**Williams, David M. and Jarrold, Christopher (2013)** ***Assessing planning and set-shifting abilities in autism: Are experimenter-administered and computerised versions of tasks equivalent?*** **Autism Research, 6 . pp. 461-467.**

Williams, D.M., & Jarrold, C. (2013). Assessing planning and set-shifting abilities in autism: Are experimenter-administered and computerised versions of tasks equivalent? *Autism Research, 6,* 461-467
DOI: 10.1002/aur.1311

# Assessing Planning and Set-shifting Abilities in Autism: Are Experimenter-administered and Computerised Versions of Tasks Equivalent?

David M. Williams
Durham University

Christopher Jarrold
University of Bristol



Acknowledgements

This study was funded by an Economic and Social Research Council UK Grant awarded to David Williams and Christopher Jarrold (RES-000-22-4125).

Correspondence concerning this article should be addressed to David Williams, School of Psychology, Keynes College, University of Kent, Canterbury, Kent, CT2 7NP. Tel. +44 (0)1227 827652. Email: d.m.williams@kent.ac.uk;

**Scientific abstract**

The performance of individuals with autism spectrum disorder (ASD) on classic measures of executive functioning may suggest that people with this disorder are impaired only when tasks are administered by an experimenter, but not when the same tasks are computer-administered. This may imply that the underlying cause of apparent executive dysfunction in ASD is a diminished ability to engage with another person/comprehend what another person expects, rather than a diminution of the control processes that typically underpin EF task performance. However, this suggestion is limited because, to our knowledge, only one study has ever directly compared the equivalence of computer-administered and standard experimenter-administered versions of EF tasks among a common sample of individuals with ASD.

In the current study, 21 children with ASD and 22 age- and IQ-matched comparison participants completed, in counterbalanced order, computerised and manual versions of both a planning task and a cognitive flexibility/set-shifting task. Contrary to expectation, results indicated that participants with ASD were equally impaired in terms of the key dependent variable on standard and computerised versions of both tasks.

Practically, these results suggest that computer-administered and experimenter-administered versions of planning and set-shifting tasks are equivalent among individuals with ASD and can be used inter-changeably in studies of EF among this population. Theoretically, these results challenge the notion that poor performance on EF tasks among school-aged children with ASD is only the result of a limited ability to engage with a human experimenter/comprehend socially-presented rules.

**Keywords**

Autism; executive functioning; planning; set-shifting; Wisconsin Card Sorting Test; Tower of London task

**Assessing planning and Set-shifting abilities in ASD: Are experimenter-administered and Computer-administered Versions of Tasks Equivalent?**

Executive functioning (EF) is an umbrella term referring to a set of abilities (including planning, set-shifting, and inhibition) that allow the flexible control of action and that are underpinned by the frontal lobes. Autism spectrum disorder (ASD) is diagnosed on the basis of limitations in behavioural flexibility, alongside diminished social-communication (American Psychological Association, 2000). According to one theory, (at least some) ASD features are caused by executive dysfunction (e.g., Damasio & Maurer, 1978; Russell, 1996). One challenge to the EF theory of ASD has been the inconsistent findings regarding EF task performance among individuals with ASD. For example, the performance of individuals with ASD is by no means always diminished even on tasks measuring those aspects of EF (e.g., planning/cognitive flexibility) that (theoretically) underpin those aspects of behaviour that are diminished in ASD (e.g., behavioural flexibility)(for a review, see Kenworthy et al., 2008).

One relatively recent explanation for inconsistent findings of executive dysfunction in ASD concerns the *format* of the EF task that participants complete in each study. A detailed analysis of the performance of individuals with ASD across studies (see Kenworthy et al., 2008) suggests that people with this disorder may show diminished performance on classic measures of planning (e.g., the Tower of London/Hanoi task) and set-shifting (e.g., the Wisconsin Card Sorting Test) only when those tasks are experimenter-administered; when those same tasks are computer-administered, performance may not be diminished among individuals with ASD. Partly on the basis of this evidence, several researchers have suggested that poor performance among people with ASD on standard experimenter-administered does not reflect a diminution of the executive control processes that typically underpin planning and set-shifting tasks. Rather, among people with ASD, diminished theory of mind/"mentalising" is responsible for poor EF task performance (see Perner & Lang, 2002; White, 2013; see Hobson & Hobson, 2011 for a related theory).

According to this argument, of which there are several variants, the difficulty for people with ASD on experimenter-administered EF tasks is in inferring the experimenter's expectations for the task. Thus, whenever an individual with ASD is engaged in a socially-mediated EF task, in which they are required to respond to socially-presented rules that require the inference of implicit information, they will perform poorly; remove the social element of the task and they will perform relatively well. Thus, from a practical perspective, standard and computer-based EF tasks may not be equivalent among individuals with ASD (see Ozonoff, 1995). However, these arguments are limited substantially by the fact that findings of differential performance have been inferred from results *across* different samples.

To our knowledge, no study has directly compared performance on computerised and experimenter-administered versions of the ToL test (or related tests) of planning among the same sample of individuals with ASD. Equally, to our knowledge, no study has found evidence of selectively-diminished cognitive flexibility on an experimenter-administered set-shifting task, but undiminished performance on a computerised version of the same task, among a common sample of ASD participants[1].

We gave a group of children with ASD and a closely (age- and IQ-) matched comparison group computerised and manual/standard versions of classic planning and set-shifting tasks. If executive dysfunction in ASD reflects underlying difficulties with social engagement/mentalising, then we should observe significant Group (ASD/comparison) × Version (Computerised/manual) interaction terms, reflecting selectively poor performance among individuals with ASD in the manual versions of each task.

# Method

## Participants

Twenty-two TD comparison participants completed both computer and standard versions of both the Tower of London (ToL) task and the WCST. Twenty-one ASD participants completed both computer and standard versions of the ToL task. Of these 21 ASD participants, 20 also completed the computer and standard versions of the WCST. One additional ASD participant completed the computer and standard versions of the WCST, but not the ToL task. Thus, statistical analyses of ToL performance were based on a slightly different sample of ASD participants than were statistical analyses of WCST performance (i.e., 2/21 ASD participants were not overlapping in the analyses). Table 1 presents baseline characteristics of each group. For ease of reading, the characteristics of ASD participants described in the table are those who were included in the ToL analyses only[2].

ASD participants had received formal diagnoses of autistic disorder or Asperger's disorder, according to conventional criteria (American Psychiatric Association, 2000; World Health Organisation, 1992). Parents of participants with ASD completed the Social Responsiveness Scale (SRS; Constantino et al., 2003). In addition, among those parents of ASD participants who agreed to be contacted over the phone, 10 also completed the Developmental, Dimensional, and Diagnostic interview (3Di; Skuse et al., 2004). In each case, participants' scores were in the ASD range on one or both of these measures. Parents of comparison children also completed the SRS. Participants in the comparison group scored below the defined cut-off for ASD on the SRS.

## Procedures

### ToL task.

We employed a version of the classic ToL task (Shallice, 1982), which consisted of five coloured disks (each a different size) that could be arranged on three individual pegs. The aim of the task was to transform one arrangement of disks (the start state) into another arrangement (the goal state) by moving the disks between the pegs, one disk at a time. To achieve this in as few moves as possible, which is the aim of the task, requires efficient planning (e.g., Owen et al., 1990). In counterbalanced order, participants completed a computerised and a manual version of the ToL task, each version comprising 12 puzzles.

In the computerised version, puzzles were presented on a 14-inch laptop screen. The goal state was visible throughout each trial at the top of the screen. Directly underneath the goal state was the puzzle for participants to complete, which always began in the appropriate starting state.

In the manual version, puzzles were completed using a pre-made set of pegs (set into a Perspex base) and disks. For each puzzle, the experimenter organised the disks into the start state and presented the apparatus to the participant, along with a picture of the disks organised into the goal state. Whichever order participants completed the versions, they completed the second version approximately one week after they completed the first version.

Across versions, the sets of puzzles were matched for difficulty in terms of the minimum number of moves required to solve each (i.e., to reach the goal state from the start state). In each set of puzzles, two problems required a minimum of two moves to reach a solution, two required a minimum of three moves, two required a minimum of four moves, two required a minimum of five moves, and two required a minimum of seven moves. One puzzle in each set required a minimum of nine moves to complete and one puzzle required a minimum of 10 moves to complete. The use of each set of 12 puzzles was counterbalanced across the computerised and manual versions of the task.

The key measure of performance on the ToL task is the number of moves taken to complete the puzzles, with fewer moves indicating more efficient planning.

**WCST.**

We employed the modified WCST (Heaton, 1976). The modified WCST comprises four stimulus cards and 48 response cards that vary on three dimensions: type of shape (triangle, cross, circle, star), number of shapes (one to four), and colour of shapes (red, green, blue, yellow). Response cards that share more than one attribute (e.g., shared number and colour) are not included (unlike in the original WCST). In the standard (experimenter-administered) version, the four stimulus cards are placed separately, face-up in front of the participant, with the response cards in a pack, face down. The participant's task is to turn over the response cards one by one, placing each card below the stimulus card it "matches"; the task is to sort the response cards into categories according to one of these dimensions, as displayed in the stimulus cards. For each participant response, the experimenter provides positive or negative feedback, but does not tell the participant explicitly what the (arbitrarily) "correct" sorting strategy is. Thus, participants must infer the sorting strategy from the experimenter's feedback. After 6 consecutive cards have been sorted correctly, the experimenter tells the participants that the rule has now changed, and that they must sort the remaining response cards utilising a different rule.

In counterbalanced order, participants completed 48 trials of the standard (experimenter-administered) version and then 48 trials of a computerised version of the task (presented on a 14-inch laptop screen)[3]. In the computerised version, the stimulus cards remained at the top of the screen and a single response card appeared at the bottom of the screen. Alongside the response card appeared the message "Click on a pile above to sort the card". The participant was required to click the mouse on the stimulus card that they believed the response card "matched". After each choice, the computer provided positive/negative feedback by displaying the word "correct" or "incorrect" at the bottom of the screen. After 6 consecutive cards have been sorted correctly, the computer displayed a message that the rule had changed, and that the participant must sort the remaining response cards utilising a different rule. The order in which manual and computerised versions were completed was counterbalanced across participants. Participants completed the second version approximately one week after they completed the first version.

Key measures of performance on the WCST are number of perseverative errors and number of non-perseverative errors, with the former type of error indicating a specific difficulty with set-shifting.

## Results

### ToL Task Performance

Table 2 shows the total number of moves taken to complete all 12 puzzles in each of the manual and computerised versions of the ToL task among ASD and comparison participants. These data were subjected to a 2 (Group: ASD/comparison) × 2 (Version: Computerised/manual) mixed ANOVA. The ANOVA yielded a marginally significant main effect of Group, $F(1, 41) = 3.91$, $p = .055$, partial $\eta^2 = .09$, indicating that, overall, planning efficiency was significantly lower among ASD than comparison participants. Neither the main effect of Version, nor the interaction between Group and Version approached significance (all *p*s > .58, all partial $\eta^2$ values <.01). Thus, participants with ASD showed an overall deficit in planning, but this was not affected by the version of the task (manual versus computerised).

### WCST Performance

Figure 1 shows the mean number of perseverative and non-perseverative errors made on the manual and computerised versions of the task among ASD and comparison participants. These data were subjected to two 2 (Group: ASD/comparison) × 2 (Version: Computerised/manual) mixed ANOVAs. In the first ANOVA, number of non-perseverative errors was included at the dependent variable. In the second ANOVA, number of perseverative errors was included.

In the first ANOVA, no significant main effects or interactions emerged (all ps > .11, all partial $\eta^2$ values < .06). Thus, the number of non-perseverative errors made on each version of the task was equivalent, and there were no significant differences between the groups in any respect.

In the second ANOVA, the main effect of Version was significant, reflecting that more perseverative errors were made among both groups on the computerised version than on the manual version, $F(1, 41) = 14.64$, $p < .001$, partial eta squared = .26. The main effect of Group was also significant, reflecting that participants with

ASD made significantly more perseverative errors across both versions of the task than did comparison participants, $F(1, 41) = 6.56$, $p = .01$, partial eta squared = .14. The interaction between Group and Version was non-significant, $F(1, 41) = 2.31$, $p = .14$, partial $\eta^2 = .05$. Thus, in terms of perseverative errors, the groups were not affected significantly differently by the task version.

## Discussion

Contrary to expectation, we found no evidence that individuals with ASD are selectively impaired on experimenter-administered measures of planning or set-shifting. In none of the statistical analyses did a significant interaction between Group (ASD/comparison) and Version (computerised/manual) emerge;
In terms of planning, participants with ASD were equally impaired on the computerised and manual versions of the ToL task. In terms of set-shifting, perseverative errors were made significantly more frequently by ASD participants than by comparison participants on *both* versions of the task. Equally, there were no significant differences between the groups in number of non-perseverative errors made on *either* version of the WCST.

The computerised and manual versions of each task were essentially equivalent, apart from the requirement to interact with experimenter and respond to socially-presented rules in the manual versions. Thus, it seems to us reasonable to assume that if the primary difficulty with EF tasks among individuals with ASD was socially-mediated performance, then we should have observed deficits in the manual versions of each task only. The fact that we did not observe this pattern of performance among a common sample of individuals with ASD provides a challenge to the theory that social engagement/mentalising limitations play a significant role in poor EF task performance.

A main methodological implication of the current study is that, with respect to planning at least, computerised and experimenter-administered versions of the ToL task are equivalent among children with ASD.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (4th edition, text revised) (DSM-IV-TR).* Washington DC: American Psychiatric Association

Constantino, J. N., Davis, S. A., Todd, R. D., Schindler, M. K., Gross, M. M., Brophy, S. L., et al. (2003). Validation of a brief quantitative measure of autistic traits: Comparison of the social responsiveness scale with the autism diagnostic interview-revised. *Journal of Autism and Developmental Disorders, 33*(4), 427–433.

Hobson, R.P., & Hobson, J. (2011). Cognitive flexibility in autism A social-developmental account. In I. Roth & P. Rezaie (Eds), *Researching the Autism Spectrum*. Cambridge: Cambridge University Press (pp. 265-283).

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.

Owen, A. M., Downes, J. J., Sahakian, B. J., Polkey, C. E., & Robbins, T. W. (1990). Planning and spatial working memory following frontal-lobe lesions in man. *Neuropsychologia, 28*(10), 1021-1034.

Ozonoff, S. (1995). Reliability and validity of the Wisconsin Card Sorting Test in studies of autism. *Neuropsychology, 9*, 491-500.

Perner, J., & Lang, B. (2002). What causes 3-year-olds' difficulty on the dimensional change card sorting task? *Infant and Child Development, 11*(2), 93-105.

Russell, J. (1996). Agency: Its role in mental development. Hove, UK: Taylor and Francis.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London B Series, 298*(1089), 199-209.

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

White, S. J. (2013). The triple I hypothesis: Taking another('s) perspective on executive dysfunction in autism. *Journal of Autism and Developmental Disorders, 43,* 114-121.

World Heath Organisation. (1993). *International classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines (10th edn.)* Geneva, Switzerland: World Heath Organisation.

## Footnotes

1. A study by Ozonoff (1995) is often cited as having found selectively-diminished performance on an experimenter-administered version of the WCST, but not on a computerised version of the WCST among the same sample of individuals with ASD. However, matters are not that straightforward. In Ozonoff's Study 2, (n = 10) ASD and (n = 11) control participants were given a manual version of the WCST at time 1 and then a computer version of the WCST one year later (thus, order of version completion was not counterbalanced, which limits interpretation of results); in fact, no between-group differences in cognitive flexibility were found on *either* version of the task. In Ozonoff's Study 3, 24 children with ASD and 24 matched controls took part. Twelve participants from each diagnostic group undertook a computerised version of the WCST and 12 undertook a manual version. Results were that highly significant group differences were found on the manual version, whereas only marginally significant differences were observed on the computer version. However, this was clearly not a truly *within*-subjects design. Indeed, no evidence was provided that the *sub*-samples of ASD and comparison children who undertook the each version of the task were matched for age and IQ. Thus, results should be treated with caution.

2. The 21 ASD participants who were included in the WCST analyses (of which 20 were also included in the ToL analyses, of course) were also well-matched on all variables with comparison participants. The $p$ values and $d$ values associated with between-group comparisons of baseline variables in the WCST sample were as follows: Age: $p = .35$, $d = 0.30$; VIQ: $p = .78$, $d = 0.09$; PIQ: $p = .51$, $d = 0.20$; SRS raw score: $p < .001$, $d = 3.34$.

3. Aspects of the data from the computerised WCST have been reported in a separate study (study details not included in anonymised manuscript), in which performance on the WCST was used as a cognitive correlate of an ability other than set-shifting.

*Table 1*: *Participant characteristics*

| | Group | | | | |
|---|---|---|---|---|---|
| | ASD (n = 21 ) | TD (n = 22) | t | p | Cohen's *d* |
| Age | 10.45 (2.10) | 10.61 (1.30) | 0.30 | .77 | 0.09 |
| VIQ[a] | 103.29 (18.04) | 105.55 (13.25) | 0.47 | .64 | 0.14 |
| PIQ[a] | 110.24 (16.41) | 107.18 (13.03) | 0.68 | .50 | 0.21 |
| SRS[b] Raw Score | 111.67 (25.63) | 33.00 (22.04) | 10.81 | < .001 | 3.30 |

[a]Established using the Wechsler Abbreviated Scale of Intelligence (Psychological Corporation, 1999);
[b]Social Responsiveness Scale

Table 2:

| | Group | |
|---|---|---|
| | ASD | TD |
| Total # moves: Computer version | 69.43 (7.89) | 66.86 (6.56) |
| Total # moves: Manual version | 69.95 (7.94) | 65.86 (5.91) |
| Total # moves: Averaged across versions | 69.70 (6.17) | 66.36 (4.82) |

Note: The more moves taken, the poorer the planning performance

Figure 1: Mean number of non-perseverative and perseverative errors made by ASD and comparison participants on the computerised and manual versions of the WCST (error bars represent 1 *SEM*)