 Acceptability ratings given by participants for each task

Task	**ASD Mean (SD)**	**Control Mean (SD)**	***t***	***p***
Toy Improvement	69.20 (21.00)	62.53 (20.65)	-0.88	.388
Story Scenes	71.33 (19.06)	66.93 (26.58)	-0.52	.606
Situations and Solutions	76.20 (21.56)	73.80 (21.50)	0.76	.762

4.7 Between Group Comparisons

The effect of varying the task conditions on performance on the various measures of creativity was compared between groups. The large number of comparisons and the small sample size mean that both Type I and Type II error are a risk for these analyses. Therefore, to convey the true precision of group difference estimates, 95% confidence intervals for standardized effect sizes are provided.

4.7.1 Toy Improvement

These results are displayed in Table 2.9. No results remained significant following Bonferroni correction.

In Condition 1, control participants scored significantly higher than the ASD group on the Originality and Best Chosen Response measures. These were both large effect sizes. There was a trend in the same direction for the Usefulness and Creativity scales, with medium and large effect sizes respectively.

In Conditions 2 and 3, however, no group differences were found on these measures. No group differences in Fluency, Flexibility or Within-Category Fluency were observed between groups in any condition.

Group differences on the Creativity Evaluation scale, measuring alignment between Best Chosen response and the highest Creativity score, were non-significant in Conditions 1 and 2. However, in Condition 3 (Context Items) there was a near significant trend for the ASD participants to be better at selecting their most creative responses. This was a large effect size.

Table 2.9 Between group comparisons in the Toy Improvement task.

Measure	Condition (description)	ASD Mean (SD)	Control Mean (SD)	Test statistic	*p*	Effect Size *d*	95% CI of Effect Size	
Fluency	1 (Baseline)	6.20 (2.86)	6.13 (3.25)	-0.06	.953	0.02	-0.69	0.74
	2 (Function items)	3.73 (1.41)	4.40 (2.33)	0.94	.352	-0.35	-1.06	0.38
	3 (Context items)	3.83 (1.68)	4.93 (2.77)	91.50^{U}	.381	-0.48	-1.19	0.26
Originality	1 (Baseline)	1.94 (0.66)	2.60 (0.76)	2.52	.018*	-0.93	-1.65	-0.15
	2 (Function items)	3.11 (0.66)	3.17 (0.43)	0.31	.761	-0.11	-0.82	0.61
	3 (Context items)	3.45 (0.76)	3.33 (0.59)	-0.47	.641	0.18	-0.55	0.89
Usefulness	1 (Baseline)	2.76 (0.84)	3.28 (0.55)	1.99	$.058^{+}$	-0.73	-1.45	0.03
	2 (Function items)	3.38 (0.82)	3.58 (0.76)	0.72	.480	-0.25	-0.96	0.47
	3 (Context items)	3.98 (0.81)	4.13 (0.80)	0.51	.612	-0.19	-0.90	0.54
Creativity	1 (Baseline)	6.18 (3.33)	9.11 (3.22)	67.00^{U}	$.059^{+}$	-0.89	-1.62	-0.12

Measure	Condition (description)	ASD Mean (SD)	Control Mean (SD)	Test statistic	*p*	Effect Size *d*	95% CI of Effect Size	
	2 (Function items)	11.59 (4.23)	11.97 (3.73)	0.27	.792	-0.10	-0.81	0.62
	3 (Context items)	13.06 (4.38)	14.27 (3.76)	0.82	.421	-0.30	-1.01	0.43
Best Chosen Response	1 (Baseline)	9.30 (5.08)	15.82 (8.30)	2.60	.015*	-0.95	-1.67	-0.17
	2 (Function items)	15.98 (5.66)	16.04 (4.56)	0.03	.975	-0.01	-0.73	0.70
	3 (Context items)	18.65 (7.04)	18.63 (5.16)	-0.01	.994	0.00	-0.71	0.72
Creativity Evaluation	1 (Baseline)	.78 (.26)	.90 (.19)	84.00 [U]	.191	-0.53	-1.24	0.21
	2 (Function items)	.91 (.16)	.93 (.10)	108.00 [U]	.829	-0.15	-0.86	0.57
	3 (Context items)	.97 (.05)	.89 (.13)	73.50 [U]	.068[+]	0.81	0.05	1.53
Flexibility	1 (Baseline)	3.07 (1.16)	3.27 (1.28)	107.00[U]	.812	-0.16	-0.88	0.56
	2 (Function items)	2.30 (0.88)	2.47 (0.90)	0.51	.612	-0.19	-0.9	0.53
	3 (Context items)	2.47	2.53	0.17	.869	-0.05	-0.77	0.66

Measure	Condition (description)	ASD Mean (SD)	Control Mean (SD)	Test statistic	*p*	Effect Size *d*	95% CI of Effect Size	
		(1.08)	(1.11)					
Within-Category Fluency	1 (Baseline)	2.02 (0.98)	1.89 (0.91)	99.50[U]	.588	0.14	-0.58	0.85
	2 (Function items)	1.75 (0.46)	1.84 (0.46)	0.56	.579	-0.20	-0.91	0.53
	3 (Context items)	1.68 (0.54)	2.05 (0.67)	1.66	.108	-0.61	-1.32	0.14

* $p < .05$
+ $p < .10$
[U] Mann-Whitney Test *U* statistic. All other reported statistics are independent samples T test.

4.7.2 Story Scenes

These results are displayed in Table 2.10. No results remained significant following Bonferroni correction.

Significant group differences were found in the Proportion of Original Units, Narrative Coherence and Creativity of stories in Condition 1, with the control group scoring more highly on these measures than the ASD group. These were all large effect sizes.

The same pattern was observed for Narrative Coherence of stories in Condition 3, with a large recorded effect size, and there was a trend towards significance for stories produced by the control group in this condition to have higher Creativity scores, reduced to a medium effect size between groups.

There were no group differences in Narrative Elaboration in any condition and no group differences observed in any measure in Condition 2.

Table 2.10 Between group comparisons in the Story Scenes task.

Measure	**Condition (description)**	**ASD Mean (SD)**	**Control Mean (SD**	**Test statistic**	***p***	**Effect Size *d***	**95% CI of Effect Size**	
Proportion Original Units	1 (Baseline)	.56 (.23)	.72 (.13)	2.39	.026*	-0.86	-1.58	-0.09
	2 (Shuffled Scenes)	.53 (.31)	.70 (.19)	79.00[U]	.164	-0.66	-1.38	0.09
	3 (Added Scenes)	.72 (.20)	.77 (.16)	0.84	.406	-0.28	-0.99	0.45

Measure	Condition (description)	ASD Mean (SD)	Control Mean (SD	Test statistic	*p*	Effect Size *d*	95% CI of Effect Size	
Narrative Coherence	1 (Baseline)	1.80 (0.53)	2.20 (0.37)	64.50[U]	.029*	-0.88	-1.60	-0.10
	2 (Shuffled Scenes)	1.67 (0.98)	1.80 (0.68)	106.00[U]	.772	-0.15	-0.87	0.57
	3 (Added Scenes)	1.80 (0.77)	2.47 (0.52)	59.00[U]	.013*	-1.02	-1.75	-0.23
Narrative Elaboration	1 (Baseline)	1.53 (0.92)	1.83 (0.65)	89.00[U]	.315	-0.38	-1.09	0.36
	2 (Shuffled Scenes)	1.67 (0.90)	1.80 (0.77)	100.00[U]	.578	-0.16	-0.87	0.57
	3 (Added Scenes)	1.87 (0.83)	2.20 (0.68)	86.00[U]	.240	-0.43	-1.15	0.30
Creativity	1 (Baseline)	2.05 (1.39)	3.00 (0.99)	2.14	.042*	-0.79	-1.51	-0.02
	2 (Shuffled Scenes)	2.15 (1.66)	2.56 (1.08)	0.81	.428	-0.29	-1.00	0.43
	3 (Added Scenes)	2.69 (1.47)	3.68 (1.33)	1.95	.062[+]	-0.71	-1.42	0.05

* $p < .05$
[+] $p < .10$
[U] Mann-Whitney Test *U* statistic. All other reported statistics are independent samples T test.

4.7.3 Situations and Solutions

These results are displayed in Table 2.11. No results remained significant following Bonferroni correction.

There were no group differences in Fluency in either condition. In Condition 1 the control group had marginally higher Within-Category Fluency than the ASD group with a small observed effect size, whereas in Condition 2 the control group showed a trend towards increased Flexibility relative to the ASD group, increasing to a medium effect size.

No group differences in Originality of responses were found in either condition. However, the control group had significantly higher Usefulness scores relative to the ASD group in Condition 2 of the task. This was a large effect size. In both conditions, the control group scored higher on the variable of Creativity than the ASD group: this effect was significant in Condition 2 with a large effect size, but only marginally significant in Condition 1 with a medium effect size.

This also corresponded with significantly higher Best Chosen Response scores for the control group in both conditions. This effect size was large. However, there was no difference in Creativity Evaluation in either condition.

Table 2.11 Between group comparisons in the Situations and Solutions task.

Measure	Condition (description)	ASD Mean (SD)	Control Mean (SD)	Test statistic	*p*	Effect Size *d*	95% CI of Effect Size	
Fluency	1 (Everyday Items)	3.33 (1.05)	3.61 (1.31)	0.64	.528	-0.24	-0.95	0.49
	2 (Unusual Items)	4.02 (2.41)	3.80 (1.19)	96.00 ^{U}	.492	0.12	-0.6	0.83
Originality	1 (Everyday Items)	2.83 (0.54)	3.09 (0.55)	1.34	.190	-0.48	-1.19	0.26
	2 (Unusual Items)	3.08 (0.62)	3.55 (0.60)	77.50 ^{U}	.146	-0.77	-1.49	-0.01
Usefulness	1 (Everyday Items)	4.41 (0.68)	4.80 (0.59)	1.67	.107	-0.61	-1.33	0.14
	2 (Unusual Items)	3.84 (0.64)	4.35 (0.53)	2.35	.026*	-0.87	-1.59	-0.10
Creativity	1 (Everyday Items)	12.74 (3.01)	15.02 (3.13)	2.03	.052^{+}	-0.74	-1.46	0.02
	2 (Unusual Items)	12.76 (4.42)	16.60 (4.26)	2.42	.022*	-0.88	-1.61	-0.11
Best Chosen Response	1 (Everyday Items)	15.49 (3.58)	19.61 (4.59)	2.73	.011*	-1.00	-1.73	-0.22
	2 (Unusual Items)	17.36 (4.85)	21.81 (5.13)	2.44	.021*	-0.89	-1.62	-0.12

Measure	Condition (description)	ASD Mean (SD)	Control Mean (SD)	Test statistic	*p*	Effect Size *d*	95% CI of Effect Size	
Creativity Evaluation	1 (Everyday Items)	.93 (.08)	.90 (.11)	-0.77	.448	0.31	-0.42	1.02
	2 (Unusual Items)	.94 (.10)	.92 (.09)	91.50 [U]	.368	0.21	-0.51	0.92
Flexibility	1 (Everyday Items)	3.73 (1.29)	4.68 (1.75)	1.68	.104	-0.62	-1.33	0.13
	2 (Unusual Items)	4.20 (1.54)	5.36 (1.60)	2.02	.053[+]	-0.74	-1.46	0.02
Within-Category Fluency	1 (Everyday Items)	1.10 (0.18)	1.14 (0.13)	71.50 [U]	.079[+]	-0.25	-0.97	0.47
	2 (Unusual Items)	1.23 (0.27)	1.27 (0.17)	87.50 [U]	.299	-0.18	-0.89	0.54

* $p < .05$
[+] $p < .10$
[U] Mann-Whitney Test *U* statistic. All other reported statistics are independent samples T test.

5 Discussion

The primary aim of this study was to design and pilot a new measure of creativity for children with ASD that prioritised ecological validity and addressed several of the limitations of divergent thinking tasks. The first part of this discussion will focus on the psychometric properties of the three new tasks and then move to consider the group performance on the tasks at this preliminary stage of development. Methodological strengths and weaknesses are highlighted throughout before addressing overall limitations and making recommendations for next steps with task development.

5.1 Reliability

5.1.1 Inter-rater Reliability (IRR)

Both the Toy Improvement and Situations and Solutions tasks trialled a subjective rating method for the Originality and Usefulness scales. Responses for each item were collated across participants and scored as a set by both the main researcher and a second independent rater. Agreement between raters ranged from 'fair' for the Toy Improvement Originality scale to 'good' for the other three scales (Cicchetti, 1994).

The variation in IRR across scales could reflect multiple factors. Firstly, the Situations and Solutions task was scored after the Toy Improvement task, and therefore practice effects and familiarity with the scoring method could have helped align scores. Secondly, improved IRR in the Situations and Solutions task could be related to task design. As all participants were required to use the same set of eight objects in their answers, judgements about the relative creativity of one idea to another were easier to standardise than in the Toy Improvement task where no restrictions were imposed. Thirdly, the Usefulness scales may be inherently easier to rate than the Originality scales in the current tasks due to the clearly defined problem context, which allow for a more straightforward assessment of whether the

idea would be helpful or practical. Originality, on the other hand, perhaps requires a more nuanced or instinctive assessment: did the idea surprise or interest you?

Although the IRR on these scales was acceptable for the current study, future versions should focus on refining the scoring method to improve reliability. Further training and the development of a more detailed manualised coding system is required. Scoring should also ideally be carried out by three independent raters (Silvia et al., 2008). Due to time and resource constraints, the IRR of the other task variables was not assessed, for example Proportion of Original Units or Narrative Elaboration on the Story Scenes task. This would be helpful in future research to determine the reliability of the scoring method as a whole and refine the coding instructions.

The main researcher was involved in both the data collection and scoring and this inevitably influenced the process despite efforts to reduce bias. The method of data collection could also be streamlined to aid the scoring process. Responses were recorded by hand by the researcher and later electronically transcribed before scoring, and this approach potentially reduced the level of contextual detail and elaboration contained within responses. Finally, the Toy Improvement task design could mirror the Situations and Solutions task by providing participants with a set of objects which they can use to improve the toy.

5.1.2 Test-retest Reliability

The Total task scores were not significantly related at the first and second testing session. There are several possible explanations for this finding. A small proportion of participants (23%) participated in the retest and this limits the deductive power of the analysis. Furthermore, the time period between sessions (*M* = 74 days) increased the risk of confounding variables, environmental effects and participant changes that could impact upon test performance.

As participants demonstrated mixed results as to whether they did better or worse across tasks at the second administration, it is unlikely that there is a single explanation for the low consistency in scores. However, this is perhaps unsurprising given the complex interplay of trait, state, environmental and socio-cultural factors that can affect creativity at any one point in time (Czikszentmihalyi, 1998; Da Costa, Paez, Sanchez, Garaigordobil & Gondim, 2015). Furthermore, it is argued that low test-retest reliability is a necessary compromise for ecological validity, as these tasks rely on novelty which is diminished by repeat testing (Henry & Bettenay, 2010). Further replication with a larger sample and at differing time points is needed to clarify the retest reliability of the new tasks.

5.2 Validity

5.2.1 Criterion Validity

Correspondence between self-reported creative endeavours and the creativity task variables was inconsistent. As an aim of the study was to develop measures of creativity that corresponded with real-life 'Little C' creative behaviours, this finding is problematic, and criterion validity is therefore an area for improvement.

Self-reported Literary Achievements corresponded with the majority of the Story Scenes measures, providing some evidence of criterion validity in this task. However, Literary Activities did not correlate with the Story Scenes, and instead significantly related to performance on other tasks. This calls into question the domain specificity of the measures.

The ICAA Total Activities and Achievements scales correlated with two measures of the Story Scenes, suggesting multiple creative talents may be implicated in storytelling. Total creative activities and achievements was also significantly related to Flexibility in the Toy Improvement task. This could indicate that children who pursue numerous creative activities in diverse fields also have a

tendency to produce ideas belonging to numerous and diverse categories, providing preliminary evidence for Flexibility as a valuable construct in determining real-world creative behaviours.

5.2.2 Construct Validity

5.2.2.1 Convergent Validity

The PECC scale correlated significantly with only one measure, the number of ideas produced on the Toy Improvement task. The lack of convergent validity across measures is an oft reported problem in creativity research (Hocevar, 1981), with measures of the creative person having low correlation with measures of creative process or products (Cropley, 2000). The weak correspondence between parental attributions of creative characteristics and children's performance on the creativity tasks could indicate the potential biases that influence perceptions of creativity. However, as the PECC was validated in a non-clinical sample it is also possible that it does not adequately capture the creative characteristics and nature of children with ASD.

Aside from parental ratings, the ratings of an expert panel were used to evaluate a sample of stories produced by participants. The CAT involves obtaining reliable subjective ratings of products from appropriate judges and holds ecological validity as a measure of creativity as it mimics real-world evaluations of creative works (Amabile, 1982). There was excellent correlation between the ratings given by a panel of five experts in creative writing and the creativity scores on the Story Scenes task. Albeit for a small sample of stories, this demonstrates promising evidence that standardised creativity ratings can be developed that agree with the gold-standard, labour-intensive technique.

The CAT lends itself to domain specific tasks where appropriate experts in the field assess the creative products. This method is therefore difficult to apply to traditional DT tasks, which are based on the assumption that creative potential is

domain general (Plucker & Renzulli, 1998) and identification of experts is therefore less obvious. An area for future development could be to adapt the tasks so as to target particular domains; for example, setting practical problem scenarios relating to different school subjects in the Situations and Solutions task. It would then be possible to determine CAT and scale agreement for these measures. Using this method to determine construct validity appears a relevant route to explore in the development of ecologically valid tools of creativity in children with ASD, as arguably these consensual expert ratings have more 'real-world' applicability than self- or parent-report scales, for example in determining exam or career prospects in a particular domain (Diener & Wright, 2014).

5.2.2.2 Discriminant Validity

Few creativity task variables correlated with ecologically valid measures of Executive Functioning (EF) as assessed by the BADS-C and Eco-TED, providing some evidence of construct validity. However, scores on the BRIEF GEC were inversely related to several task variables measuring the quality of the creative product, indicating conceptual overlap that warrants exploration.

There is some evidence linking executive functions such as working memory and inhibition to creativity, as measured by subjective ratings in DT tasks (Benedek, Jauk, Sommer, Arendasy & Neubauer, 2014). In the current study, it is possible that the observed relationship between parent-reported behaviours of executive dysfunction and the child's quality of creative output is also mediated by higher-order executive functions. For example, in the Story Scenes task, narrative coherence could be conceptually linked to the ability to plan a response and hold the different story elements in mind to link them together cohesively. This would correspond to several items on the BRIEF, such as 'When given three things to do, remembers the first or the last' (Item 2) or 'Gets caught up in details and misses the big picture' (Item 28).

To tease apart these relationships it would be necessary to record more behavioural indices in the tasks, such as the length of time paused between receiving the start of the story and continuing the narrative. It was not feasible to report on all variables comprehensively in the current study; but it is perhaps unsurprising that multiple functions are implicated in the creativity tasks, given that ecologically valid measures do not aim to be 'pure' measures of a construct (Burgess et al., 2006).

FSIQ was found to be significantly related to several creativity variables, predominantly in the Story Scenes and Situations and Solutions tasks. This could mean that these tasks in particular are cognitively demanding, requiring reasoning and problem-solving abilities, or that performance is dependent upon a degree of prior knowledge or education level. It is also possible that in some cases scoring was implicitly biased towards children with a higher IQ and language proficiency; for example, longer stories using a sophisticated vocabulary receiving a higher Narrative Elaboration score. These ideas are speculative, and it would be necessary to collect additional variables, such as story length, to determine how a higher IQ influenced performance and control for the effects.

The relationship between intelligence and creativity has a complex and controversial history (Runco, 2014). The commonly held "threshold theory" posits that below a certain level of cognitive ability there is a correlation between IQ and creative potential; above it there is not (Guildford, 1967). The exact threshold may depend on the criterion of creativity adopted (Jauk et al., 2013). However, the evidence for threshold theory is disputed and a meta-analysis found only a small positive correlation between IQ and creativity (r = .174; Kim, 2005). This was moderated by age and type of creativity test rather than by different IQ thresholds.

It is notable that in the current study, the size of the reported correlations is larger than typically reported in the literature. Therefore, although the relationship between IQ and the creativity variables in this study may not be straightforward, it is

concerning. Given the significantly higher IQ of the control group than the children with ASD in this study, group differences should therefore be interpreted with caution. Further research with well-matched control and ASD groups is needed to disentangle the relationship between creativity and IQ and clarify what level of IQ threshold is operating in the current tasks, if any.

5.3 Group Performance on Creativity Tasks

A second aim of this study was to compare the performance of children with ASD relative to controls on the newly developed measures of creativity. For ease of discussion of results in the Toy Improvement and Situation and Solutions tasks, the variables are broadly grouped into the following categories: creative process variables (fluency, flexibility, within-category fluency); creative product (originality, usefulness, creativity); and creative comprehension (best chosen response, creativity evaluation).

As this was an initial exploratory analysis seeking to gain an understanding of performance on the numerous variables in the different task conditions, a large number of group comparisons were carried out. This increased the chance of Type I error. The relatively small sample size also reduced the statistical power of the tests. As aforementioned, the clinical and control groups were not matched on IQ and this is a further consideration when discussing the findings.

Overall, the interpretation of results at this stage is tentative and points to directions for future study as opposed to drawing conclusions of the nature of creativity in children with ASD.

5.3.1 Toy Improvement

This measure was adapted from a task commonly used in the Torrance Tests of Creative Thinking (TTCT; Torrance, 1974). Condition 1 replicated the traditional instructions, whereas in Conditions 2 and 3 participants were required to adapt the toy elephant for a particular function or context, respectively.

This task was liked the least by all participants; however ratings were acceptable and above the midpoint of the scale. Significant group differences were observed in Condition 1 only. The creativity of the best chosen response was higher for children without ASD and this was in line with this group producing significantly more original ideas on average. These measures were calculated independent of fluency and so are not simply an effect of the non-ASD children being more productive. Furthermore, fluency was found not to differ significantly between groups.

Interestingly, these group differences diminished in Conditions 2 and 3. Whilst the difference in originality scores was large in Condition 1, it became negligible in Conditions 2 and 3. The same pattern was observed in the creativity scores of the best chosen response. One interpretations of such findings, that could be formally tested in future using a larger sample, is that the scaffolding provided by the task instructions diminishes group differences in creativity.

Furthermore, from inspection of group means, scores on all creative product and creative comprehension variables appeared to improve in tandem for both groups, irrespective of diagnosis. This could suggest the modifications to the task helped all children to be more creative. As the number of ideas produced on average actually decreased in Conditions 2 and 3 relative to Condition 1 for both groups, it appears that sheer productivity is unable to account for the increased creativity. The mechanism for creativity – or answer to the question of *how* participants improved the quality of their ideas – is unclear from these preliminary results. However, these task modifications appear to provide a better platform from which to study this question.

The creativity evaluation scale is ecologically valid as a metric of creativity as it mirrors the process reported by creative individuals outside the lab, as described by Csikszentmihalyi (1998): the ability to know ideas that are good and pursue them, versus ideas that are bad and discard them. He argues that to do so

requires the internalisation of the rules and opinions of the socio-cultural domain in which one creates. Although not a significant finding, it is perhaps of note that in Condition 3 the ASD participants were better than controls at selecting their most creative ideas to improve the toy elephant in the given problem contexts: the garden; and the hospital.

It is possible that familiarity with a hospital setting could have meant the ASD participants had a better problem appreciation than control participants and were therefore better able to select their most original and appropriate responses. In future research it therefore might be useful to ask participants to explain the reasons for selecting their responses to better understand their level of problem appreciation and contextual understanding.

5.3.2 Story Scenes

In this task participants were asked to generate four stories using comic strip scene cards. In Condition 2 the cards were shuffled whereas in Condition 3 additional blank cards were added in.

In the baseline Condition 1, the control participants produced a significantly higher proportion of original content and created more coherent narratives than the children with ASD. This resulted in a higher creativity score overall for this condition. Stories told by control participants were also significantly more coherent in Condition 3, although this did not translate into significantly higher creativity scores. As discussed earlier, performance on these task variables was related to FSIQ and scores on the BRIEF, which might go some way to explain group differences. What is interesting, however, is that task manipulations appeared to reduce group differences in creativity even without controlling for these confounding variables.

The addition of blank scenes in Condition 3 appeared to encourage the children with ASD to produce more original story elements, so that only a small group difference was observed from the control participants, relative to a large

difference in the other conditions. Furthermore, there were no group differences found in narrative elaboration. Without overstating the results at this stage, one interpretation of the findings could be that children with ASD have potential for creative storytelling comparable to their peers, given the appropriate prompts and tools to develop their narratives. Further adaptations such as asking participants to write their stories rather than speak them aloud, or using an electronic interface, may help improve narrative coherence scores (Dillon & Underwood, 2012).

This task differed from the other two in that it was specific to the literary domain and did not resemble a DT task. This added complexity in defining variables that could map onto the relevant creative domains, and usefulness in particular. Although the design of domain specific tasks is merited in the interests of ecological validity, it does appear problematic for producing standardised scales of creativity that can be utilised across tasks and domains.

5.3.3 Situations and Solutions

In this task participants were required to generate creative solutions to everyday problems (Condition 1) and unusual problems (Condition 2) using a given set of objects. This task was the favourite overall for both groups of children.

The control group scored more highly on the creative product variables of usefulness and creativity. This effect was most pronounced in Condition 2 when faced with unusual scenarios. Interestingly, the creative processes of flexibility and within-category fluency were differentially used by the control group relative to the ASD group depending upon the condition. In Condition 2, the control group showed a tendency to use more object categories in their answers relative to children with ASD; whereas in Condition 1, there was a trend for children without ASD to generate a larger number of ideas using the same object. This was not simply a by-product of generating more ideas overall as fluency did not differ between groups.

This pattern of findings requires replication in a larger and well-matched sample to determine whether this is a genuine and generalizable between-group difference. The hypothesis to test could be that there are different optimal strategies for generating creative ideas depending upon the condition: for novel scenarios a more exploratory and inclusive approach is beneficial, jumping between and incorporating multiple categories; whereas for familiar scenarios where obvious ideas may spring to mind more readily for each object, it may be more beneficial to persevere within a particular category until these ideas are exhausted and creative uses emerge.

What could potentially differentiate performance in the children with and without ASD in this task, is not the preference for one strategy over another, but the ability to use these strategies flexibly to suit task demands. This would correspond with the dual pathway model of creative ideation (Nijstad et al., 2010). A tentative hypothesis could be that children with ASD do not have an impairment in either pathway (flexibility or persistence), but rather in knowing when one strategy might be preferential over another.

This would need to be elucidated by further research. It could then be possible to introduce task manipulations to attempt to induce the use of either pathway and assess the effect on creative output, for example, by reducing the number of objects available to use to solve the problem (encouraging use of the persistence pathway) or rather, limiting the number of times each object can be used (encouraging use of the flexibility pathway).

In both conditions of this task children without ASD nominated more creative ideas as their best responses. However, as both groups were just as good at evaluating the relative creativity of their own responses, it appears this group difference stems from the children with ASD producing less creative responses on average. Creative comprehension therefore appears to be comparable between

groups; this could be further investigated by asking participants why they selected their best response, as recommended in the Toy Improvement task.

5.4 Strengths and Limitations

A strength of this study lies in the design of tasks with varying conditions to allow comparison between groups *and* examination of the interaction between group and task condition. In this way it was possible to illustrate the interplay between external and internal factors influencing creativity. Another strength was in the steps taken to address several shortcomings of traditional DT tasks. In particular the design of tasks that are more representative of real-life creative contexts and therefore hold more ecological validity in the measurement of the creative process and products in children, with and without ASD.

A second advantage of this study is the development of a new scoring method and variables of creativity to more comprehensively assess the creative product and process, including not only the ideation phase but also the evaluation of ideas. Although inter-rater reliability should be improved for the originality and usefulness scales and assessed for other variables, certain scales showed promising construct and criterion validity by corresponding to the expert consensus opinion (CAT) and self-reported creative activities and achievements, respectively. These are arguably the most reliable and authentic benchmarks against which to compare a psychometric measure of creativity given their ecological validity, real-world applicability and generalisability (Amabile, 1982; Hocevar, 1981). This feature also makes the tasks more clinically useful as they provide a fuller assessment of the creative profile in ASD and therefore enable targeted adaptations and interventions.

A clear limitation to this study is the sample size, which limits the statistical power of the study and the generalisability of the results. A sample size of 40 participants in each group would have allowed sufficient power (.80) to detect a

medium effect size (r = 0.50) in correlational analyses of the psychometric properties of the measure, as well as detect large between group differences (d = .80) at the .01 level (Cohen, 1992). This more stringent alpha value is necessary due to the large number of comparisons carried out. Due to time and resource limitations it was not possible to recruit this desired number of participants and therefore extension of the current study with a larger sample is warranted.

Another main limitation of this study is that the clinical and control groups were not matched on IQ. Although the ASD group had an IQ within the normal range, the control group had an above-average IQ and were predominantly recruited from the same school with a mainly white middle-class demographic. Given the influence of these socio-cultural and personal variables on creativity (Csikszentmihalyi, 1998) it is important to have well matched groups to control for these confounding effects. However, a tension seems to exist between having a rigorously controlled and well matched experimental design, versus having a representative and diverse sample, reflecting the heterogeneity within the ASD population.

Finally, there are several recommendations for improving the creativity tasks borne out of the research process. Although pen and paper tasks are the mainstay of the psychometric tradition in creativity research, computerised tasks are favoured by children with ASD (Davis, Dautenhahn, Powell & Nehaniv, 2010) and would also streamline the data collection and scoring process. The task script and items could also be improved to provide more explicit information and instructions to be creative (Silvia et al., 2008), moving away from the traditional instruction in DT tasks to 'list as many ideas as you can' and instead prizing idea quality.

As a future consideration it would also be interesting to adapt the tasks for assessment in a group context; this may be particularly relevant for creative domains that rely on team interaction, such as drama and sports. These

suggestions could provide both useful and promising avenues to explore in the development of ecologically valid measures of creativity in children with ASD.

5.5 Clinical Implications

The findings of this exploratory investigation can be used to guide clinicians working with children with ASD and encourage the application of creative therapeutic tools within a clinical context. Creative problem-solving is a successful framework for interventions with children (Treffinger, 1995) and therapies such as CBT incorporate problem-solving approaches to enable adaptive coping in novel situations. The preliminary results of this study may indicate that given the appropriate framework and support, children with ASD can engage in these creative approaches in therapy.

When working in this way with children with ASD, it would appear helpful to clearly define the problem context and parameters, provide explicit instructions and offer encouragement and praise to help children generate creative solutions independently. It may also be beneficial to provide concrete prompts and the 'building blocks' to help get started on tasks as was done in this study, for example, providing the start of the story in the Story Scenes task and the objects to use to solve the problems in the Situations and Solutions tasks. When generating ideas, children with ASD should be encouraged to evaluate the problem context and select their best ideas based on this information. Circular questioning can also be introduced to help children consider the optimal solution from other perspectives.

Comic strips and storytelling techniques are commonly used interventions in ASD, for example Social Stories (Gray, 1998). The format of the Story Scenes task could be used as a template to create specific stories tailored to the individual's particular problem or social situation. This could be done collaboratively between the clinician and child by providing the initial story outline or scenes and then asking

the child to elaborate upon this or fill in the missing pictures to create alternative endings.

5.6 Conclusion

This paper reports on the development and piloting of three new ecologically valid tasks of creativity in children with ASD. The tasks were designed to address limitations of the most widely used psychometric measures of creativity, divergent thinking tasks, by mirroring real-life creative contexts and developing a new subjective scoring method to more fully assess the profile of creativity in ASD.

Overall the tasks were received favourably by the child participants. Test-retest reliability was low and although interrater reliability was satisfactory, there remains scope for improvement. Criterion validity was variable but best evidenced in the Story Scenes and Situations and Solutions tasks, corresponding to literary activities and achievements. Correspondence with expert panel ratings in the Story Scenes task lends support to the construct validity of the task; however, this is diminished by the significant relationship with both IQ and EF. Although preliminary, between-group comparisons indicated that modifying task conditions had some impact on the creative process and product. However, these findings require replication in a larger sample with a control group matched for IQ.

This research project reflects the first cycle of a creative process in task development. Although development of ecologically valid measures is increasingly prioritised in ASD research (Kenworthy et al., 2008), the real-world value and utility of the new creativity measures must ultimately be determined by the experts in the field: the ASD population they serve.

6 References

Amabile, T. M. (1982). A consensual assessment technique. *Journal of personality and social psychology*, *43*, 997-1013.

Amabile, T. M. (2014). Big C, Little C, Howard, and Me: Approaches to Understanding Creativity. In H. E. Gardner, M. L. Kornhaber & E. Winner (Eds.), *Mind, Work, and Life: A Festschrift on the Occasion of Howard Gardner's 70th Birthday, Volume 1*, (pp. 5–25). Cambridge, MA: CreateSpace Independent Publishing Platform.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Baer, J., & McKool, S. S. (2009). Assessing creativity using the consensual assessment technique. In C. S. Schreiner (Ed.), *Handbook of research on assessment technologies, methods, and applications in higher education* (pp. 65-77). London: IGI Global.

Baron, I. S. (2007). Review of Behavioural Assessment of the Dysexecutive Syndrome for Children (BADS-C). *Child Neuropsychology*, *13*(6), 539–542.

Barron, F. (1955). The disposition toward originality. *The Journal of Abnormal and Social Psychology*, *51*(3), 478.

Benedek, M., Jauk, E., Sommer, M., Arendasy, M., Neubauer, A. (2014b). Intelligence, creativity, and cognitive control: The common and differential involvement of executive functions in intelligence and creativity. *Intelligence*, 46, 73–83.

Best, C., Arora, S., Porter, F., & Doherty, M. (2015). The relationship between subthreshold autistic traits, ambiguous figure perception and divergent thinking. *Journal of autism and developmental disorders*, *45*(12), 4064-4073.

Bristow, E. (2016) *Developing and Piloting a New Measure of Executive Functioning for Children with Autism Spectrum Disorder (ASD).* Doctoral Thesis (D.Clin.Psy), University College London.

Burgess, P. W., Alderman, N., Forbes, C., Costello, A., LAURE, M. C., Dawson, D. R., ... & Channon, S. (2006). The case for the development and use of "ecologically valid" measures of executive function in experimental and clinical neuropsychology. *Journal of the international neuropsychological society*, *12*(02), 194-209.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment*, *6*(4), 284.

Cicchetti, D. V., & Sparrow, S. S. (1990). Assessment of adaptive behaviour in young children. In J. J. Johnson and J. Goldman (Eds.), *Developmental assessment in clinical child psychology: A handbook* (pp. 173-196). New York: Pergamon Press.

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155.

Craig, J., & Baron-Cohen, S. (1999). Creativity and imagination in autism and Asperger syndrome. *Journal of Autism and Developmental Disorders*, *29*(4), 319-326.

Cramond, B., Matthews-Morgan, J., Bandalos, D., & Zuo, L. (2005). A report on the 40-year follow-up of the Torrance Tests of Creative Thinking: Alive and well in the new millennium. *Gifted Child Quarterly*, *49*(4), 283-291.

Cropley, A. J. (2000). Defining and measuring creativity: Are creativity tests worth using?. *Roeper review*, *23*(2), 72-79.

Csikszentmihalyi, M. (1998) Implication of a Systems Perspective for the Study of Creativity. In R. J. Sternberg (Ed.), *Handbook of Creativity* [Online] (pp. 313-336). Cambridge: Cambridge University Press. Retrieved from: Cambridge Books Online http://dx.doi.org/10.1017/CBO9780511807916.018

Da Costa, S., Páez, D., Sánchez, F., Garaigordobil, M., & Gondim, S. (2015). Personal factors of creativity: A second order meta-analysis. *Revista de Psicología del Trabajo y de las Organizaciones*, *31*(3), 165-173.

Davis, M., Dautenhahn, K., Powell, S., & Nehaniv, C. (2010). Guidelines for researchers and practitioners designing software and software trials for children with autism. *Journal of Assistive Technologies*, *4*(1), 38-48.

Demetriou, E. A., Lampit, A., Quintana, D. S., Naismith, S. L., Song, Y. J. C., Pye, J. E., ... & Guastella, A. J. (2017). Autism spectrum disorders: a meta-analysis of executive function. *Molecular psychiatry*.

Diener, M. L., Wright, C. A., Smith, K. N., & Wright, S. D. (2014). Assessing Visual-Spatial Creativity in Youth on the Autism Spectrum. *Creativity Research Journal, 26*(3), 328-337.

Drevdahl, J. E. (1956). Factors of importance for creativity. *Journal of Clinical Psychology*, *12*(1), 21-26.

Dillon, G., & Underwood, J. (2012). Computer mediated imaginative storytelling in children with autism. *International Journal of Human-Computer Studies*, *70*(2), 169-178.

Emslie, H., Wilson, FC, Burden, V., Nimmo-Smith, I., & Wilson, BA (2003). Behavioural Assessment of the Dysexecutive Syndrome for Children (BADS-C). London: Harcourt Assessment/The Psychological Corporation.

Engel-Yeger, B., Josman, N., & Rosenblum, S. (2009). Behavioural Assessment of the Dysexecutive Syndrome for Children (BADS-C): An examination of construct validity. *Neuropsychological Rehabilitation*, *19*(5), 662-676.

Field, A. (2009). *Discovering statistics using SPSS*. London: Sage publications.

Fitzgerald, M. (2004). *Autism and creativity: Is there a link between autism in men and exceptional ability?* Hove, UK: Routledge.

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). Test review behavior rating inventory of executive function. *Child Neuropsychology*, *6*(3), 235-238.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 38*(5), 581–586.

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*(11), 1337-1345.

Goodman, A., & Goodman, R. (2009). Strengths and difficulties questionnaire as a dimensional measure of child mental health. *Journal of the American Academy of Child and Adolescent Psychiatry*, *48*(4), 400–3.

Gray, C. A. (1998). Social stories and comic strip conversations with students with Asperger syndrome and high-functioning autism. In E. Schopler, G. B. Mesibov, & L. J. Kunce (Eds.), *Asperger syndrome or high-functioning autism?* (pp. 167-198). Boston, MA: Springer.

Grayson, D. (2004). Some myths and legends in quantitative psychology. *Understanding Statistics, 3*(1), 101–134.

Guilford, J. P. (1967). *The nature of human intelligence.* New York, NY, US: McGraw-Hill.

Henry, L. A., & Bettenay, C. (2010). The assessment of executive functioning in children. *Child and Adolescent Mental Health*, *15*(2), 110-119.

Hill, E. L., & Bird, C. M. (2006). Executive processes in Asperger syndrome: Patterns of performance in a multiple case series. *Neuropsychologia*, *44*(14), 2822–2835.

Hocevar, D. (1981). Measurement of creativity: Review and critique. *Journal of Personality assessment*, *45*(5), 450-464.

Homack, S. R., & Reynolds, C. R. (2007). *Essentials of assessment with brief intelligence tests.* Hoboken, NJ: John Wiley & Sons, Inc.

Irby, S. M., & Floyd, R. G. (2013). Test Review: Wechsler Abbreviated Scale of Intelligence. *Canadian Journal of School Psychology*, *28*(3), 295–299.

Jauk, E., Benedek, M., Dunst, B., & Neubauer, A. C. (2013). The relationship between intelligence and creativity: New support for the threshold hypothesis by means of empirical breakpoint detection. *Intelligence*, *41*(4), 212-221.

Jauk, E., Benedek, M., & Neubauer, A. C. (2014). The road to creative achievement: A latent variable model of ability and personality predictors. *European journal of personality*, *28*(1), 95-105.

Kenworthy, L., Yerys, B. E., Anthony, L. G., & Wallace, G. L. (2008). Understanding executive control in autism spectrum disorders in the lab and in the real world. *Neuropsychology review*, *18*(4), 320-338.

Kim, K. H. (2005). Can only intelligent people be creative? A meta-analysis. *Journal of Secondary Gifted Education*, *16*(2-3), 57-66.

Ledger-Hardy, L. (2017). *Developing and Piloting a New Parent-Report Measure of Executive Function for Children with Autism Spectrum Disorder (ASD).* Doctoral thesis (D.Clin.Psy), University College London.

Liu, M. J., Shih, W. L., & Ma, L. Y. (2011). Are children with Asperger syndrome creative in divergent thinking and feeling? A brief report. *Research in Autism Spectrum Disorders*, *5*(1), 294-298.

Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., & Schopler, E. (1989). Autism diagnostic observation schedule: A standardized observation of communicative and social behavior. *Journal of autism and developmental disorders*, *19*(2), 185-212.

Mayer, R. E. (1999). Fifty Years of Creativity Research. In R. J. Sternberg (Ed.), *Handbook of Creativity* [Online] (pp. 449-460). Cambridge: Cambridge University Press. Retrieved from: Cambridge Books Online http://dx.doi.org/10.1017/CBO9780511807916.024

Mednick, S. (1962). The associative basis of the creative process. *Psychological review*, *69*(3), 220.

Mumford, M. D., Baughman, W. A., Threlfall, K. V., Supinski, E. P., & Costanza, D. P. (1996). Process-based measures of creative problem-solving skills: I. Problem construction. *Creativity Research Journal, 9*(1), 63-76.

Nijstad, B. A., De Dreu, C. K., Rietzschel, E. F., & Baas, M. (2010). The dual pathway to creativity model: Creative ideation as a function of flexibility and persistence. *European Review of Social Psychology*, *21*(1), 34-77.

Nijstad, B. A., Stroebe, W., & Lodewijkx, H. F. (2002). Cognitive stimulation and interference in groups: Exposure effects in an idea generation task. *Journal of experimental social psychology*, *38*(6), 535-544.

Norbury, C. F., & Bishop, D. V. (2003). Narrative skills of children with communication impairments. *International Journal of Language & Communication Disorders*, *38*(3), 287-313.

Okuda, S. M., Runco, M. A., & Berger, D. E. (1991). Creativity and the finding and solving of real-world problems. *Journal of Psychoeducational assessment*, *9*(1), 45-53.

Plucker, J. A., & Renzulli, J. S. (1998) Psychometric Approaches to the Study of Human Creativity. In R. J. Sternberg (Ed.), *Handbook of Creativity* [Online] (pp. 35-61). Cambridge: Cambridge University Press. Retrieved from: Cambridge Books Online http://dx.doi.org/10.1017/CBO9780511807916.005

Pullinger, J. (2017). *Developing and piloting an ecologically valid measure of executive function for children with autism: a function-led approach.* Doctoral Thesis (D.Clin.Psy), University College London.

Rietzschel, E. F., Nijstad, B. A., & Stroebe, W. (2007). Relative accessibility of domain knowledge and creativity: The effects of knowledge activation on the quantity and originality of generated ideas. *Journal of experimental social psychology*, *43*(6), 933-946.

Runco, M. A. (2014). *Creativity: Theories and Themes: Research, Development, and Practice* (Second Edition). San Diego: Academic Press. Retrieved from https://doi.org/10.1016/B978-0-12-410512-6.00001-1.

Runco, M. A., & Bahleda, M. D. (1986). Implicit theories of artistic, scientific, and everyday creativity. *The Journal of Creative Behavior*, *20*(2), 93-98.

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal*, *24*(1), 92-96.

Runco, M. A., Johnson, D. J., & Bear, P. K.(1993). Parents' and teachers' implicit theories of children's creativity. *Child Study Journal, 23*, 91–113.

Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., … & Richard, C. A. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts*, *2*(2), 68.

Skuse, D. H., James, R. S., Bishop, D. V., Coppin, B., Dalton, P., Aamodt-Leeper, G., ... & Jacobs, P. A. (1997). Evidence from Turner's syndrome of an imprinted X-linked locus affecting cognitive function. *Nature, 387,* 705-708.

Skuse, D. H., Mandy, W. P., & Scourfield, J. (2005). Measuring autistic traits: heritability, reliability and validity of the Social and Communication Disorders Checklist. *The British Journal of Psychiatry*, *187*(6), 568-572.

Skuse, D., Warrington, R., Bishop, D., Chowdhury, U., Lau, J., Mandy, W., & Place, M. (2004). The developmental, dimensional and diagnostic interview (3di): A novel computerized assessment for autism spectrum disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *43*(5), 548-558.

Simonton, D. K. (2017) Big-C versus little-c creativity: Definitions, implications, and inherent educational contradictions. In A. R. Beghetto, & B. Sriraman (Eds.), *Creative Contradictions in Education* (pp. 3-19). Switzerland: Springer International Publishing.

Stein, M. I. (1953). Creativity and culture. *The journal of psychology*, *36*(2), 311-322.

Sternberg, R. J., & Lubart, T. I. (1996). Investing in creativity. *American psychologist*, *51*(7), 677.

Torrance, E. P. (1974). *The Torrance Tests of Creative Thinking-Norms-Technical Manual Research Edition-Verbal Tests, Forms A and B- Figural Tests, Forms A and B.* Princeton, NJ: Personnel Press.

Torrance, E. P. (1981). Predicting the Creativity of Elementary School Children (1958-80)—and the Teacher Who" Made a Difference". *Gifted Child Quarterly*, *25*(2), 55-62.

Treffinger, D. J. (1995). Creative problem solving: Overview and educational implications. *Educational Psychology Review*, *7*(3), 301-312.

Turner, M. A. (1999). Generating Novel Ideas: Fluency Performance in High-functioning and Learning Disabled Individuals with Autism. Journal of Child Psychology and Psychiatry, 40(2), 189-201.

Wallach, M. A. (1976). Tests tell us little about talent: Although measures of academic skills are widely used to determine access to contested educational opportunities, especially in their upper ranges they lack utility for predicting professional achievement. *American Scientist*, *64*(1), 57-63.

Wallach, M. A., & Kogan, N. (1965). *Modes of thinking in young children: A study of the creativity-intelligence distinction.* New York: Holt, Rinehart, & Winston.

Wallach, M. A., & Wing, C. W., Jr. (1969). *The talented student: A validation of the creativity intelligence distinction.* New York: Holt, Rinehart & Winston.

Weisberg, R. (1986). *Creativity: Genius and other myths*. New York: Freeman and Co.

Weisberg, R. W. (2006). Modes of expertise in creative thinking: Evidence from case studies. In K. A. Ericsson, N. Charness, P. J Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance*, (pp. 761-787). New York: Cambridge University Press.

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence - Second Edition.* Pearson.

White, S. J., Burgess, P. W., & Hill, E. L. (2009). Impairments on “open-ended” executive function tests in autism. *Autism Research*, *2*(3), 138-147.

White, S., Hill, E., Happé, F., & Frith, U. (2009). Revisiting the strange stories: Revealing mentalizing impairments in autism. *Child development*, *80*(4), 1097-1117.

Wing, L. (1981). Asperger's syndrome: a clinical account. *Psychological medicine*, *11*(01), 115-129.

Wing, L., & Gould, J. (1979). Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification. *Journal of autism and developmental disorders*, *9*(1), 11-29.

Zeng, L., Proctor, R. W., & Salvendy, G. (2011). Can traditional divergent thinking tests be trusted in measuring and predicting real-world creativity?. *Creativity Research Journal*, *23*(1), 24-37.

Part 3: Critical Appraisal

1 Introduction

This paper provides a critical appraisal of the empirical research undertaken as part of the Doctorate in Clinical Psychology. I initially expand upon the limitations already discussed in Part 2 before offering qualitative reflections about the process of developing the creativity tasks and scoring method. I then discuss more conceptual themes that emerged through exploring the field of creativity and share personal insights and developments that occurred by engaging in the research process.

2 Sampling Limitations

2.1 Background and Personal Characteristics

As discussed in the empirical study (Part 2 Discussion) the groups were not matched on IQ. There are numerous factors that can affect creativity. On the level of the person, the relationship with IQ is debated (Kim, 2005). However, it seems inappropriate to consider these variables on an individual level without situating the creative 'person' within their broader socio-cultural context, considering also the interaction of a number of personal and environmental variables in the expression of creativity.

I found the integrated "systems perspective" of creativity proposed by Csikszentmihalyi (1998) helpful when reflecting on the sample characteristics in this study. I thought not only of how IQ might have affected task performance directly, but how a higher than average IQ might come attached to other social factors that can influence creativity. Csikszentmihalyi refers to the idea of "cultural capital", including factors such as socio-economic status, parental education and interest in creative domains, which can bestow an individual with advantages to realise their creative potential.

We did not control for these variables in this study, and nor would it be practical or realistic to control for every potential confounding variable in the study of creativity. Furthermore this would reduce ecological validity and generalisability of results. However, on a broader scale it is important to consider how an individual may possess creative potential and yet never be recognised as creative due to the inability to translate this into a tangible product, perhaps due to familial or social factors that inhibit this expression.

Creative achievement requires sufficient opportunity to actualise creative potential and relies on more than individual creativity alone (Cropley, 2000). This might also go some way to explaining the variable correlations between the creativity tasks and the Inventory of Creative Activities and Achievement scale, as creative thinking must also be combined with motivation, economic resources and social opportunities to be turned into a creative behaviour. This complicates the picture and cannot be portrayed by a linear relationship.

2.2 Recruitment and Sample Size

Recruitment and data collection procedures were considerable determinants in limiting the sample size. Recruitment of new ASD participants was carried out with a research assistant based at the specialist social and communication disorders clinic and this was a valuable and helpful link with families and children. The plan was also for the research assistant to help with data collection and we developed a training programme whereby she observed me delivering the test battery on two occasions and then I observed her carrying out a testing session.

Unfortunately, due to scheduling demands, the research assistant was only able to test two participants independently. I therefore carried out the majority of the data collection, across a total of 35 testing sessions. This was beneficial in several ways in that it ensured consistency and standardisation across test sessions; gave me greater experience with the practical matters of task administration and

recording; and greater exposure to the clinical child population. However, given the length of the test battery it also placed considerable demands on my personal resources and data collection inevitably became an enormous and consuming task.

A large proportion of time was also spent in the coordination of research sessions with parents, which required flexibility in arranging visits at home and school in the evenings and at weekends. Mirroring my clinical experiences during my training placement in a CAMHS setting, I was reminded of the importance of working within the family system to engage families in clinical or research work, considering the numerous roles within a system and competing demands for time and resources, as well as the differing beliefs about creativity, ASD, research and the NHS more generally that could affect participation (Gross & Goldin, 2008; Reder & Fredman, 1996; Stateva et al., 2012). I am incredibly grateful to the families that participated and the time and effort they contributed to the study.

3 Administration and Scoring

3.1 Computerised Tasks

Designing computer-mediated creativity tasks could be advantageous for several reasons. Children with ASD are motivated by the use of computers (Dillon & Underwood, 2012) and this would also increase ecological validity of the tasks due to the familiarity and popularity of technology with children in this age group. It would also maximise efficiency of the administration and scoring process, allowing data from a larger sample to be collected more easily and reducing problems with experimenter blinding if tasks were completed remotely.

As this was a preliminary study, it was not feasible to take this step in task development. However, there may also be several limitations in changing the administration of the tasks. Carrying out the data collection face to face highlighted the quality of this human interaction versus a screen interface. Over

the course of data collection, several ideas and responses generated by children in the tasks stood out in my mind. Whilst it is possible that certain personal features of the children could have biased my opinion, what these ideas held in common was not my response to the child, but my response to the idea: surprise!

This element of surprise, or something that generates interest or curiosity, appears to be a defining qualitative characteristic of originality (Bruner, 1962; Simonton, 2012). This may be a defining interpersonal feature of creativity. Csikszentmihalyi (1998) suggests "whether an idea or product is creative or not does not depend on its own qualities, but on the effect it is able to produce in others who are exposed to it… what we call creativity is a phenomenon that is constructed through an *interaction between producer and audience*." (p. 314). I therefore wonder whether something important in the measurement of creativity would be lost by removing the interpersonal element to the testing environment and it would seem important to preserve this condition even with the addition of technology.

3.2 Development of a Subjective Scoring Method

Potentially the major dilemma faced in the development of these measures was to define and condense the variables of interest. It felt difficult to strike a balance between inclusivity and practicality in an exploratory investigation and I was cautious to narrow the focus too soon given that this was a measure development study.

Producing too many variables would increase the chance of Type I error and over-burden the scoring process; whereas too few could risk Type II error and not doing justice to the study aims. My aim was to collect data that could more comprehensively assess the profile of creativity in ASD; however, there were

many angles from which this task could be approached and little extant literature of creativity and ASD in which to ground these decisions.

Furthermore, based on variation in creativity scales and reporting methods in the literature (see Part 1 Literature Review), it felt helpful to replicate some of the variables most frequently used in Divergent Thinking tasks (Fluency, Flexibility) to allow comparison across studies. However, other scales (Originality, Usefulness) were newly developed and used a subjective rating method, as these have advantages over traditional scoring methods (Silvia et al., 2008) and adopt a more ecologically valid approach.

Using such an approach in practice required considerable training and time. Furthermore, it relied on the inclusion of several raters, which can increase the cost of research. As all ratings are made *relative to the item and response set,* scores are not transferable across the task or indeed study samples, which can limit extension and replicability of findings. It is also not possible to extract a criterion or norm-referenced manual as is commonly used in neuropsychological tasks (Levin, 1994)**.**

However, whilst a simpler scoring system may be preferable for research purposes, this does not make it more accurate, helpful or valid. Furthermore, although traditional DT scoring methods profess objectivity, a degree of interpretation is always required by the rater, for example in deciding whether a response is innovative or incorrect (Bishop & Norbury, 2005).

Creativity is necessarily context dependent; therefore it is not a fixed construct across time and place (Lubart & Sternberg, 1998). If hallmarks of creativity are flexibility, openness to experience and being able to adapt a product for the demands and values of the audience (Csikszentmihalyi, 1998), then a tool measuring creativity must do the same and cannot presume a one-size-fits-all approach. This does not mean that rating scales cannot be standardised; but rather, they should be interpreted meaningfully in the given task context.

As a further assessment of construct validity it may be helpful to compare the newly developed subjective rating methods against the traditional norm-referenced or uniqueness rating scales in further research with the creativity tasks.

4 Attitudes Towards Creativity

During the research process I encountered different views about what it means to be creative and *who* can be creative. These beliefs and assumptions have been observed in child participants and parents, in friends and colleagues, and also in myself. It is important to understand these implicit theories as they form the basis of information processing biases and stereotypes; and can be used to inform the planning and evaluation of efforts to foster creativity (Plucker & Renzulli, 1998).

4.1 Research as a Creative Process

On a personal level, I had not previously been involved with neuropsychological measure development nor worked with an ASD child population prior to training. This resulted in a steep learning curve and initially left me doubting my capacity to produce something both original and useful when I felt neither the area of research nor clinical population was an area of expertise.

In many ways, the research task at hand mirrored the research question of how to produce something creative and evaluate its potential. As the project progressed, I increasingly noticed the evidence of what I was reading in the creativity literature in my first-hand experience of the research process. The research process moved in parallel with a creative process from idea generation, to evaluation, through implementation and finally analysis (Zeng, Proctor & Salvendy, 2011). Although prior to this project I would not have described myself as a creative person; the process of developing a measure of creativity has undoubtedly been a creative endeavour.

4.2 Children's Self-Perceptions of Creativity

A common theme that emerged over the course of the research is that one must be good at something to be considered creative, and level of achievement is a direct inference of creativity. For example, when completing the Measure Acceptability Scale participants frequently remarked "I wasn't very good at that and so I didn't like it as much". Participants received praise and encouragement for every idea expressed in the creativity tasks and therefore these self-evaluations are likely to be influenced by other factors. Furthermore, in filling out the Inventory of Creative Activities and Achievements, participants often self-nominated their top creative achievements as any pursuit in which they had received the most accolades, for example, being a member of the football team.

Whilst this provides anecdotal evidence that creativity is a socially validated construct (in that one cannot be recognised as creative without the verification of external judges e.g. Amabile, 1982; Csikszentmihalyi, 1998) it also portrays how linked these judgements are to self-evaluation and enjoyment of a creative activity. It also demonstrates how creativity can be conflated with the idea of expertise. The implication is that individuals are unlikely to pursue creative domains where they do not receive positive feedback or 'high marks', which in turn reduces their experience and practice within a domain and diminishes the opportunities to develop a creative interest into a creative achievement (Simonton, 2014).

4.3 Parents' Implicit Views

These views can also be reinforced or challenged by parents. Creativity is implicitly associated with positive characteristics in children by parents and teachers and these ideas form standards against which to evaluate children's performances and behaviours (Runco, Johnson & Bear, 1993). They are likely also to influence expectations and influence parental behaviour towards the child, which can in turn facilitate or inhibit creativity.

The parents I met over the course of data collection held a variety of views about their children's creativity and it would be unfair to take a parsimonious stance and categorise these into ASD versus non-ASD group beliefs. However, it is notable that words such as 'imaginative' and 'playful' are frequently ascribed as creative characteristics in children (see Appendix G for items on the PECC) yet are attributed as deficits in ASD. This could bias parent's expectations of their child's abilities or force them to take a more defensive stance in asserting their child's creative strengths.

4.4 Deficit Narrative of ASD

The dominant narrative in research in ASD appears limitation rather than strength focused (Dinishak, 2016) and I wondered how this is filtered into the internalised discourses of consumers of research and the ASD community. This idea did not sit comfortably with me and at times I also questioned the clinical utility of having the creativity tasks sit within a broader battery of tests designed to capture 'everyday problems' and 'executive dysfunction'. This perhaps also influenced the angle I took throughout this research in pointing out the limitations with existing measures and methods before pointing to any limitations within individuals or attempting to make generalisations about a group as a whole.

Certainly what has struck me through this research is the variation in performance in all the children who participated, irrespective of diagnosis. ASD is often characterised by the within-group heterogeneity and further exploration of individual differences and the factors that mediate this could be a more fruitful avenue of research than focusing on group comparisons that mask individual differences (Shallice & Evans, 1978). This would require a shift in the dominant research paradigm (Towgood, Meuwese, Gilbert, Turner & Burgess, 2009).

4.5 Creativity in the School System

On a broader level, I have also reflected on the expectations enforced by the education system and wondered whether this encourages or inhibits creativity in children. The classroom environment is fundamental in engaging or suppressing creativity (Beghetto & Kaufman, 2014) and the emphasis in many schools is on academic attainment and conformity of thinking, as opposed to innovation and breaking of conventional rules.

As worded by Sir Ken Robinson (2006) during a seminal talk about creativity and the school systems, “We don't grow into creativity, we grow out of it. Or rather, we get educated out of it.” The importance of creativity is widely acknowledged across domains on an individual, academic, organisational and societal level. I believe that the value in developing measures that can reliably and validly measure creativity in children therefore lies in their potential to assess influences upon creativity and factors that can limit or realise creative potential, to help concentrate efforts on fostering creativity.

5 Challenges and Future Directions of Research

In designing the tasks I gave much thought to the optimal conditions that would promote creativity in children, for example in the design of tasks that are untimed, fun and game-like (Wallach & Kogan, 1965). Outside of a test setting it can be difficult to actualise these ideal conditions and they were perhaps hardest to replicate in my own research environment where there were many internal and external pressures, time and resource constraints, evaluations and competing demands that influenced the creative process. I therefore found the clinical element of the data collection both a rewarding and helpful part of the research process as it reintroduced the necessary playful component and refreshed my creative thinking.

Task development was not a linear process but iterative and cyclical. However, at some point a decision needed to be made to move to next the stage

rather than revising current ideas and formalise the tasks. Upon reflection, there are several elements I would have changed about the tasks. However, this knowledge was gained by progressing through the various stages of the research and the wealth of insight gained through data collection. This speaks to finding the balance in experiential versus conceptual knowledge during measure design: or taking an ecologically valid versus a top-down approach.

Due to the paucity of research in creativity and ASD (see Part 1 Literature Review), there was not a clear behavioural template upon which to build the tasks from the bottom-up, as had been the method of task development in the earlier Eco-TED tasks (Bristow, 2016; Pullinger, 2017). One method of task development was therefore to adapt existing creativity measures and attempt to translate these into a real-world context, as with the Toy Improvement task. The other two tasks used measures within the Eco-TED as the starting point. The Story Scenes task was based upon the earlier Storytelling task (Pullinger, 2017) and the Situations and Solutions task took parent-reported descriptions of common behavioural problems in children with ASD (Ledger-Hardy, 2017) and presented these as problems requiring a creative problem-solving approach.

Other tasks were thought of in the initial ideation phase of the project that resembled real-life creative tasks and contexts but later had to be abandoned due to the difficulties in standardising scoring. For example, one idea was to ask participants to design a new app to help with a given problem; whilst this yielded interesting and enthusiastic responses during piloting, it was not practical to evaluate the data in a meaningful way due to a lack of domain knowledge. This speaks to the difficulty in devising a standardised scoring instrument for domain-specific measures.

An alternative method to the development of tasks and variables could be to use a phenomenological approach and carry out a preliminary investigation into the views and experiences of creativity in the ASD community, including children,

parents and teachers. This could involve asking questions about the ways that children with ASD use creativity in the school or home environment; the observed strengths and weaknesses in creativity; and how they would assess a creative product as *original* and *useful* in an everyday sense.

It would also be interesting to ask carers and teachers how they attempt to foster creativity in the home and classroom environment, with the aim of developing task instructions that mirror real life demands. Although this was beyond the scope of the current project, this could be both a useful and meaningful way to involve the ASD community in research and maximize ecological validity by prioritizing the voice of the participant and those with lived experience (Pellicano, Dinsmore & Charman, 2014).

6 References

Amabile, T. M. (1982). A consensual assessment technique. *Journal of personality and social psychology*, *43*, 997-1013.

Beghetto, R. A., & Kaufman, J. C. (2014). Classroom contexts for creativity. *High Ability Studies*, *25*(1), 53-69.

Bishop, D. V., & Norbury, C. F. (2005). Executive functions in children with communication impairments, in relation to autistic symptomatology: I: Generativity. *Autism*, *9*(1), 7-27.

Bristow, E. (2016) *Developing and Piloting a New Measure of Executive Functioning for Children with Autism Spectrum Disorder (ASD).* Doctoral Thesis (D.Clin.Psy), University College London.

Bruner, J. S. (1962). The conditions of creativity. In H. Gruber, G Terrell, & M. Wertheimer (Eds.), *Contemporary Approaches to Creative Thinking* (pp. 1-30). New York, NY: Atherton Press.

Cropley, A. J. (2000). Defining and measuring creativity: Are creativity tests worth using?. *Roeper review*, *23*(2), 72-79.

Csikszentmihalyi, M. (1998) Implication of a Systems Perspective for the Study of Creativity. In R. J. Sternberg (Ed.), *Handbook of Creativity* [Online] (pp. 313-336). Cambridge: Cambridge University Press. Retrieved from: Cambridge Books Online http://dx.doi.org/10.1017/CBO9780511807916.018

Dillon, G., & Underwood, J. (2012). Computer mediated imaginative storytelling in children with autism. *International Journal of Human-Computer Studies*, *70*(2), 169-178.

Dinishak, J. (2016). The deficit view and its critics. *Disability Studies Quarterly*, *36*(4).

Gross, V., & Goldin, J. (2008). Dynamics and dilemmas in working with families in inpatient CAMH services. *Clinical Child Psychology and Psychiatry*, *13*(3), 449-461.

Kim, K. H. (2005). Can only intelligent people be creative? A meta-analysis. *Journal of Secondary Gifted Education*, *16*(2-3), 57-66.

Ledger-Hardy, L. (2017). *Developing and Piloting a New Parent-Report Measure of Executive Function for Children with Autism Spectrum Disorder (ASD).* Doctoral thesis (D.Clin.Psy), University College London.

Levin, H. S. (1994). A guide to clinical neuropsychological testing. *Archives of neurology*, *51*(9), 854-859.

Lubart, T. I., & Sternberg, R. J. (1998). Creativity across Time and Place: life span and cross-cultural perspectives. *High Ability Studies*, *9*(1), 59-74.

Pellicano, E., Dinsmore, A., & Charman, T. (2014). What should autism research focus upon? Community views and priorities from the United Kingdom. *Autism*, *18*(7), 756-770.

Plucker, J. A., & Renzulli, J. S. (1999). Psychometric approaches to the study of human creativity. In R. J. Sternberg (Ed.), *Handbook of Creativity* [Online] (pp. 35-61). Cambridge: Cambridge University Press. Retrieved from: Cambridge Books Online http://dx.doi.org/10.1017/CBO9780511807916.005

Pullinger, J. (2017). *A meta-analytic review of generativity in autistic spectrum disorders.* Doctoral thesis (D.Clin.Psy), University College London.

Reder, P., & Fredman, G. (1996). The relationship to help: Interacting beliefs about the treatment process. *Clinical Child Psychology and Psychiatry*, *1*, 457–467.

Robinson, K. (2006). *Do schools kill creativity?* [Video file]. Retrieved from https://www.ted.com/talks/ken_robinson_says_schools_kill_creativity/Runco, M. A., Johnson, D. J., & Bear, P. K. (1993). Parents' and teachers' implicit theories of children's creativity. *Child Study Journal, 23*, 91–113.Shallice, T.,

& Evans, M. E. (1978). The involvement of the frontal lobes in cognitive estimation. *Cortex*, *14*(2), 294-303.

Simonton, D. K. (2012). Taking the US Patent Office criteria seriously: A quantitative three-criterion creativity definition and its implications. *Creativity research journal*, *24*(2-3), 97-106.

Simonton, D. K. (2014). Creative performance, expertise acquisition, individual differences, and developmental antecedents: An integrative research agenda. *Intelligence*, *45*, 66-73.

Stateva, M., Minton, J., Beckett, C., Doolan, M., Ford, T., Kallitsoglou, A., & Scott, S. (2012). Challenges recruiting families with children at risk of anti-social behaviour into intervention trials: lessons from the Helping Children Achieve (HCA) study. *Journal of Children's Services*, *7*(4), 285-302.

Sternberg, R. J., & Lubart, T. I. (1996). Investing in creativity. *American psychologist*, *51*(7), 677.

Towgood, K. J., Meuwese, J. D., Gilbert, S. J., Turner, M. S., & Burgess, P. W. (2009). Advantages of the multiple case series approach to the study of cognitive deficits in autism spectrum disorder. *Neuropsychologia*, *47*(13), 2981-2988.

Wallach, M. A., & Kogan, N. (1965). *Modes of thinking in young children: A study of the creativity-intelligence distinction.* New York: Holt, Rinehart, & Winston.

Zeng, L., Proctor, R. W., & Salvendy, G. (2011). Can traditional divergent thinking tests be trusted in measuring and predicting real-world creativity?. *Creativity Research Journal*, *23*(1), 24-37.

Appendix A. Study Quality Evaluation Tool

Criteria		YES (2)	PARTIAL (1)	NO (0)
1	Question / objective clearly described?			
2	Study design evident and appropriate?			
3	Method of selection for clinical and control groups clearly explained and appropriate?			
4	Clinical and control group characteristics sufficiently described?			
5	Diagnostic and screening procedures for ASD detailed and appropriate for both groups? (For clinical group, DSM criteria plus at least one other measure.)			
6	Clinical and control groups matched on age, gender and FSIQ?			
7	Outcome measure well defined and dependent variable(s) clearly identified?			
8	Sample size appropriate/ sufficient power?			
9	Analytic methods described/justified and appropriate?			
10	Rater(s) blinded to diagnosis/ability?			
11	Mean and standard deviation reported for main results?			
12	Results reported in sufficient detail?			
13	Conclusions supported by the results?			

Appendix B. Task Script for Toy Improvement, Story Scenes and Situations and Solutions Tasks

Task Script
Version 5, Oct 2017

Toy Improvement Task

Materials:
- Stuffed toy elephant
- Stop watch

In this task, participants are asked to generate as many ideas as possible to improve a toy. In subsequent trials they are asked to improve the toy for a particular use.

The examiner gives the following instructions:

> *[Places toy in front of child].* ***This is a toy elephant. It is small and grey. You can have a look at it, touch it or pick it up if you like.***

Orientation trial:

> ***Can you tell me one other thing about this elephant that you can see?***

If child correctly names feature, move on to main task. If incorrect or no response, say:

> ***The elephant is fluffy.***

Then say:

> ***Let's play a game with the elephant.***

Main task

Item 1

> ***I would like you to tell me as many ways you can think of that would make this toy elephant more fun to play with. It can be anything you like, just tell me all the ways to make it better.***

Start timing once instructions have been read out. Give the participant plenty of encouragement. If the child is silent for 15 seconds, prompt:

> ***Those are good ideas. Can you think of anything else? It can be anything at all.***

Following 3 consecutive incorrect responses i.e. where the child appears to have misunderstood the instructions (e.g. listing descriptive features of the elephant without suggesting improvements, naming other animals), remind the child of the instructions. Do this only <u>once</u> per item.

1

Discontinue when the child indicates they have no other ideas or have been silent for 30 seconds following prompt. Then say:

> ***Well done. Which are your two best ideas?*** *[Wait for child response].*

Item 2

> ***Now I would like you to imagine that this toy is designed to help babies learn to walk. Can you tell me all the ways this toy could be made better for that? It can be anything you like, just tell me all the ways you can think of.***

Item 3

> ***Now let's imagine that this toy is designed for teachers to use in school. Can you tell me all the ways this toy could be made better for that? It can be anything you like, just tell me all the ways you can think of.***

Item 4

> ***Now let's imagine that this toy is designed for use in the garden. Can you tell me all the ways this toy could be made better for that? It can be anything you like, just tell me all the ways you can think of.***

Item 5

> ***Now let's imagine this toy is designed for use in hospital. Can you tell me all the ways this toy could be made better for that? It can be anything you like, just tell me all the ways you can think of.***

Complete the Measure Acceptability Scale at the end of the task.

Task Script
Version 5, Oct 2017

Story Scenes Task

Materials:

- Story Packs 1 and 2 laminated cards, incl blank cards (10x10cm size)
- Audio recorder
- Stopwatch

In this task participants are asked to tell a story using the story picture cards displayed in a given sequence. In subsequent trials the cards are placed in an alternating sequence and the story is lengthened using blank cards. Participants are required to tell a new story each time, incorporating all the story pictures in sequence. The beginning of the story is provided. The stories must be based on the story picture cards but can add in novel or imaginative elements not shown on the cards.

Item 1 (Story Pack 1)

Place cards from Story Pack 1 in front of participant in the order shown below. Place down each card one at a time in sequence from first to last scene.

Once cards placed in front of child, say:

> ***Here is a comic strip made up of different pictures. I would like you to tell me a story using the pictures. I will start the story*** *[point to first card]* ***and then you can carry it on. You should include all the pictures in order*** *[point to cards in order]* ***and the story should finish here*** *[point to last card].* ***The story must use the pictures but you can add in other things you can't see in the pictures too. Try to be as creative as you can! I am going to record the story using this*** *[show the audio recorder]* ***but this is not a timed task so take time if you need it.***
>
> ***OK, now listen to the start of the story. You can take some time to think about it.***
>
> ***There is a girl called Alex in her bedroom. It is her day off school. She has got dressed and is ready to start her day.***
>
> ***Now you carry on the story when you are ready…***

Start audio recorder once instructions read out. Record time taken to start story after instructions given . If no response after 45 seconds, give prompt:

This is quite tricky, isn't it? Would you like me to give you some help?

If child indicates they would like help, begin the prompting procedure below. Record the number of prompts given and the time taken to respond following each prompt. Discontinue task after 3 consecutive non-responses. If child indicates they would like more time, allow 30 more seconds and then begin to administer prompts.

1. ***Ok, let's start/look here.*** *[point to first picture]* ***. What can you see in the picture?***

   Following child's response say ***Great! Now see if you can carry on with the story, tell me what could be going on in the pictures.***

   [Allow 20 seconds without any response before moving to prompt 2.]

2. ***OK, I can see…*** *[insert description e.g. books on the floor].* ***What else can you tell me?***

   Following child's response say ***Great! Now see if you can carry on with the story, tell me what could be going on in the pictures.***

   [Allow 20 seconds without any response before moving to prompt 3.]

3. ***OK, I'll tell you what I think is going on…*** *[Insert story action e.g. Alex reads her books and then goes downstairs].*

   What do you think happens next? *[Point to next picture].*

   Following child's response say ***Great! Now see if you can carry on with the story, tell me what could be going on in the pictures.***

   If child does not respond after 20 seconds say, ***OK, this is a hard story. Let's finish here and move on.***

If during storytelling the child pauses for 20 seconds, say:

Well done. What happens next?

If the child does not respond within 20 seconds, offer to help and begin the prompting procedure as outlined above, starting at the relevant picture.

Terminate task when either a) discontinue rule met or b) five minutes have passed following start of task (i.e. start of recording following instructions). After 4:30 seconds, say:

> ***OK, I'm going to stop you for a moment. I can see you are trying very hard with this story but we don't have much time. Can you finish the story now so we can move onto the next part?***

Allow 30 more seconds and then say:

> ***Let's finish there. Well done!***

Item 2 (Story Pack 1)

Collect the cards and shuffle them, then place back down in the new order show below.

Once cards placed in front of child, say:

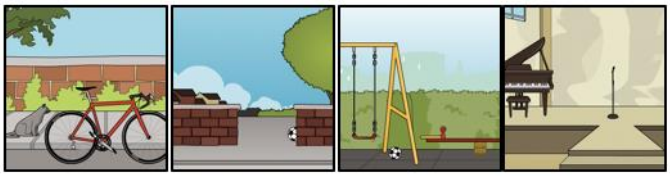
> ***Now here is a new comic strip. It is made of the same pictures, but in a different order. Please tell me a different story using these pictures. I will start the story again*** *[point to picture on the left]* ***and then you carry it on, finishing here*** *[point to picture on right].* ***The story must use the pictures but you can add in other things you can't see in the pictures too. Try to be as creative as you can! I will record it like last time.***
>
> *[Check understanding of instructions before proceeding.]*
>
> ***OK, now listen to the start of the story. You can take some time to think about it.***
>
> ***There is a boy called James. He is at his friend Olly's house. They have done their homework and now are playing a game.***
>
> ***Now you carry on the story when you are ready…***

Start audio recorder once instructions read out. Record time taken to start story after instructions given.

If no response after 45 seconds, administer prompts as described above. Discontinue after 3 consecutive non-responses or once five minutes have passed.

Item 3 (Story Pack 2)

Place cards from Story Pack 2 in front of participant in the order shown below. Place down each card one at a time in sequence from first to last scene.



Once cards placed in front of child, say:

> ***Let's play again with different pictures. I would like you to tell me a story using the pictures in the comic strip. I will start the story*** *[point to first card]* ***and then you can carry it on. You should include all the pictures in order*** *[point to cards in order]* ***and the story should finish here*** *[point to last card].* ***The story must use the pictures but you can add in other things you can't see in the pictures too. Remember, try to be as creative as you can!***
>
> ***OK, now listen to the start of the story. You can take some time to think about it.***
>
> ***Hina is riding her bicycle after school. She is planning to go to the park to play football.***
>
> ***Now you carry on the story when you are ready…***

Start audio recorder once instructions read out. Record time taken to start story after instructions given . If no response after 45 seconds, give prompt:

> ***This is quite tricky, isn't it? Would you like me to give you some help?***

If child indicates they would like help, begin the prompting procedure below. Record the number of prompts given and the time taken to respond following each prompt. Discontinue task after 3 consecutive non-responses. If child indicates they would like more time, allow 30 more seconds and then begin to administer prompts.

1. ***Ok, let's start/look here*** *[point to first picture]*. ***What can you see in the picture?***

Following child's response say ***Great! Now see if you can carry on with the story, tell me what could be going on in the pictures.***

[Allow 20 seconds without any response before moving to prompt 2.]

2. ***OK, I can see a …*** *[insert description e.g. cat on the pavement].* ***What else can you tell me?***

   Following child's response say ***Great! Now see if you can carry on with the story, tell me what could be going on in the pictures.***

   [Allow 20 seconds without any response before moving to prompt 3.]

3. ***OK, I'll tell you what I think is going on…*** *[Insert story action e.g. Hina has got off her bike to help the cat].*

   What do you think happens next? *[Point to next picture].*

   Following child's response say ***Great! Now see if you can carry on with the story, tell me what could be going on in the pictures.***

   If child does not respond after 20 seconds say, ***OK, this is a hard story. Let's finish here and move on.***

If during storytelling the child pauses for 20 seconds, say:

> ***Well done. What happens next?***

If the child does not respond within 20 seconds, offer to help and begin the prompting procedure as outlined above, starting at the relevant picture.

Terminate task when either a) discontinue rule met or b) five minutes have passed following start of task (i.e. start of recording following instructions). After 4:30 seconds, say:

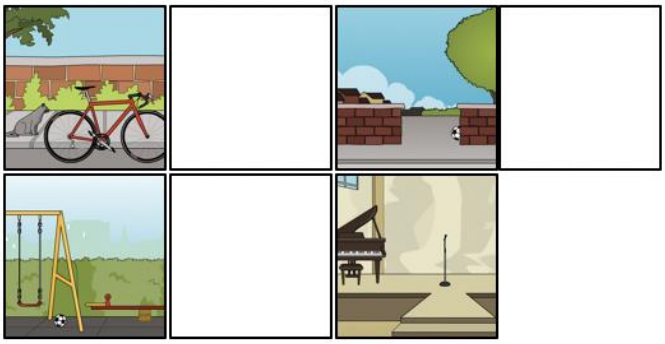
> ***OK, I'm going to stop you for a moment. I can see you are trying very hard with this story but we don't have much time. Can you finish the story now so we can move onto the next part?***

Allow 30 more seconds and then say:

> ***Let's finish there. Well done!***

Item 4 (Story Pack 2 + blank cards)

Leave the story cards in front of the child and add in three blank cards, as shown below.



Then say:

> ***Now the comic strip has some pictures missing. There is a picture missing from the beginning*** *[point]****, middle*** *[point]* ***and end*** *[point]****. You need to fill in the missing pictures to complete the comic strip. Please tell me a <u>different</u> story using all the pictures. I will start the story again*** *[point to picture on the left]* ***and then you carry on. Remember to finish the story here*** *[point to picture on right]* ***and be as creative as you can!***
>
> *[Check understanding of instructions before proceeding.]*
>
> ***OK, now listen to the start of the story. You can take some time to think about it.***
>
> ***Enzo is late to a concert. He is cycling fast when he falls off his bike.***
>
> ***Now you carry on the story when you are ready…***

<u>Start audio recorder</u> once instructions read out. Record time taken to start story after instructions given .

If no response after 45 seconds, administer prompts as described above. Discontinue after 3 consecutive non-responses or once five minutes have passed.

Collect cards once item complete.

Complete the Measure Acceptability Scale at the end of the task.

Situation and Solutions Task

Materials:

- Object stimulus cards with pictures and text label (10x10cm size)
- Stop watch

In this task participants are asked to generate solutions to everyday situations. Some of the problems are realistic whereas others require more imagination. Five common objects are provided which can be used to solve the problems. The objects can be used independently or in combination. The objects can be used in both usual or unusual ways.

Main task

> ***This game is about finding ways to help people. I am going to tell you about everyday situations. I am also going to give you eight different objects. You should think of as many ways to help as you can by using the objects. The objects can be used in whichever way you like. You can use them together and you can use the same object again, but you can only use the objects you have been given. Does that make sense?*** *[If child does not understand then repeat instructions].*
>
> ***Can you please repeat the instructions back to me?***

If child misses part of instructions, say:

> ***Well done. Remember [insert missing instruction].***

Record any instructions missed.

> ***Ok, let's start.***

Read out the following instructions before placing the object stimulus card in front of the child.

 Book	 Clock	 Pen	 Balloon
 Apple	 String	 Box	 Bell

I am going to start by showing you the objects. *[Point to first object on stimulus card].* ***What is this?***

*If child answers correctly, say "**Well done"** and proceed to next item. If child answers incorrectly, name the object and proceed to the next item. Do not tell the child what the object is used for. [Record the child's responses to naming these items.] Then turn the cards over and say,*

Now can you tell me what objects are on the cards?

*Once the child has named all items or cannot recall anymore, say **"Well done"** and turn the cards over. Point to any cards the child missed and name them by saying **"Good job. You also have a [insert missing object name]."***

Now I am going to tell you the first situation. I will read it out twice, so please listen carefully. Once I have finished, you should tell me as many solutions as you can think of using the objects here *[point].* ***Ready?***

Read item twice and then start timing. If child asks for clarification then repeat instructions but do not stop timing. If child is silent for 15 seconds, then prompt:

Those are good ideas. Can you think of anything else? It can be anything at all using the given objects.

Discontinue when the child indicates they have no other ideas or have been silent for 30 seconds following prompt. Then say:

Well done. Which are your two best ideas? *[Wait for child response].*

*If the child uses an object not provided, then say **Remember you can only use the objects on the cards in front of you.** Repeat this prompt only once per item.*

Following each item, say:

Well done. Let's move on to the next one. Remember you can use any of the objects you have been given in any way you like. You can use them more than once and as many as you want at the same time. Tell me as many ideas as you can.

Item 1

Emma often gets told off by her Mum for having a messy room. Emma has decided to tidy her room but she doesn't know where to start. What could Emma do?

Item 2

Harry likes to play puzzles and games inside. At playtime he has to play outside with his classmates. He wants to join in the games with them but does not know how. What could Harry do?

Item 3

Amelia is on a shopping trip with her Mum in the supermarket. Amelia has a pet hamster that she loves and takes with her everywhere. Suddenly, Amelia realises her pet hamster has run away in the supermarket. What could Amelia do?

Item 4

Mohammed and Ruben are going to their friend's Birthday party. When they arrive, they realise that it is a fancy dress party. They are not wearing costumes! What can they do?

Item 5

Josh finds it hard to get ready for school in the morning. He needs to get dressed, have breakfast, brush his teeth and have his school bag ready on time. The night before school Josh decides to make a plan to help himself get ready for school the next morning. What could Josh do?

Item 6

Farida wants to grow flowers in the garden for a school project but her neighbour's rabbit keeps eating them! What can Farida do?

Complete the Measure Acceptability Scale at the end of the task.

Appendix C. Scoring Criteria for the Toy Improvement Task and the Situations and Solutions Task

Fluency

The total number of responses, including repetitions but excluding errors.

Errors

There were several categories of errors. Redundant errors were classified as responses that were unable to be interpreted in the given context or were not clear in relation to scenario (e.g. ‘put the bell in the box because you don’t need that’). These types of errors possibly indicated that the instructions were misunderstood. Irrelevant errors were nonsensical or random responses. For the Situations and Solutions task a further category of error was defined. Non-object errors were solutions that used additional objects or none of the target objects, (e.g. ‘put the hamster in a cage’). Some items referenced other objects, for example a schoolbag in item 5; if the response included these additional objects as well as the target objects then this was not scored as an error.

Repetitions

Any repetitions or only minor variations of a previous response within a set (e.g. ‘make it bigger’ and ‘make it gigantic’).

Originality

Each response was given a score of 1 to 7 indicating how original it was, with 1 being not at all original and 7 being highly original. Originality was defined as the degree to which an idea was ‘unusual, surprising and interesting’. This definition was used as opposed to a measure of statistical infrequency to capture the multiple qualitative aspects of originality and overcome the issue with unique but mundane responses receiving the highest scores.

Usefulness

Each response was given a score of 1 to 7 indicating how useful it was, with 1 being not at all useful and 7 being highly useful. Usefulness was defined in relation to how

‘appropriate, feasible and functional’ the idea was. Raters were further asked to consider the prompts ‘how well does it address the problem?’ and ‘would it work?’ in relation to each response.

Flexibility

In the Toy Improvement task, flexibility was scored in relation to the number of thematic categories encompassed by the response set (n = 9). These categories were defined by the main researcher and then verified by a second independent rater. Any discrepancies or disagreements were resolved through discussion between raters. This also helped to refine the category descriptions, set out in Table A1.

In the Situations and Solutions task, flexibility was scored by counting the number of different object categories used in the response set, yielding a maximum score of 8 if all object categories were incorporated. This method was used based on trials of different scoring methods with sample data, which found that thematic categories broadly mapped onto object functions (for example, the theme ‘timekeeping’ corresponded to the clock) and therefore counting object categories allowed a simpler approach. When use of an object was implied but not named it was counted anyway (e.g. ‘write in the book’ would score 2 for both book and pen).

Within-category fluency

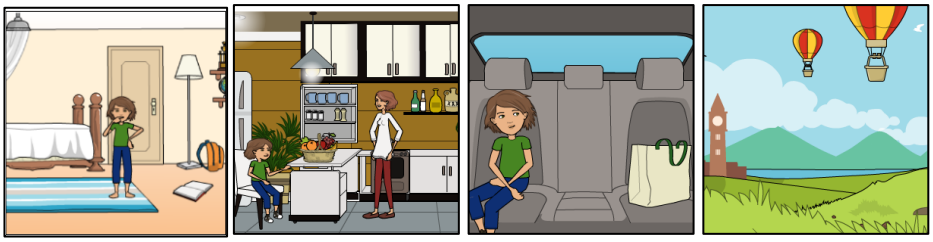
The average number of responses within each category was calculated by dividing the total number of responses (fluency or object frequency) by the number of different categories used (flexibility). Higher scores indicated a higher frequency of ideas within the same category.

Table A1. Description of thematic categories used in the Toy Improvement Task.

Code	Label	Description
1	manipulations	ways to move and position elephant e.g. move its legs
2	materials/ texture	changes to fabric, touch e.g. waterproof
3	features/ characteristics	changes to size, limbs and appearance (light, sound, smell, taste) e.g. prettier eyes
4	accessories/ additions	incorporating new aesthetic elements with elephant e.g. give it a hat
5	equipment/ tools	incorporating new functional elements with elephant e.g. give it wheels
6	transformations	changing characteristics/features of elephant into another object e.g. elephant alarm clock that makes elephant noises
7	animation/ action	performing an action independently, life-like e.g. swim
8	interactive/ sensory	detects and responds to input/environment or uses senses e.g. talk and responds back
9	uses/places	ways it can be used in existing form / within surroundings e.g. take it outside

Appendix D. Cartoon Scenes and Basic Plot Outlines for Items in the Story Scenes Task

Item 1



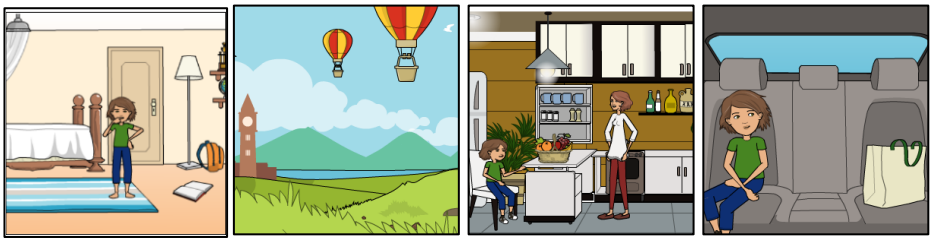
There is a girl called Alex in her bedroom. It is her day off school. She has got dressed and is ready to start her day.

She goes downstairs to the kitchen. Her Mum [or substitute other character] is in the kitchen. She eats (fruit for) breakfast. She talks to her Mum about what they will do today.

They get into the car/ they are driving in the car. There is a bag in the back of the car. Alex does not know what is inside the bag/ it is a surprise.

She sees hot air balloons out of the window/ they go on a hot air balloon ride.

Item 2



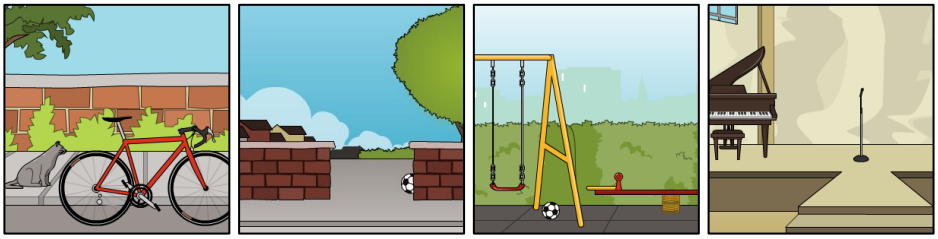
There is a boy called James. He is at his friend Olly's house. They have done their homework and now are playing a game.

They play a game outside and there are hot air balloons/ they imagine a game with hot air balloons.

They go downstairs/ come inside to eat dinner [or other meal]. They talk to Olly's Mum.

James gets in the car/ gets picked up to go home. There is a bag in the backseat of the car.

Item 3



Hina is riding her bicycle after school. She is planning to go to the park to play football.

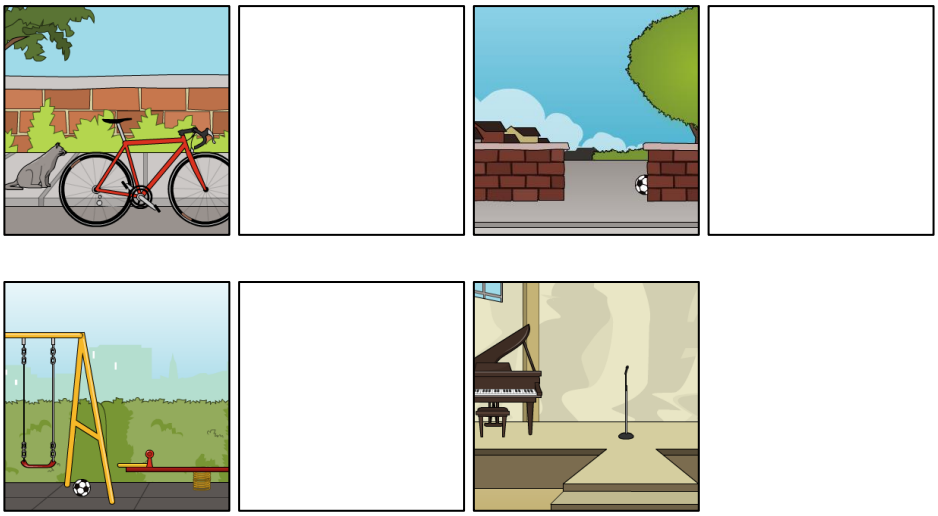
She gets off her bike on the pavement next to a cat.

She sees a ball behind the wall. She kicks/ plays with the ball.

The ball lands in the park next to the swings. She plays in the park (on the swings/ see-saw/ with the football).

Hina leaves the park and goes to a music concert/ performance.

Item 4



Enzo is late to a concert. He is cycling fast when he falls off his bike.

He walks with his bike (and a cat) along the road. He walks past a wall and there is a football behind the wall. He kicks/ plays with the football.

He plays in the park (on the swings/ see-saw/ with the football).

He continues his journey and arrives at the concert. He is just in time to watch the concert/ he is too late and has missed the concert.

Appendix E. Narrative Coherence and Narrative Elaboration Scales Used to Score the Story Scenes Task, Taken from Dillon and Underwood (2012)

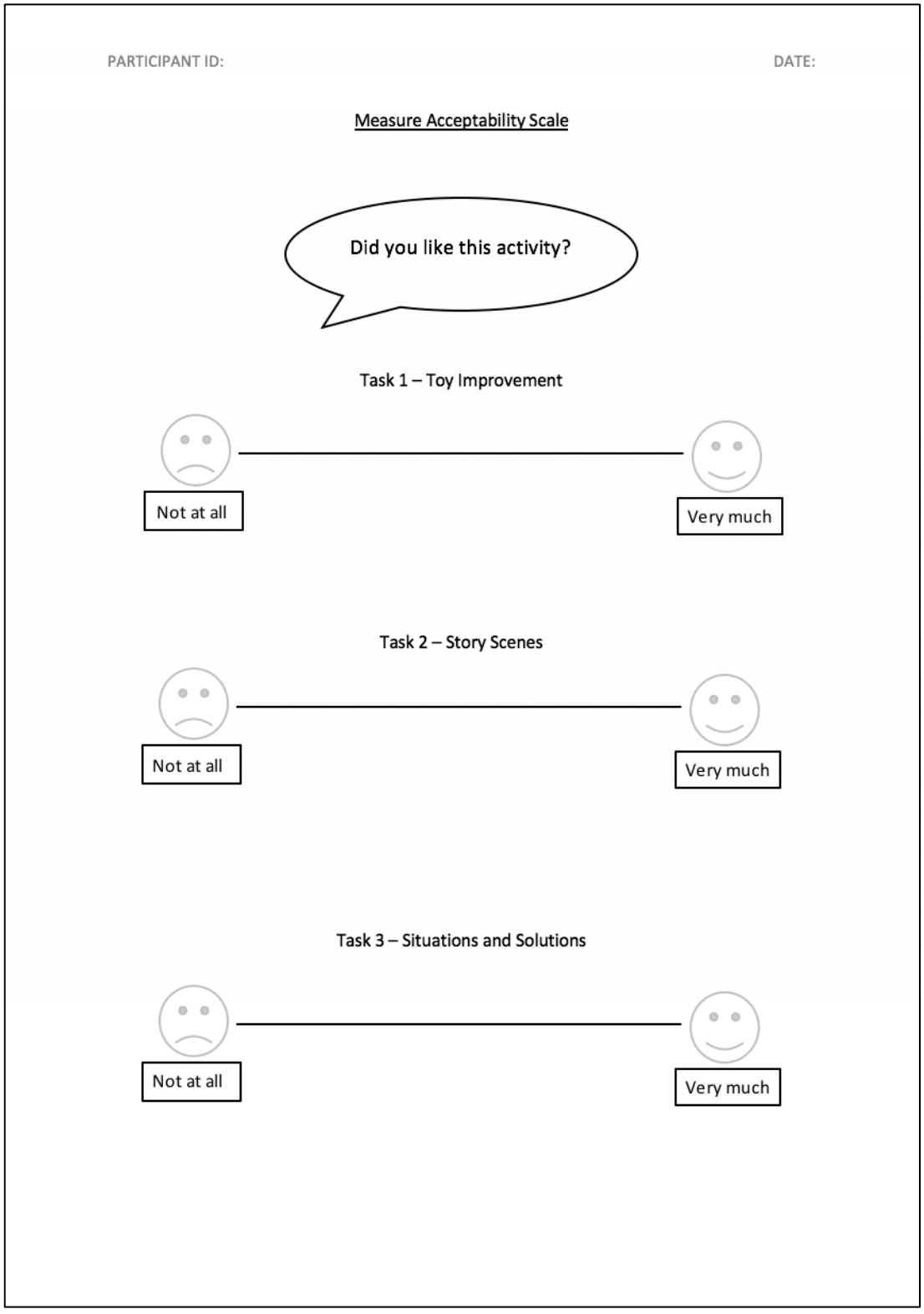
Narrative Coherence scale

Score	Definition
0 – no coherence	No mention of appropriate action
1 – weak	Mention of fragments of the action but with no logical order
2 – moderate	Explains the action in a logical manner with features linked together, but fails to acknowledge all salient features
3 – good	Explains action in a coherent logical manner with the use of all salient features including story resolution

Narrative Elaboration scale

Score	Definition
0 – no elaboration	Of either characters or story development
1 – weak	Some evidence of the development of character description and/or story beyond action given, although no elaboration of such detail given
2 – moderate	Elaboration of either story *or* character with elaboration in coherent format and not fragmented
3 – good	Coherent development of both story and character beyond original starting point

Appendix F. Measure Acceptability Scale



Appendix G. Parent's Evaluation of Children's Creativity (Runco, Johnson & Bear, 1993)

Not included due to copyright restrictions.

Appendix H. Inventory of Creative Activities and Achievements (Jauk, Benedek, Dunst & Neubauer, 2013)

Not included due to copyright restrictions.

Appendix I. ECO-TED Task Script for Schoolbag, Lego and Alternating Sequence Tasks

School Bag Task
(Originally relating to item 6)

Items required:

- School bag task poster
- Sticky labels
- Stop watch
- Lesson prompt cards

Participants are asked to plan for activities at school the next day. They will be presented with a poster (placed on the table directly in front of them) depicting 81 objects, including those they need for the given school activities, as well as distractor items which they do not need. Sticky labels are used to show which items they have selected. After each trial, the examiner **removes the labels.**

The examiner gives the following instructions:

Look at this poster – it has lots of pictures on it. Let's see if you can find some things on the poster.

Scanning trial 1:

Can you show me a jam jar?

If they point to the jar say:

Well done. When you want to pick something on the poster, you need to stick one of these on it to show me you have chosen it. Try sticking one on the jam jar.

[If they cannot find the jam jar within 60 seconds the examiner points it out, and proceeds to the next scanning trial]

Scanning trial 2:

Hand the child the sticker dispenser and say:

Now find the sunglasses.

[If child does not put the sticker on, remind them that they need to do this whenever they pick an item]

[If they cannot find the sunglasses within 60 seconds the examiner points them out, and proceeds to the next scanning trial]

Scanning trial 3:

Now find the Clock

[If child does not put the sticker on, remind them that they need to do this whenever they pick an item]

[If the child fails all three scanning items, do not proceed with the task]

Auditory memory trial 1:

Now I am going to ask you to find three things, and put stickers on them all. Please find a teddy bear, a rubber duck and some tennis balls.

[If participant cannot remember what they need to find do not prompt them. Instead, say: "I can't tell you again. Just try to remember what I said, and if you can't remember just guess"]

[If participant finds the three items within 120 seconds, proceed to the main task. Otherwise, attempt auditory memory trial 2]

Auditory memory trial 2:

Please find a scarf, some sun cream and a blue pencil.

[If participant cannot remember what they need to find do not prompt them. Instead, say: "I can't tell you again. Just try to remember what I said, and if you can't remember just guess"]

[If participant finds the three items within 120 seconds, proceed to the main task. Otherwise, discontinue task]

Main task:

Now we will play a game using the poster. I will ask you to imagine that you have certain lessons and activities at school tomorrow. You will need to choose things from the poster to bring to school for these lessons and activities. You can pack as much or as little as you like, but only pack exactly what you will need for each activity. Do not pack anything that you won't need.

I'll have a go at the game to show you how it is played.

Examiner takes practice prompt card and puts it in from of them. They then say:

The card shows me I have a French lesson tomorrow. So…I need to choose from the poster what I will need.

Examiner takes their time looking over whole poster and then says:

I'm going to put a sticker on this French dictionary because I'll need it for the French lesson *[put sticker on]*. ***I mustn't put a sticker on this sun lotion because I won't need it for the French lesson.***

Now it's your turn – but before you start do you have any questions about this game?

Remember, you can pack as much or as little as you like for any activity, but don't pack things you won't need for that activity.

If at any stage the participant selects items they would not need for the lesson in question i.e. if they select an item not listed in the "correct items" section of the score sheet, you may prompt them ONCE ONLY over the course of the whole task:

Remember, you can pack as much or as little as you like, but don't pack things you won't need.

⇨ **Discontinue rule:** Discontinue task if child is unsuccessful on 3 consecutive trials i.e. if they do not select at least one of the correct items specified on the score sheet within 120 seconds.

Task Script
Version 8: February 2016

Item 1

Show the participant Card 1 (i.e. Science). Say:

Okay, you're now going to choose what you'll need for your Science lesson. In this lesson you're going to be looking at plant leaves using a magnifying glass. What do you need?

[If child does not use stickers record their answers anyway, but prompt them to use stickers for subsequent trials. For subsequent trials prompt the participant to use stickers as needed. Also prompt the participant to tell you when they are finished if they do not do this of their own accord.]

For all items record the correct and incorrect items packed. Also record the time taken for the participant to select their items, starting timing as soon as the instructions are complete. You may *repeat the instructions once* if the participant asks but start recording as soon as the instructions are complete. Also note whether the participant paused between hearing the instructions and starting to select their items (see score sheet).

Item 2

Show the participant Card 2 (i.e. break time). Say:

It's break time. A friend in your class is bringing in a cake to celebrate their birthday. They asked you to bring in some balloons for them. What do you need?

Item 3

Show the participant Card 3 (i.e. Art). Say:

> ***You have Art. Your teacher will ask you to draw a ladybird. You are only allowed to use colouring pencils for this. The ladybird should look like a real ladybird as much as possible. What do you need?***

Item 4

Show the participant Card 4 (i.e. break time). Say:

> ***It's break time. You will go to a shop near school to buy a bottle of water. It will probably rain during break time. What do you need?***

Item 5

Show the participant Card 5 (i.e. lunch). Say:

> ***At school tomorrow, you will have a packed lunch from home. You're not allowed chocolate or crisps, but you need to have a sandwich and two items containing fruit. What do you need?***

Item 6

Show the participant Card 6 (i.e. P.E.). Say:

> ***It's P.E. You can choose between tennis and football. You'll need shoes, socks, shorts and a top: white for tennis, red for football. School provides everything else. What do you need to bring to school?***

[If the participant asks/tries to prepare for both football and tennis, ask them to choose and prepare for one option only]

Item 7

Show the participant Card 7 (i.e. end of school day). Say:

> ***After P.E., it's the end of the school day. You could go to the school library – you've got a world atlas to return, and you want to get a book about the rainforest out. You've also got a five pence fine to pay off. Or you could go to homework club. You'll be saving your answers on a computer, but need to bring your maths textbook and calculator. You should bring a drink too. What do you need?***

[If the participant asks/tries to prepare for both the library and homework club, ask them to choose and prepare for one option only]

Lego Task
Task relating to searching and retrieving (originally relating to item 14 on list):

Items required:

- Two bowls of Lego
- Spare Lego pieces for demonstration/learning
- Cup
- 6 small animal figures (3 in each Lego bowl)
- Stop watch

In this task participants will be asked to sort and select Lego. They will first be given the instructions below and shown 2 bowls containing 65 pieces of Lego each and three small animals. *NB the bowls should contain a greater quantity of each piece than the amount requested so that participants need to recall how many pieces were asked for i.e. if asked to find one piece; the bowl should contain at least two of these pieces.* Say:

> ***In a moment I'd like you to help me find some Lego pieces that are hidden in these bowls. I want you to find the following pieces as quickly as you can:***
>
> - ***A piece like this one*** *[show a spare 3x2 orange piece and put on the table in front of the participant]*
> - ***2 pieces like this one*** *[show a spare 4x2 white piece and put on the table in front of the participant]*

> - ***3 pieces like this one*** *[show a spare 2x2 brown piece and put on the table in front of the participant]*
>
> ***I want to be sure you know what to find because once we start I won't be able to remind you. Can you remember what I asked you to look for?***

The participant must demonstrate that they understand which pieces they are searching for. They can do this by referring to the example pieces on the table (e.g. 3 of this one) or by describing the length, width and colour of the pieces. If the participant gets any of the pieces wrong repeat the list in the following way until they are able to remember them correctly **(up to three times only – discontinue if they are unable to learn the list)**:

> ***We need:***
>
> - ***A piece*** *like this one [show the spare 3x2 orange piece and put it back on the table in front of the participant]*
> - ***Two pieces like this one*** *[show the spare 4x2 white piece and put it back on the table in front of the participant]*
> - ***Three pieces like this one*** *[show the spare 2x2 brown piece and put it back on the table in front of the participant]*

Then say:

> ***Once you find each piece please put it in this cup*** *[show participant the cup].*

> ***There is also something else you need to do during this game. There are some animals like this one*** *(show them a spare example)* ***hidden in the bowls. For each bowl, choose your favourite animal and put it by this cup.***
> ***Okay, can you tell me what you need to do in this game?***

Ensure child knows: (1) they have to look for specific Lego pieces; (2) these are the pieces that are in front of them on the table **(note: participant does not need to repeat their description/list of these pieces)**; (3) they should put the pieces in the cup when they find them; (4) they also need to identify a favourite animal from each bowl; (5) and place these by the cup.

If any of these elements are missing from the child's answer, explain that element to the child.

Next remove the target Lego examples from view.

Allow the participant to continue until they tell you they have finished the task. If they ask for a reminder or help, say:

> ***I can't give you a reminder. Just do your best and let me know when you've finished.***

Record the contents of the cup, whether they have selected animals correctly, and the time taken.

⇨ **Discontinue rule:** Discontinue task if child is unable to grasp what is required after all prompts given or fails to start task within 60 seconds of all the instructions being given.

Alternating Sequence Task
(Originally relating to item 8)

Items required:

- Alternating Sequence Stimulus Cards
- Alternating Sequence Response Sheet
- Black felt tip pen
- Stopwatch

For this task the participants will be asked to copy and complete a pattern of alternating shapes on a piece of paper. There will be three items and the patterns will be of increasing complexity. The child will have to draw the pattern from left to right of the response sheet without taking their pen off the paper.

Item 1/Pattern 1:
Place Pattern 1 stimulus card in front of the participant. Say the following:

> ***I am going to start by showing you a pattern on a piece of paper. You draw the pattern like this*** *(trace over the pattern so that the participant can see how the pattern is drawn without removing the pen from paper).* ***I want you to copy the pattern underneath exactly as you see it without taking your pen off of the paper. Start here*** *(point to left hand side of paper)* ***and finish here*** *(point to right side of paper)****. Be as quick as you can. You can start when I say go.***

Ensure that the response page is positioned in front of the child, with the stimulus card above it, and place a black felt tip pen in front of the participant.

> ***You should draw the pattern here.*** *(*Point to the space for item 1/pattern 1 on the response sheet)
> ***Ready? Go.***

Start timing as soon as you have said "go". Stop timing once the participant reaches the end of the paper or once two minutes has elapsed. Record time on the score sheet and proceed to item 2.

A pattern correctly drawn from left to right of the page without removing pen from paper should be given a score of '2'. If the participant does not complete the pattern (score '0') or removes the pen from the paper (score '1'), give the following reminder:

> ***Remember you need to copy the pattern exactly as it is without taking your pen off of the paper. Let's try another one.***

If the participant is observed to perseverate (draws one of the shapes repeatedly) this should be recorded in the 'observations' section on the score sheet.

Item 2/Pattern 2
Place Pattern 2 stimulus card in front of the participant. Say the following:

> ***Here is the next pattern I would like you to draw. You should draw it in the space here*** *(point to correct place on response sheet)****. Remember draw it just like the picture without taking your pen off of the paper. You should do it as quickly as you can.***
>
> ***Go.***

Start timing as soon as you have said "go". Stop timing once the participant reaches the "finish" line or once two minutes as elapsed. Record time on the score sheet along with the awarded score for the pattern (0/1) and proceed to item 3.

Item 3/Pattern 3
Place Pattern 3 stimulus card in front of the participant. Say the following:

> ***Let's try another. Go.***

Start timing as soon as you have said "go". Stop timing once the participant reaches the "finish" line or once two minutes as elapsed. Record time on the score sheet along with the awarded score for the pattern (0/1). If the participant is unsure how the pattern is drawn, they should not be shown how to do it but should instead be encouraged to work it out themselves.

⇨**Discontinue Rule:** All three items should be attempted regardless of the participant's success on previous items.

Appendix J. Letter Confirming Favourable Ethical Opinion from the Research Ethics Committee and Email Confirmation from the Health Research Authority

NHS
Health Research Authority

London - Westminster Research Ethics Committee
4 Minshull Street
Manchester
M1 3DZ

Tel: 0207 104 8012

Please note: This is the favourable opinion of the REC only and does not allow the amendment to be implemented at NHS sites in England until the outcome of the HRA assessment has been confirmed.

05 October 2017

Ms Polly Cocker
Camden and Islington NHS Foundation Trust
4th Floor, East Wing
St Pancras Hospital
4 Saint Pancras Way
London
NW1 0PE

Dear Ms Cocker

Study title:	**Validating a new ecologically valid measure of executive functioning for children with autism spectrum disorder (ASD)**
REC reference:	**15/LO/1332**
Protocol number:	**N/A**
Amendment number:	**2**
Amendment date:	**18 August 2017**
IRAS project ID:	**170531**

The above amendment was reviewed by the Sub-Committee in correspondence.

Favourable opinion

The amendment consisted of changes to the study design, including changes to the study measures.

The recruitment process seemed to involve passing a consent form to the child's teacher. The Sub-Committee was unclear on whether this was a consent to be contacted form or a consent to take part in the study form.

When contacted by email the researchers explained it was a consent to participate form. Parents were asked to provide contact details on the form so the researchers could get in touch with them to arrange a time to carry out the testing.

The Sub-Committee were grateful for the clarification and found the consent form acceptable.

Since some people might decide not to take part, the Sub-Committee asked that the sentence "To find out more information and take part in the study..." be changed to "To find out more information..." in the advert.

The researchers submitted a revised advert which the Sub-Committee found acceptable.

The members of the Committee taking part in the review gave a favourable ethical opinion of the amendment on the basis described in the notice of amendment form and supporting documentation.

Approved documents

The documents reviewed and approved at the meeting were:

Document	*Version*	*Date*
Copies of advertisement materials for research participants [for Clinical Participants]	1	18 August 2017
Letters of invitation to participant [Parents - control]	6	18 August 2017
Letters of invitation to participant [Parents - GOSH clinical]	5	18 August 2017
Letters of invitation to participant [Parents - other clinical]	6	18 August 2017
Notice of Substantial Amendment (non-CTIMP)	2	18 August 2017
Participant information sheet (PIS) [and assent form - children]	4	18 August 2017
Participant information sheet (PIS) [and consent form - parents clinical]	8	18 August 2017
Participant information sheet (PIS) [and consent form - parents control]	7	18 August 2017
Research protocol or project proposal [clean]	4	18 August 2017
Research protocol or project proposal [tracked]	4	18 August 2017
Validated questionnaire [The Inventory of Creative Activities and Achievements]		18 August 2017
Validated questionnaire [Parent's evaluation of children's creativity]		

Membership of the Committee

The members of the Committee who took part in the review are listed on the attached sheet.

Working with NHS Care Organisations

Sponsors should ensure that they notify the R&D office for the relevant NHS care organisation of this amendment in line with the terms detailed in the categorisation email issued by the lead nation for the study.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

We are pleased to welcome researchers and R & D staff at our Research Ethics Committee members' training days – see details at http://www.hra.nhs.uk/hra-training/

15/LO/1332:	**Please quote this number on all correspondence**

Yours sincerely

[redacted]

Mr Robert Goldstein
Chair

E-mail: [redacted]

Enclosures: List of names and professions of members who took part in the review

Copy to: Mr Elliott Dickens, Great Ormond Street Hospital
Dr William Mandy, University College London
Smaragda Agathou

From: "AMENDMENTS, Hra (HEALTH RESEARCH AUTHORITY)" [redacted]
Date: 12 October 2017 at 12:09:25 BST
To: [redacted]
Cc: [redacted]
"Re[redacted]
Subject: RE: 170531; 15/LO/1332/AM02 - SA2 - Assessment of Amendment Complete

Dear Dr Mandy and Polly Cocker

Further to the below, I am pleased to confirm that HRA Approval has been issued for the referenced amendment, following assessment against the HRA criteria and standards.

The sponsor should now work collaboratively with participating NHS organisations in England to implement the amendment as per the below categorisation information. This email may be provided by the sponsor to participating organisations in England to evidence that the amendment has HRA Approval.

Pl[redacted]nts@nhs.net for any queries relating to the assessment of this amendment.

Kind regards

Maeve

Maeve Ip Groot Bluemink
Assessor
Health Research Authority
E. [redacted]
W. www.hra.nhs.uk

Appendix K. Information Sheets and Consent Forms

Information sheet and consent forms 1 – for parents/carers of children with ASD

Information sheet and consent forms 2 - for parents/carers of typically developing children

Information sheet and consent forms 3 - for both children with and without ASD

1. Information sheet and consent form for parents/carers of children with ASD

V8 18/08/17

Great Ormond Street Hospital for Children NHS Foundation Trust

UCL

PARENT/GUARDIAN INFORMATION SHEET AND CONSENT FORM FOR CHILDREN WITH A DIAGNOSIS OF AUTISM

Developing a Measure of Planning and Organisation for Children with ASD

We would like to invite you and your child to take part in our research study. Before you and your child decide whether you would like to take part, it is important for you to know why the research is being done and what it will involve. Please take time to read this information sheet carefully and discuss it with others if you wish. If there is anything that is not clear, or if you would like more information, please do not hesitate to contact us.

Why is the study being done?

There has been lots of research that has suggested that people with Autism Spectrum Disorders (ASD) can have difficulties with executive functioning. This is a term used to describe the many tasks our brains perform that are necessary to think, act, and solve problems. Executive functioning includes tasks that help us learn new information, remember and retrieve information we've learned in the past, and use this information to solve problems of everyday life.

There are a number of tests currently available that aim to assess a child's executive functioning. The problem with these tests is that they have not been specifically designed for children with ASD. For this reason the tests can sometimes miss some of the everyday difficulties that are seen in individuals with the diagnosis. Through our research we are hoping to develop a new test that more accurately assesses these difficulties so that we can gain a better idea of how executive functioning is affected in those with ASD.

What will happen if we take part?

If you agree to take part in this research, your child will be seen by one of the study researchers at the Social Communications Disorder Clinic at Great Ormond Street Hospital, or at your home depending on what is most convenient for you. This meeting will last no longer than one hour. During that time, your child will do some games and puzzles that look at how they think and process information. The games have been designed to be fun.

You will be asked to fill in some questionnaires about your child's behaviour, communication and feelings. These are widely used and should take no longer than one hour to complete. These can be completed whilst your child takes part in the games and puzzles or in your own time.

In addition we will also ask for your permission to access some of the routine information collected as part of your child's assessment at the Social Communications Disorders clinic. This will include information on your child's diagnosis and their IQ score. If you give permission we will liaise directly with your child's care team to collect this information.

A small number of children will be asked to take part in the games and puzzles for a second time. This shorter follow-up session will take place around a month after the first visit and will take no longer than one hour. Only your child will need to take part in this session.

An information sheet for your child has been provided. Please talk about the study with your child. We will also make sure that your child understands what he/she will be doing and give them an opportunity to ask any questions that they may have.

As a small thank you for taking part in our study we will offer your child a £5 voucher.

What are the possible disadvantages and risks of taking part?

Whilst we expect that most children will enjoy the puzzles and games, it is possible that some children may find them hard work or frustrating to complete. We will offer regular breaks and give your child the opportunity to stop at any time should this happen. They will

also have the chance to talk to the researcher's about how they found taking part once finished.

What are the potential benefits?

We hope that our findings will help to develop a better measure of executive functioning for children with ASD. There is no immediate benefit for the children taking part in the study, but we hope that their help will be beneficial to other children in the future.

Does my child have to take part in this study?

It is up to you and your child whether or not you take part in this study. If you do decide to take part, you will be asked to sign a consent form. If you decide now, or at a later date, that you do not wish to participate in this research you are free to withdraw at any time without giving a reason. Even if you are happy for your child to take part, he or she will still decide for themself. It will be explained to your child that he/she can choose to withdraw from the study at any time, without giving a reason. We want to make sure that everyone is happy when taking part in our project. We would also like to stress that if you decide not to take part in the research; it will not in any way affect the care that your child receives.

Will taking part in this study remain confidential?

All information collected from you and your child during the course of this research will be kept strictly confidential. No one, other than the researchers involved in the study, will have access to your or your child's personal details or any of the information provided to the Service. This information will be kept in locked cabinets and stored anonymously at University College London (UCL).

Who has reviewed this study?

All research in the NHS is looked at by an independent group of people, called a Research Ethics Committee, to protect your interests. This study has been reviewed and given a favourable opinion by the Westminster Research Ethics Committee.

What will happen to the results of the research?

The information collected from children with a diagnosis of ASD will be compared to a group of children without ASD to see whether the test of executive functioning is useful in differentiating between those with and without the diagnosis. The findings of the study will be written up as part of a doctoral thesis. However, names and other identifying information will be removed. The results of the study may be presented at national and international conferences and published in academic journals. Neither you nor your child will be personally identified in any reports or publications of the research. If you wish, a summary of the findings can be sent to you via post or email once the study is complete.

How to contact the researchers

If you have any further questions or would like assistance at any point during the study, please feel free to contact Polly Cocker at UCL on [redacted] or email [redacted] In the case of a complaint, please contact Dr Will Mandy via [redacted] We are happy to talk through any questions with you.

Thank you for taking the time to read this information sheet.

Your help makes our research possible!

University College London holds insurance against claims from participants for harm caused by their participation in this clinical study. Participants may be able to claim compensation if they can prove that UCL has been negligent. However, if this clinical study is being carried out in a hospital, the hospital continues to have a duty of care to the participant of the clinical study. University College London does not accept liability for any breach in the hospital's duty of care, or any negligence on the part of hospital employees. This applies whether the hospital is an NHS Trust or otherwise.

Great Ormond Street NHS Hospital for Children
NHS Foundation Trust

UCL
Polly Cocker
Department of Clinical, Health and Educational Psychology
University College London
1-19 TORRINGTON PLACE, LONDON, WC1E 7HB
Tel [redacted]
Email [redacted]

Please tick (√) appropriate box:

☐ **Yes**, my child and I are happy to participate in this study

☐ **No**, we do not want to participate in this study.

If Yes, please complete the following:

(Please initial box)

☐ I have read the Information Sheet.

☐ I understand that I am free to withdraw my child from the study at any time without giving a reason.

☐ I understand that my child is free to withdraw from the study at any time without giving a reason.

☐ I give consent to be sent some questionnaires to complete regarding my child.

☐ I am happy to be contacted again to take part in further research.

☐ I am happy to be contacted for a second time to arrange a shorter follow-up session.

☐ I have had the opportunity to ask any questions I wish to ask.

☐ I have the contact details of the research team in case I have any queries in the future.

Child's Name: _____ Parent's Name: _____

Parent/Guardian Signature: _____ Date: _____

Researcher Signature: _____ Date: _____

Contact Details (these will remain confidential and only be used to send questionnaires and arrange a session to meet with your child):

Address: _____

_____

Tel No: _____

PLEASE PROVIDE AN EMAIL ADDRESS IF YOU WOULD LIKE TO BE SENT A SUMMARY OF THE FINDINGS ONCE THE STUDY IS COMPLETED

Email: _____

2. Information sheet and consent form for parents/carers of typically developing children

V7 18/08/17

Great Ormond Street Hospital for Children NHS Foundation Trust

UCL

PARENT/GUARDIAN INFORMATION SHEET AND CONSENT FORM OF TYPICALLY DEVELOPING CHILDREN

Developing a Measure of Planning and Organisation for Children with ASD

We would like to invite you and your child to take part in our research study. Before you and your child decide whether you would like to take part, it is important for you to know why the research is being done and what it will involve. Please take time to read this information sheet carefully and discuss it with others if you wish. If there is anything that is not clear, or if you would like more information, please do not hesitate to contact us.

Why is the study being done?

Please note that we are contacting you because we are keen to recruit comparison children who do not have a diagnosis of Autistic Spectrum Disorder (ASD). These children will form part of our control group.

Executive functioning is a term used to describe the many tasks our brain performs that are necessary to think, act, and solve problems. It includes tasks that help us learn new information, remember and retrieve information we've learned in the past, and use this information to solve problems of everyday life.

Children who find these things difficult can struggle in different aspects of their life. For the purpose of our study we are particularly interested in looking at thinking, learning and planning in children with ASD. There are currently a number of tests that aim to assess these skills but the problem with those already available is that they have not been specifically designed for children with ASD. For this reason they can miss some of the everyday difficulties that are seen in individuals with the diagnosis. Through our research we are hoping to develop a new test that more accurately assesses thinking, learning and planning in children with ASD.

We need a control group so that we can compare how well the children in the control group do on the test compared to the children with ASD. If the children in the control group do better on the test then we

will know that our test is good at differentiating between children with and without the disorder. That is why we'd like your child to take part.

What will happen if we take part?

If you agree to take part in this research, your child will be seen by Polly (study researcher) at school or at home, depending on which location is more convenient for you. The session will last for a maximum of two hours. During that time, your child will do some games and puzzles that look at how they think and process information. The games have been designed to be fun.

You will be asked to fill in some questionnaires about your child's behaviour, communication, strengths and weaknesses. These are simple parent-report questionnaires which are widely used and should take you no longer than one hour to complete.

A small number of children will be asked to take part in the games and puzzles for a second time. This shorter follow-up session will take place around a month after the first visit and will take no longer than one hour. Only your child will need to take part in this session.

An information sheet for your child has been provided. Please talk about the study with your child. We will also make sure that your child understands what he/she will be doing and give them an opportunity to ask any questions that they may have.

As a small thank you for taking part in our study we will offer your child a £5 voucher.

What are the potential benefits?

We hope that our findings will help to develop a more reliable measure of executive functioning for children with ASD. There is likely to be no immediate benefit for the children taking part in the study, but we hope that their help will be beneficial to other children in the future.

Does my child have to take part in this study?

It is up to you and your child whether or not to take part in this study. We kindly ask you to complete the attached form and return it to your child's teacher indicating whether you would/would not like your child to take part. If you do decide to take part but later change your mind you are free to withdraw at any time without giving a reason. Even if you are happy for your child to take part, he or she will still decide for themself. It will be explained to your child that he/she can choose to withdraw from the study

at any time, without giving a reason. We want to make sure that everyone is happy when taking part in our project.

Will taking part in this study remain confidential?

All information collected from you and your child during the course of this research will be kept strictly confidential. No one, other than the researchers involved in the study, will have access to your or your child's personal details or any of the information provided to the Service. This information will be kept in locked cabinets and stored anonymously at University College London (UCL).

Who has reviewed this study?

All research in the NHS is looked at by an independent group of people, called a Research Ethics Committee, to protect your interests. This study has been reviewed and given a favourable opinion by the Westminster Research Ethics Committee.

What will happen to the results of the research?

The information collected from children with a diagnosis of ASD will be compared to a group of children without ASD to see whether the test of executive functioning is useful in differentiating between those with and without the diagnosis. The findings of the study will be written up by as part of a doctoral thesis. However, names and other identifying information will be removed. The results of the study may be presented at national and international conferences and published in academic journals. Neither you nor your child will be personally identified in any reports or publications of the research. If you wish, a summary of the findings can be sent to you via post or email once the study is complete.

How to contact the researchers

If you have any further questions or would like assistance at any point during the study, please feel free to contact Polly Cocker at UCL on [redacted] or email [redacted] In the case of a complaint, please contact Dr Will Mandy via [redacted] We are happy to talk through any questions with you.

Thank you for taking the time to read this information sheet.

Your help makes our research possible!

University College London holds insurance against claims from participants for harm caused by their participation in this clinical study. Participants may be able to claim compensation if they can prove that UCL has been negligent. However, if this clinical study is being carried out in a hospital, the hospital continues to have a duty of care to the participant of the clinical study. University College London does not accept liability for any breach in the hospital's duty of care, or any negligence on the part of hospital employees. This applies whether the hospital is an NHS Trust or otherwise.

Great Ormond Street NHS Hospital for Children
NHS Foundation Trust

UCL
Polly Cocker
Department of Clinical, Health and Educational Psychology
University College London
1-19 TORRINGTON PLACE, LONDON, WC1E 7HB
Tel: [redacted]
Email: [redacted]

Please tick (√) appropriate box:

☐ **Yes**, my child and I are happy to participate in this study.

☐ **No**, we do not want to participate in this study.

If Yes, please complete the following:

(Please initial box)

☐ I have read the Information Sheet.

☐ I understand that I am free to withdraw my child from the study at any time without giving a reason.

☐ I understand that my child is free to withdraw from the study at any time without giving a reason.

☐ I give consent to be sent some questionnaires to complete regarding my child.

☐ I am happy to be contacted again to take part in further research.

☐ I am happy to be contacted for a second time to arrange a shorter follow-up session.

☐ I have had the opportunity to ask any questions I wish to ask.

☐ I have the contact details of the research team in case I have any queries in the future.

Child's Name: _____ Parent's Name: _____

Parent/Guardian Signature: _____ Date: _____

Researcher Signature: _____ Date: _____

Contact Details (these will remain confidential and only be used to send questionnaires and arrange a session to meet with your child)**:**

Address: _____

Tel. No: _____

PLEASE PROVIDE THE FOLLOWING DETAILS IF YOU WOULD LIKE TO BE SENT A SUMMARY OF THE FINDINGS ONCE THE STUDY IS COMPLETED

Email: _____

3. Information sheet and consent form for children with and without ASD

V4 18/08/17

INFORMATION SHEET FOR CHILDREN & YOUNG PEOPLE

Great Ormond Street Hospital for Children NHS Foundation Trust

UCL

Can you help us?

We are looking for some young people to take part in our research study. This page tells you a bit about the study and we would be really grateful if you could have a read and see if you'd like to take part. If you are not sure about any of the words or have any questions please ask the researcher or talk about it with a member of your family.

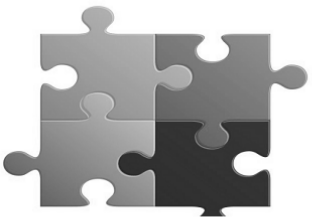
What is the study about?

We know that everyone thinks differently. Some people find it hard to learn new things and other people find it easy. Some people are good at solving puzzles whilst other people are good at telling stories. Most people have some things they are quite good at and others that they are not so good at.

We have designed some new games and puzzles that can help us to look at the different way young people think and do things. We would like you to give our games and puzzles a go to see if they work!

What will I need to do?

If you like the idea of taking part in our study then the researcher will visit you at your home, school or at Great Ormond Street Hospital. You will get to have a go at some games and puzzles which we hope you'll find fun.



We'll also ask the person who takes care of you to fill in some

INFORMATION SHEET FOR CHILDREN & YOUNG PEOPLE

questions about you. Things you are good at and things that you like doing.

As a small thank you for taking part in our study we will offer you a £5 voucher.

Why ask me?

We are asking you because we want to test out our puzzles and games on young people who are between 8 and 12 years old.

What will it be like to take part?

We hope our games will be fun but sometimes you might find them a bit tricky. Not everyone will be able to finish them all. If you get tired or need a rest then you can ask to stop.

Do I have to take part?

No - it is up to you and the person who looks after you. If you do want to take part, we will ask you and your parent/carer to tick and sign a form. If you change your mind that's OK, you just have to tell us and you can stop at any time. You do not have to take part in this study.

Will anyone know how I do?

Our study is confidential. This means that no one will know how well you did in the puzzles and games.

Questions?

If you have any questions or would like to talk more about taking part you can ask to speak to Polly (the researcher) or another member of the research team.

V4 18/08/17

INFORMATION SHEET FOR CHILDREN & YOUNG PEOPLE

Great Ormond Street Hospital for Children NHS Foundation Trust

UCL

- ☐ I know that I don't have to take part if I don't want to
- ☐ If I change my mind I can just tell my parent or Polly
- ☐ It's OK to ask my parent/carer some questions about me
- ☐ I am happy to take part in the games and puzzles twice if needed to

Please put a circle around No or Yes to tell us if you want to take part

No

Yes

Signed……

Please print your name……

Appendix L. Invitation Letters to Parents

Invitation letter 1 – for parents/carers of children with ASD

Invitation letter 2 – for parents/carers of typically developing children

1. Invitation letter to parents/carers of children with ASD

V5 18/08/17

Great Ormond Street Hospital for Children NHS
NHS Foundation Trust

UCL
Polly Cocker
Department of Clinical, Health and Educational Psychology
University College London
1-19 TORRINGTON PLACE, LONDON, WC1E 7HB

Tel: [redacted]
Email: [redacted]

Invitation for your child to take part in a study

Dear Parent / Guardian

We would like to invite you and your child to take part in some research that we are conducting at the Social Communication Disorders Clinic at Great Ormond Street Hospital. We are approaching you as you have previously participated in this study and have given permission to be contacted to take part in further research.

As you may know, we are developing tests specifically designed for children with Autism Spectrum Disorders (ASD) to measure abilities planning and organising behaviour. We want to find out whether the test works and is able to measure these things accurately. A more detailed explanation of the study can be found on the attached information sheet. Since you and your child took part, several new tasks have been developed and we would like to invite you to take part in these too.

This part of study will involve no more than one hour of yours and your child's time, during which they will be asked to do a number of new games and puzzles that aim to assess how they think, learn and remember things. These are designed to be as fun as possible. You will be asked to complete one questionnaire about your child's behaviour. This can be organised for a time and place of your choosing to minimise inconvenience to you and your child.

A small number of children and parents will be contacted a second time to arrange a shorter follow-up meeting. Only your child will take part in this second visit and will repeat the puzzles and games, which should take no longer than one hour. As a thank you your child will receive a £5 voucher.

We would be grateful if you could spare some time to read through the attached information sheet and speak with your child about whether or not they would be happy to

take part in the research. If you and your child are happy to take part then please contact Polly (the researcher) or return the enclosed consent form so that we can contact you. Contact details can be found at the bottom of this page.

Please do not hesitate to contact us should you have any questions or require further information.

Thank you for your time and for considering taking part in our research.

Polly Cocker
Trainee Clinical Psychologist
(Researcher)

Dr Will Mandy
Clinical Psychologist
(Chief Investigator)

2. Invitation letter to parents/carers of typically developing children

Great Ormond Street Hospital for Children NHS Foundation Trust

V6 18/08/17

[redacted] UCL
Polly Cocker
Department of Clinical, Health and Educational Psychology
University College London
1-19 TORRINGTON PLACE, LONDON, WC1E 7HB
Tel: [redacted]
Email: [redacted]

Invitation for your child to take part in a study

Dear Parent / Guardian

We would like to invite you and your child to take part in some research that we are conducting at Great Ormond Street Hospital and UCL. We are approaching you as you have previously participated in this study and might want to take part in further research.

As you may know, the aim of the research is to develop tests that hope to accurately assess planning and organisation in children with Autistic Spectrum Disorder (ASD). We are looking to recruit children that do not have a diagnosis of ASD to form part of our comparison group. A more detailed explanation of the study can be found on the attached information sheet. Since you and your child took part, several new tasks have been developed and we would like to invite you to take part in these too.

This part of study will involve no more than one hour of yours and your child's time, during which they will be asked to do a number of new games and puzzles that aim to assess how they think, learn and remember things. These are designed to be as fun as possible. You will be asked to complete a questionnaire about your child's behaviour. This can be organised for a time and place of your choosing to minimise inconvenience to you and your child.

As a thank you your child will receive a £5 voucher. A small number of children and parents will be contacted a second time to arrange a shorter follow-up meeting. Only your child will take part in this second visit and will repeat the puzzles and games, which should take no longer than one hour.

We would be grateful if you could spare some time to read through the attached information sheet and speak with your child about whether or not they would be happy to take part in the research. If you and your child are happy to take part then please contact

Polly (the researcher) or return the enclosed consent form so that we can contact you. Contact details can be found at the bottom of this page.

Please do not hesitate to contact us should you have any questions or require further information.

Thank you for your time and for considering taking part in our research.

Yours Sincerely,

Polly Cocker

Trainee Clinical Psychologist

(Researcher)

Dr Will Mandy

Clinical Psychologist

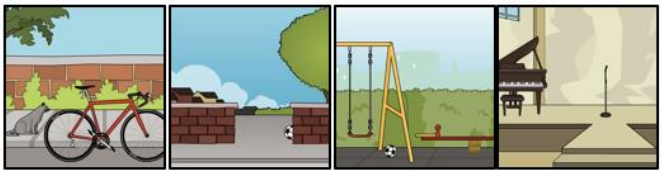
(Chief Investigator)

Appendix M. Instructions for Expert Raters in the Consensual Assessment Technique

Instructions

Child participants (aged 8 to 14) were given cartoon scenes forming a comic strip and asked to make up a story using the pictures. They were told to be as creative as they can. They had to use the comic strip to form the story but could add in other things too. They were all shown the same comic strip and given the following start to the story:

"Hina is riding her bicycle after school. She is planning to go to the park to play football."



Please rate the **creativity** of the following eight stories on a

1.0 – 7.0 scale (where 1 is the least creative and 7 the most creative).

Write your scores next to the participant numbers below.

Some things to note:

- Please use your own independent expert judgement of what is creative in this domain
- The stories should be rated relative to each other, not to any external standard
- Please use the full range of the scale

P09	
P11	
P16	
P21	
P29	
P31	
P57	
P62	