This is an intriguing article focusing on the potential role of manual motor speed as a marker of BD. This topic is of particular relevance as increasing evidence supports the notion that FTT may be an endophenotype in schizophrenia. Little is however known about FTT in bipolar disorder. I have a few queries regarding the study methodology and interpretation of the results.

-Introduction: please provide some background related to the neural networks associated with FTT. For instance, the involvement of primary motor cortex, cerebellum SMA…and potential reduction in D2 activity in caudate and putamen that is often associated with reduced motor skills in other populations, e.g. older individuals. Also a connection between motor skills and higher order cognitive function should be made, even if briefly.

-Methods:the FTT task included an “only dominant” vs “alternating dominant/non-dominant finger”. Did you explore results by dividing them into “simple tapping” and “complex tapping”? the latter involves additional processing as one has to initiate this “alternating movement” and therefore make a decision. Could the authors comment on this?

-did you monitor mood over the course of the study? Did you correlate these measures with FTT? Could you please address this in the manuscript.

-do you have IQ and simple/complex processing measures for BD and BD rel? Did you consider including them in the manuscript and interpret all these measures together? I think this could be helpful and could at least be mentioned as a future direction.

-Statistical analyses: please explain how you explored variables, e.g. their distribution etc. Also mention that you conducted correlational analyses (they appear in the results section but are not mentioned in methods). What kind of ANOVA did you conduct (MANOVA, MANCOVA, repeated measures, mixed linear modeling)?Which software did you use? Which covariates did you include in your ANOVAs or correlations (if they are partial correlations)? When testing T3 measures did you include baseline measures as covariates? Was medication load (e.g. patients on 1,2 or more medications) taken into account?

-Also I wonder how the authors corrected for “time” in their analyses. Providing correlations of measures at different time points may not be address this sufficiently.

-Results: please provide a graph and/or discuss how participants’ performance changed over the course of the task (I would expect a certain procedural learning effect). Could you please clarify if there was a difference in learning effect across groups? I am asking this because you mentioned that differences in FTT total between BD patients and BD rel disappeared over time. I am curious about your interpretation

-Figure 1. Add caption to explain what the graph is meant to represent. Please consider either transforming this graph into a histogram, boxplot, scatterplot for individual groups. In its current form this graph is unclear and may lead to think that you conducted regression analyses which I think you did not?

-Figure 2. Please provide a caption, and please consider improving the format of your line graph. For instance, there could be dots, triangles and squared dots to distinguish the 3 groups (esp. for readers who will not read the color version of the manuscript. Also is this graph showing means and SD (please clarify)? Please mention all acronyms in your figures. A rule of thumb is that the reader should be able to understand graphs without referring to the text.

-Table 1. Please provide caption and statistics such as F, X2, p values to enable comparisons across the three groups. Same comment applies to table 2.

-Table 3. I can only intuitively understand what this matrix refers to. Please add a very clear caption and please include Ns too.

- Table 3. The standard way to report significance threshold can be .05, .001, or .01 (the reference to p=0 as reported in Table 3 is almost never used). Since a p value cannot tell readers the strength or size of an effect or relationship, I wonder if the authors compared coefficients of correlation between groups? If so could you please report what they found or why they did not make such comparisons?

MINOR POINTS

Please mention in your abstract what kind of statistical analyses you performed