The effect of self-monitoring on Wisconsin Card Sorting Test performance in euthymic patients with bipolar disorder: a pilot study

<u>Short Title</u>: Self-monitoring in bipolar disorder

Lucy J. Robinson [1,*], *BA (Hons), BSc (Hons), PhD, DClinPsy, AFBPsS*
Academic Clinical Fellow (Psychology)

John M. Gray [1], *MA (Hons), MPhil, PhD*
Visiting Researcher (Psychology)

I. Nicol Ferrier [1,2], *BSc (Hons), MD (Hons), FRCP, FRCPsych*
Emeritus Professor of Psychiatry

Peter Gallagher [1,2], *BSc (Hons), MPhil, PhD, AFBPsS*
Lecturer in Neuropsychology

Affiliations:

[1] *Institute of Neuroscience (Academic Psychiatry), Newcastle University, UK.*

[2] *Newcastle University Institute for Ageing, Newcastle University, UK.*

*Corresponding author:

Dr Lucy Robinson, Academic Psychiatry, Campus for Ageing and Vitality, Newcastle University, Westgate Road, Newcastle upon Tyne, NE4 5PL, UK

tel: +44 (0)191 208 1392   e-mail: Lucy.Robinson2@newcastle.ac.uk

Word count (body of text): 3,891   Abstract: 198

Abstract

Objectives: Euthymic patients with bipolar disorder (BD) show cognitive impairment in measures of executive function. The use of non-pharmacological strategies to aid cognitive function has not been widely-explored in BD. In schizophrenia, concomitant verbalisation (*self-monitoring*) while performing tests of executive function improved performance. The present pilot study assesses the effects of self-monitoring during performance of the Wisconsin Card Sorting Test (WCST) compared to standard administration in patients with BD.

Method: Thirty-six euthymic patients with BD and 42 healthy controls participated. Twenty patients with BD and 20 controls received standard administration and 16 patients and 22 controls were asked to self-monitor while performing the test.

Results: ANCOVA revealed a significant 'group by administration' interaction. Patients who received the standard administration showed significant impairment relative to healthy controls on several WCST indices (trials administered: p=0.012, η_p^2=0.17; trials to first category: p=0.046, η_p^2=0.11; failure to maintain set: p=0.003, η_p^2=0.23). BD patients who self-monitored performed significantly better than patients receiving the standard administration (trials to first category: p=0.020, η_p^2=0.17) and showed no significant differences in performance compared to controls.

Conclusion: Self-monitoring deserves further investigation as a tool that may be helpful for patients with BD. Further exploration of the utility, generalisability and stability of the effects of self-monitoring is needed.

Introduction

There is substantial evidence that patients with bipolar disorder (BD) show cognitive impairment in the euthymic state. Several meta-analyses have indicated moderate to large deficits in verbal memory, and moderate impairment in visual memory, executive function, psychomotor speed and attention (Arts, Jabben, Krabbendam, & van Os, 2008; Bora, Yucel, & Pantelis, 2009; Kurtz & Gerraty, 2009; Robinson et al., 2006; Torres, Boudreau, & Yatham, 2007). The reasons behind cognitive impairment in BD are not fully understood and there has been a paucity of research investigating whether cognitive dysfunction can be improved in euthymic patients (Deckersbach et al., 2009; Khan et al., 2004).

Until recently the potential role of pharmacotherapy in managing cognitive impairment in BD had received more attention than non-pharmacological approaches (Burdick, Braga, Goldberg, & Malhotra, 2007). There is now growing interest in investigating non-pharmacological methods (Bonnin, Torrent, Vieta, & Martinez-Aran, 2014; Bowie, Gupta, & Holshausen, 2013; Demant, Almer, Vinberg, Kessing, & Miskowiak, 2013; Dickstein, Cushman, Kim, Weissman, & Wegbreit, 2015; Sanches, Bauer, Galvez, Zunta-Soares, & Soares, 2014). One of the first studies in the area investigated a 14-session cognitive remediation programme in an open trial in 18 patients (Deckersbach et al., 2010). The programme had the dual aims of reducing inter-episode symptoms and improving cognitive function in euthymic BD patients. Significant decreases in depressive symptoms, work dysfunction, and (questionnaire-rated) executive dysfunction were evident at the end of the three month treatment block and maintained at follow-up. A more recent randomized controlled study examined the efficacy of weekly group-based cognitive remediation treatment over 12 weeks and found the treatment was associated with improvement in 'sharpness/mental acuity' and quality of life, but minimal change in objective cognitive function. In the largest study to date in BD (Sole et al., 2015; Torrent et al., 2013), a 21-week a functional remediation programme was associated with greater improvement in psychosocial function than treatment as usual, although this failed to reach statistical significance and again there was no significant effect of treatment on a broad battery of neurocognitive tests. This is notably in contrast to the positive neurocognitive effects of these programmes

shown in schizophrenia (McGurk, Twamley, Sitzer, McHugo, & Mueser, 2007). Therefore, while the results of such approaches are promising, it remains unclear whether alternative approaches may also be needed that impact cognitive function.

Of the experimental interventions used in schizophrenia to improve cognition, some have involved enhanced instructions or even telling participants how to perform the task (e.g. (Green, Ganzell, Satz, & Vaclav, 1990)). The task most-commonly used for this has been the Wisconsin Card Sorting Test (WCST). This executive function test requires participants to sort cards according to a principle that is not given to them. They must ascertain the principle using trial and error, using the feedback about correct and incorrect sorts to deduce the rule, and then adjust behaviour when the rule changes. This approach can be criticised in that, if the aim of an intervention is to identify whether patients can perform the task under optimal conditions, the fundamental nature of the task needs to remain unchanged (Goldberg & Weinberger, 1994). Part of the standard administration of the WCST involves withholding the fundamental principle of the task; informing participants of the rules alters the task demands markedly. One approach that is not affected by this limitation to the same degree is *self-monitoring* – that is, asking a participant to verbalise their reasoning for the way in which they are sorting the cards. Asking patients with Schizophrenia to self-monitor has been shown to improve performance on the WCST (Perry, Potterat, & Braff, 2001). Interestingly, self-monitoring seems to enhance performance on tasks requiring more complex (executive) processes, but impairs performance on simpler tasks (Harvey, Galletly, Field, & Proeve, 2009). It may therefore be possible to identify the specific processes through which self-monitoring operates by identifying the performance indices most enhanced by it.

Performance on the WCST has been shown to be impaired in patients with BD (Bourne et al., 2013; Robinson et al., 2006). It lends itself to use in a self-monitoring context, as it involves complex processing (involving concept formation, response inhibition, set-shifting and memory), it is self-paced with minimal instructions, and it produces a diverse range of outcome measures. The present pilot study aims to explore the effect of asking euthymic patients with Bipolar Disorder to self-monitor while performing the Wisconsin Card Sorting Test. Performance will be compared against patients who received the standard

administration, as well as two non-patient comparison groups who also received either the standard or self-monitoring version. It is hypothesized that self-monitoring in euthymic patients will lead to a smaller decrement in performance than seen under standard administration.

Materials and Methods

Participants

Thirty seven patients with BD participated. Patients were recruited from secondary and tertiary psychiatric services throughout the North East of England. Inclusion criteria comprised aged between 18-65, a DSM-IV diagnosis of bipolar disorder (confirmed by a psychiatrist using the mood disorders section of the Structured Clinical Interview for DSM-IV (First, Spitzer, Williams, & Gibbon, 1995)) and currently euthymic (scored ≤7 on both the 17-item Hamilton Depression Rating Scale (Hamilton, 1960) and the Young Mania Rating Scale (Young, Biggs, Ziegler, & Meyer, 1978) which was prospectively verified for one month before testing (for full details see (Thompson et al., 2005))). Exclusion criteria comprised current alcohol misuse or dependence, history of head injury with loss of consciousness lasting more than 5 minutes, known neurological illness or relevant major medical illness, ECT within the last 6 months, and learning disability or difficulty with fluent use of the English language. Patients were not excluded for use of psychotropic medication or for comorbid anxiety disorders (comorbidities were assessed using the Mini-International Neuropsychiatric Interview (Sheehan et al., 1998)).

Forty two healthy controls were recruited via local advertisements. They were subject to the same exclusion criteria as the patient sample with the addition of no personal history of psychiatric illness and no family history of BD in a first degree relative. The study was reviewed and given ethical approval by Newcastle Research Ethics Committee. All participants gave written informed consent to participate.

Materials

All participants were administered a computerised version of the 128-card version of the Wisconsin Card Sorting Test (Heaton, Chelune, Talley, Kay, & Curtiss, 1993). The test involves matching virtual cards to one

of four base cards using an undisclosed sorting principle. The participant is informed after each sort whether the card was sorted correctly or incorrectly and they must deduce the sorting principle from this feedback. Unknown to the participant, the sorting category changes after ten consecutive correct sorts and the participant must switch from the established sorting principle and deduce the new rule. After ten consecutive correct responses to the new category, the rule switches again and so on until either five shifts in category are made successfully or until all 128 cards have been dealt. The task was administered on a PC running Windows XP and responses were made via a 15" CTX resistive touch screen monitor.

There are a number of outcome measures derived from this task that were examined in the current study: total number of trials administered, number of correct responses, number of trials until the first category is achieved (i.e. ten consecutive correct responses), total perseverative errors, total non-perseverative errors, number of categories achieved, number of conceptual level responses (responses that matched one of the four base cards on at least one possible sorting principle), failure to maintain set (the number of times after five consecutive correct sorts to a category that the participant fails to achieve the category), and an index measuring 'learning to learn' (the average decrease in errors made with each successive category). More outcome measures can be derived from this task, but we selected those measures with relatively greater mathematical independence from one another (i.e. indices that contain other indices were not analysed, such as total errors, which is the sum of perseverative and non-perseverative errors; although it is acknowledged that some overlap remains in the measures chosen, e.g. total trials administered includes perseverative and non-perseverative errors).

In the self-monitoring version of the task, the instructions were modified to ask the participant after each card sort "Why did you put that there?". The aim was to encourage the participant to 'think aloud' and to reason through their choices. The question was asked in a friendly manner and the participant was not given any further verbal prompting or feedback other than that offered automatically by the computer. Participants' responses to the question were not recorded.

Procedure

This was a pilot study embedded within a larger study. Patients and controls were recruited serially; all of the standard administration group completed the study before the self-monitoring participants. Demographic differences between groups that resulted from this recruitment method were controlled for in analyses. The Wisconsin Card Sorting Test was administered as part of a larger battery (details of the general executive measures are shown in supplementary table 1). Overall, 21 patients and 20 controls received the standard administration of the task, and 16 patients and 22 controls were asked to self-monitor.

Data Analysis

Data were analysed using SPSS version 21.0. The four groups (control standard, control self-monitoring, bipolar standard, bipolar self-monitoring) were compared using ANCOVA with two between subjects factors each with two levels (*group*: patient or control; *condition*: standard administration or self-monitoring) with years of education as a covariate. In order to preserve statistical power, depressive symptoms were not entered as a covariate in these analyses, but were included in subsequent pairwise comparisons (see below). The outcome of interest was the group x condition interaction, although main effects of group and condition are reported in table 2 for completeness. A statistically significant interaction was followed up with pairwise comparisons between each of the groups using ANCOVA to covary for years of education as well as mood when applicable. For comparisons between patients and controls, clinician-rated scores were used as the covariate. For comparisons between the patient groups, self-rated depression scores were selected as the more appropriate covariate, as there were significant differences in depression between these groups on these ratings. Mood was not included as a covariate for comparisons between control groups, as the groups did not differ on these ratings. A significance level of $p \leq 0.05$ was used. A Bonferroni correction was considered too conservative (Shi, Pavey, & Carter, 2012) for the pilot nature of the study and was not implemented. Effect sizes were calculated using Cohen's d. Effect sizes for indices where a higher score indicated worse performance (e.g. error or reaction time indices) were converted (multiplied by -1) such that a positive effect size always indicates a poorer performance by the patient group. Conventions for Cohen's d are that 0.2<d<0.5 is a small effect, 0.5≤d<0.8 is a medium effect,

and d≥0.8 is a large effect. Some demographic variables had missing data owing to recording errors (maximum of n=5 across all participants). No data was missing for the task. However some of the mood ratings were missing, which is reflected in the reduced degrees of freedom in ANCOVA analyses including mood as a covariate.

Results

Demographics and clinical characteristics

The demographics of the groups are shown in **Error! Reference source not found.**. There were no significant differences between the groups in age, gender, handedness, weekly alcohol consumption, proportion of smokers, premorbid IQ, clinician-rated manic symptoms or self-rated manic symptoms (all p>0.2). There was an overall difference between the groups in years of formal education ($F_{3,74}$=3.11, p=0.031), and post hoc LSD tests indicated that the controls who received standard administration had significantly more education than the bipolar patients who received standard administration (p=0.021) and the control patients who received the self-monitoring intervention (p=0.015). Years of education was therefore included as a covariate in subsequent analyses. There were significant between-group differences in clinician-rated and self-rated depressive symptoms ($F_{3,74}$=13.62, p<0.001; $F_{3,73}$=18.77, p<0.001 respectively). As is often found even in clinically euthymic patients, post-hoc tests indicated both patient groups had significantly higher clinician-rated depressive symptoms than the control groups (all p<0.001), but were not significantly different to one another (p=0.175). Patients in the self-monitoring group had significantly higher self-rated depressive symptoms than all other groups (all p<0.001). There were no other statistically significant differences between groups for self-rated depressive symptoms.

WCST performance and self-monitoring

Performance of the groups on the Wisconsin Card Sorting Test is shown in **Error! Reference source not found.**. The 'group x condition' interaction was statistically significant for four of the test indices: number of

trials administered ($F_{1,73}$=4.92, p=0.030, η_p^2=0.06), total trials correct ($F_{1,73}$=4.29, p=0.042, η_p^2=0.06), trials to first category ($F_{1,73}$=7.21, p=0.009, η_p^2=0.09), and failure to maintain set ($F_{1,73}$=8.03, p=0.006, η_p^2=0.10). These statistically significant differences for the interaction effect remained for three of the four indices after controlling separately for self-report depressed mood (number of trials: $F_{1,67}$=7.46, p=0.008, η_p^2=0.10; trials to 1st category: $F_{1,67}$=11.15, p=0.001, η_p^2=0.14; failure to maintain set: $F_{1,67}$=7.63, p=0.007, η_p^2=0.10) and clinician-rated depressed mood (number of trials administered: $F_{1,68}$=5.21, p=0.026, η_p^2=0.07; trials to 1st category: $F_{1,68}$=7.78, p=0.007, η_p^2=0.10; failure to maintain set: $F_{1,68}$=8.84, p=0.004, η_p^2=0.12). The differences for total trials administered was no longer statistically significant after covarying for self- rated mood ($F_{1,67}$=2.44, p=0.123, η_p^2=0.04), although a trend remained after covarying for clinician-rated mood ($F_{1,68}$=3.66, p=0.06, η_p^2=0.05). There were no significant interaction effects for the other indices (perseverative errors, non-perseverative errors, conceptual level responses, or categories achieved). These are illustrated in figure 1.

Pairwise contrasts

Comparing performance under standard administration (covarying for years of education and clinician-rated depressive symptoms) BD patients were significantly worse than controls on each of the three indices (trials administered: $F_{1,35}$=7.01, p=0.012; η_p^2=0.17, 1-β=0.73); trials to first category: $F_{1,35}$=4.30, p=0.046; η_p^2=0.11, 1-β=0.52); failure to maintain set: $F_{1,35}$=10.27, p=0.003; η_p^2=0.23, 1-β=0.88). Effect sizes for the differences are shown in table 3.

Under the condition of self-monitoring, there was no significant performance difference between BD patients and controls (with the same covariates). There were no significant differences between these two groups (trials administered: $F_{1,32}$=0.57, p=0.456; η_p^2=0.02, 1-β=0.11); trials to first category: $F_{1,32}$=0.82, p=0.373; η_p^2=0.03, 1-β=0.14); failure to maintain set: $F_{1,32}$=0.94, p=0.34; η_p^2=0.03, 1-β=0.16).

The two patient groups were compared against one another (again including years of education and self-rated depressive symptoms as covariates). Self-monitoring was associated with significantly better

performance for the trials to first category ($F_{1,34}$=6.05, p=0.020; η_p^2=0.17, 1-β=0.66). Differences on the trials administered did not reach statistical significance ($F_{1,34}$=2.58, p=0.11; η_p^2=0.08, 1-β=0.34) and a trend was observed in failure to maintain set ($F_{1,34}$=3.19, p=0.084; η_p^2=0.10, 1-β=0.41).

Finally, the two control groups were compared against one another (covarying for years of education – as there were no differences between the groups in mood ratings, this was not entered as a covariate in these analyses to preserve statistical power). In the control groups, unlike the patient groups, self-monitoring was associated with significantly poorer performance with more trials administered ($F_{1,38}$=4.23, p=0.047; η_p^2=0.10, 1-β=0.52). There was a trend for poorer performance on the other two indices (trials to first category: $F_{1,38}$=4.01, p=0.052; η_p^2=0.10, 1-β=0.50; failure to maintain set: $F_{1,38}$=3.97, p=0.054; η_p^2=0.10, 1-β=0.49).

General executive function

Data from tests assessing broader executive/working memory performance is included in supplementary table 1. ANOVA main effects of group (p<0.05) were observed for Category fluency and the Digit Symbol Substitution Test (DSST). Post hoc analyses revealed that in category fluency, controls in the standard condition performed significantly more poorly than BD patients (p<0.05) although there were no significant differences between the two control groups or the two patient groups. For the DSST, the controls in the self-monitoring condition performed significantly better than both BD patient groups (p<0.05), although again there were no significant differences between the two control groups or the two patient groups.

Discussion

The aim of the present study was to investigate the effect of a self-monitoring intervention compared to standard administration of the WCST in euthymic BD. The results indicate that the hypothesised

performance deficit in BD compared to healthy controls was only observed under standard administration – no group differences were observed under self-monitoring. Direct comparison within BD revealed that self-monitoring was associated with significantly improved 'trials to first category' and a trend towards a reduced 'failure to maintain set' compared to standard administration. The effect sizes indicated that the standard administration was associated with medium to large effect size differences when contrasted with control group performance, whereas the self-monitoring group showed small to medium effect size differences when compared against controls.

It is not clear from the present findings why or how self-monitoring is beneficial in patients. It is possible that self-monitoring encourages recruitment of cognitive resources that are necessary for good performance but that for some reason are not being recruited automatically. Or it may be that self-monitoring recruits processes that are otherwise intact to 'shore up' or compensate for those which are lacking or dysfunctional (we have previously suggested the notion of verbal scaffolding of impaired visuo-spatial functions in depressed BD; see Gallagher et al., 2015). Interestingly the effects of the self-monitoring intervention were most pronounced for failure to maintain set and trials to first category. Previous studies have suggested that self-monitoring may exert effects via reducing distractibility and increasing focus on the task at hand (Perry et al., 2001). The test indices most affected in the present study suggest that patients were helped to remember which category they were sorting by and were able to form the concepts more efficiently. It should also be noted that the self-monitoring process led to a reduction in performance in healthy controls (a strong trend in several indices, but significant for 'trials administered'). We can only speculate that in this case the act of verbalisation may have elicited some degree of distraction or interference with the typically-adopted, efficient processes. In one previous study, Rossell & David (1997) examined the effects of seven different versions of the WCST (including a range of alternative stimuli and training techniques) on performance of patients with schizophrenia and older healthy controls. The authors found that significant performance improvements, particularly in the schizophrenia group, were observed when verbalizing responses or with category modification. However, one manipulation – teaching categories – actually led to a significant increase in perseverative errors in controls but a

significant improvement in patients. Although this was the only outcome measure where this occurred, there were other manipulations that led to no change in either group. It would be of particular interest for future studies to audio-record and transcribe the methods adopted by participants in such studies (under non-directed, free-verbalization, as used in the present study), to better understand the approaches that aid performance in patients and controls, particularly those that may lead to opposite effects.

A more general mechanism by which self-monitoring may be supporting performance is by creating a 'social accountability' for the decisions made, enforcing reasoning to become conscious and deliberate. It is difficult to test this conjecture, as it impossible to know the reasoning processes underway in the absence of verbalisation. However, it is possible that by requiring verbalisation, thoughts which may otherwise remain unstructured are structured so as to make sense to someone else, and that this in itself is helpful for ongoing processing. In the developmental psychology literature, the observation that young children frequently talk aloud to themselves led to the conjecture that this ego-centric speech is used functionally for problem solving (Vygotsky, 1962). The rate of ego-centric speech almost doubles when children encounter a difficulty in a task, indicating that it "becomes an instrument of thought in the proper sense – in seeking and planning the solution of a problem." (p. 16 (Vygotsky, 1962)) It is possible that encouraging the reactivation of this strategy through self-monitoring may also reap benefits for adults with BD faced with complex tasks involving multiple cognitive processes. Interestingly, the same effect was not evident in the control sample; self-monitoring was associated with poorer performance. This was unexpected and potentially suggests that controls experienced the self-monitoring as additional cognitive load which hindered rather than enhanced their performance. This highlights the importance of developing specific strategies for the target population - strategies that help individuals with BD may not be generally helpful to all (and vice versa).

Further studies are needed using a hypothesis-driven approach either using a task or tasks with outcome indices that are closely linked to known underlying cognitive processes, or taking advantage of the factor structure of the WCST to identify the way in which task performance is supported by self-

monitoring. But even in the absence of this understanding, if the effects can be shown to lead to improvement in a test-retest design and to generalise to other tasks or everyday activities, then it shows promise as a means of remediating some of the cognitive deficit in people with BD. Interestingly, studies are currently underway which are assessing the efficacy of combining cognitive remediation with pharmacological agents (such as the NMDA receptor partial agonist, d-cycloserine) in the treatment of cognitive deficits in BD (Breitborde et al., 2014).

Do the current findings have any implications for understanding the nature of the underlying deficit in BD? It is tempting to conclude, as has been done for patients with schizophrenia, that a positive impact of psychological interventions on cognitive function militates against the likelihood that the deficit is due to immutable structural (or functional) damage or a fundamentally biological cause (Summerfelt, Alphs, Funderburk, Strauss, & Wagman, 1991). However, caution is required in accepting this possibility. Authors have suggested that the fact performance on tests such as the WCST can be improved using short-term experimental manipulations may stem in part from the fact that all cognitive measures are an imperfect estimate of the underlying functions (Goldberg & Weinberger, 1994). They are not only subject to measurement error, but may also suffer "functional redundancy" (p. 294 (Goldberg & Weinberger, 1994)). That is, a (statistically) normal result may be possible even if the underlying cognitive systems used by the task are damaged as the task does not require the underlying circuitry to be used to its full capacity. In other words, improved or normalised performance does not mean underlying deficits have been eliminated. They may simply have been obviated or sufficiently 'scaffolded' to enable a 'normal' performance from a system that still remains dysfunctional. For example, in previous work in bipolar depression we have found verbal scaffolding may support performance on some visuo-spatial processes that are amenable to verbalisation or verbal coding (Gallagher, Gray, & Kessels, 2015; Gallagher, Gray, Watson, Young, & Ferrier, 2014). Although of significant relevance for those trying to clearly delineate the exact nature of the cognitive deficit in BD (and potentially identify a biological substrate), it may be of less relevance to the pragmatic question of whether simple remediation strategies can be useful for patients to find a way around the difficulties they face.

It is also necessary to bear in mind that it is not the case that self-monitoring normalised performance on the task. Firstly, the cross-sectional nature of the data does not permit this conclusion. This was not a within-subjects design, and although the between-subjects design avoided the contamination from practice effects or carry-over effects that would arise in test-retest or crossover designs, it remains impossible to determine whether performance changed from impaired to unimpaired in the patients who received the self-monitoring intervention. Secondly, in general the self-monitoring group showed no significant differences from either the control group or the patients who received standard administration (the aforementioned indices excepted). This suggests their performance was intermediate between the two groups rather than normalised to the point that there was a consistent significant performance advantage over and above patients who received standard administration. Also, the effect sizes observed in the patients who received the modified administration contrasted with controls who received the standard administration still remained in the small to moderate range. In a larger sample, these differences would have reached statistical significance, and would result in a deficit of a similar magnitude to that generally found for verbal fluency tasks or short term memory measures in euthymic patients with BD (Robinson et al., 2006).

Some of the limitations of the current findings have already been mentioned, most notably the cross-sectional design. Additionally patients were not randomised to standard administration or self-monitoring; the groups were tested serially. Fundamental differences in the groups may thus have accounted for part of the findings rather than merely the altered administration itself, although where group differences were known covariates were included in statistical analyses to control for them. The small sample size may also have affected statistical power and replication in a larger sample is necessary. Similarly, to preserve statistical power, p-values were not adjusted for multiple tests and differences on some indices would not remain statistically significant after using, for example, a Bonferroni correction. Finally, recording the responses of the participants would have provided valuable information for understanding how participants reasoned through the task. In general the self-monitoring request was

well-received, but there may be alternative ways of encouraging verbalisation which do not involve

repetitive questioning.

Despite these limitations, there is nonetheless evidence that a straightforward and simple strategy

was associated with a smaller performance decrement on WCST. This is an optimistic message suggesting

that there may be strategies that are useful for patients to lessen the impact of their cognitive difficulties.

Replication of these findings randomising patients to groups is required, as well as further work

incorporating other measures to clarify to what extent it can be useful for patients in tackling difficulties

arising from cognitive dysfunction in everyday life.

References

Arts, B., Jabben, N., Krabbendam, L., & van Os, J. (2008). Meta-analyses of cognitive functioning in euthymic bipolar patients and their first-degree relatives. *Psychological Medicine, 38*(6), 771-785

Bonnin, C. M., Torrent, C., Vieta, E., & Martinez-Aran, A. (2014). Restoring functioning in bipolar disorder: functional remediation. *Harv Rev Psychiatry, 22*(6), 326-330. doi: 10.1097/hrp.0000000000000062

Bora, E., Yucel, M., & Pantelis, C. (2009). Cognitive endophenotypes of bipolar disorder: a meta-analysis of neuropsychological deficits in euthymic patients and their first-degree relatives. *Journal of Affective Disorders, 113*(1-2), 1-20

Bourne, C., Aydemir, Ö., Balanzá-Martínez, V., Bora, E., Brissos, S., Cavanagh, J. T. O., . . . Goodwin, G. M. (2013). Neuropsychological testing of cognitive impairment in euthymic bipolar disorder: an individual patient data meta-analysis. *Acta Psychiatrica Scandinavica, 128*(3), 149-162. doi: 10.1111/acps.12133

Bowie, C. R., Gupta, M., & Holshausen, K. (2013). Cognitive remediation therapy for mood disorders: rationale, early evidence, and future directions. *Can J Psychiatry, 58*(6), 319-325

Breitborde, N. J., Dawson, S. C., Woolverton, C., Dawley, D., Bell, E. K., Norman, K., . . . Moreno, F. (2014). A randomized controlled trial of cognitive remediation and d-cycloserine for individuals with bipolar disorder. *BMC Psychol, 2*(1), 41. doi: 10.1186/s40359-014-0041-4

Burdick, K. E., Braga, R. J., Goldberg, J. F., & Malhotra, A. K. (2007). Cognitive dysfunction in bipolar disorder: future place of pharmacotherapy. *CNS Drugs, 21*(12), 971-981. doi: 21122 [pii]

Deckersbach, T., Nierenberg, A. A., Kessler, R., Lund, H. G., Ametrano, R. M., Sachs, G., . . . Dougherty, D. (2009). Cognitive Rehabilitation for Bipolar Disorder: An Open Trial for Employed Patients with Residual Depressive Symptoms. *CNS Neurosci Ther*. doi: CNS110 [pii]

10.1111/j.1755-5949.2009.00110.x [doi]

Deckersbach, T., Nierenberg, A. A., Kessler, R., Lund, H. G., Ametrano, R. M., Sachs, G., . . . Dougherty, D. (2010). Cognitive rehabilitation for bipolar disorder: An open trial for employed patients with residual depressive symptoms. *CNS Neurosci Ther, 16*(5), 298-307. doi: 10.1111/j.1755-5949.2009.00110.x

Demant, K. M., Almer, G. M., Vinberg, M., Kessing, L. V., & Miskowiak, K. W. (2013). Effects of cognitive remediation on cognitive dysfunction in partially or fully remitted patients with bipolar disorder: study protocol for a randomized controlled trial. *Trials, 14*, 378. doi: 10.1186/1745-6215-14-378

Dickstein, D. P., Cushman, G. K., Kim, K. L., Weissman, A. B., & Wegbreit, E. (2015). Cognitive remediation: potential novel brain-based treatment for bipolar disorder in children and adolescents. *CNS Spectr*, 1-9. doi: 10.1017/s109285291500036x

First, M. B., Spitzer, R. L., Williams, J. B. W., & Gibbon, M. (1995). *Structured Clinical Interview for DSM-IV (SCID-IV)*. New York: Biometrics Research Department, New York State Psychiatric Institute.

Gallagher, P., Gray, J. M., & Kessels, R. P. C. (2015). Fractionation of visuo-spatial memory processes in bipolar depression: a cognitive scaffolding account. *Psychological Medicine, 45*, 545-558

Gallagher, P., Gray, J. M., Watson, S., Young, A. H., & Ferrier, I. N. (2014). Neurocognitive functioning in bipolar depression: a component structure analysis. *Psychol Med, 44*(5), 961-974. doi: 10.1017/s0033291713001487

Goldberg, T. E., & Weinberger, D. R. (1994). Schizophrenia, training paradigms, and the Wisconsin Card Sorting Test redux. *Schizophrenia Research, 11*(3), 291-296

Green, M. F., Ganzell, S., Satz, P., & Vaclav, J. F. (1990). Teaching the Wisconsin Card Sorting Test to schizophrenic patients. *Archives of General Psychiatry, 47*(1), 91-92

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology Neurosurgery and Psychiatry, 23*, 56-62

Harvey, K. E., Galletly, C. A., Field, C., & Proeve, M. (2009). The effects of verbalisation on cognitive performance in schizophrenia: A pilot study using tasks from the Delis Kaplan Executive Function System, 1-9.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test Manual*. Lutz, FL: Psychological Assessment Resources Inc.

Khan, A., Ginsberg, L. D., Asnis, G. M., Goodwin, F. K., Davis, K. H., Krishnan, A. A., & Adams, B. E. (2004). Effect of lamotrigine on cognitive complaints in patients with bipolar I disorder. *Journal of Clinical Psychiatry, 65*(11), 1483-1490

Kurtz, M. M., & Gerraty, R. T. (2009). A Meta-Analytic Investigation of Neurocognitive Deficits in Bipolar Illness: Profile and Effects of Clinical State. *Neuropsychology, 23*(5), 551-562

McGurk, S. R., Twamley, E. W., Sitzer, D. I., McHugo, G. J., & Mueser, K. T. (2007). A meta-analysis of cognitive remediation in schizophrenia. *American Journal of Psychiatry, 164*(12), 1791-1802. doi: 164/12/1791 [pii]

10.1176/appi.ajp.2007.07060906 [doi]

Perry, W., Potterat, E. G., & Braff, D. L. (2001). Self-monitoring enhances Wisconsin Card Sorting Test performance in patients with schizophrenia: performance is improved by simply asking patients to verbalize their sorting strategy. *J Int Neuropsychol Soc, 7*(3), 344-352

Robinson, L. J., Thompson, J. M., Gallagher, P., Goswami, U., Young, A. H., Ferrier, I. N., & Moore, P. B. (2006). A meta-analysis of cognitive deficits in euthymic patients with bipolar disorder. *Journal of Affective Disorders, 93*(1-3), 105-115

Sanches, M., Bauer, I. E., Galvez, J. F., Zunta-Soares, G. B., & Soares, J. C. (2014). The Management of Cognitive Impairment in Bipolar Disorder: Current Status and Perspectives. *Am J Ther*. doi: 10.1097/mjt.0000000000000120

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., . . . Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry, 59*, 22-33; 34-57

Shi, Q., Pavey, E. S., & Carter, R. E. (2012). Bonferroni-based correction factor for multiple, correlated endpoints. *Pharm Stat, 11*(4), 300-309. doi: 10.1002/pst.1514

Sole, B., Bonnin, C. M., Mayoral, M., Amann, B. L., Torres, I., Gonzalez-Pinto, A., . . . Torrent, C. (2015). Functional remediation for patients with bipolar II disorder: improvement of functioning and

subsyndromal symptoms. *Eur Neuropsychopharmacol, 25*(2), 257-264. doi: 10.1016/j.euroneuro.2014.05.010

Summerfelt, A. T., Alphs, L. D., Funderburk, F. R., Strauss, M. E., & Wagman, A. M. (1991). Impaired Wisconsin Card Sort performance in schizophrenia may reflect motivational deficits. *Archives of General Psychiatry, 48*(3), 282-283

Thompson, J. M., Gallagher, P., Hughes, J. H., Watson, S., Gray, J. M., Ferrier, I. N., & Young, A. H. (2005). Neurocognitive impairment in euthymic patients with bipolar affective disorder. *British Journal of Psychiatry, 186*, 32-40

Torrent, C., Bonnin, C. d. M., Martínez-Arán, A., Valle, J., L. Amann, B., González-Pinto, A., . . . Vieta, E. (2013). Efficacy of functional remediation in bipolar disorder: a multicenter randomized controlled study. *American Journal of Psychiatry, 170*(8), 852-859. doi: doi:10.1176/appi.ajp.2012.12070971

Torres, I. J., Boudreau, V. G., & Yatham, L. N. (2007). Neuropsychological functioning in euthymic bipolar disorder: a meta-analysis. *Acta Psychiatrica Scandinavica, Supplementum*(434), 17-26

Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA: MIT.

Young, R., Biggs, J., Ziegler, V., & Meyer, D. (1978). A rating scale for mania: reliability, validity and sensitivity. *Br J Psychiatry, 133*(5), 429-435

Figure caption:

Figure 1a)-d) Mean performance by the groups on the four indices showing a significant interaction

between group and administration type (error bars represent standard error of the mean)