The Version of Scholarly Record of this Article is published in Clinical Psychological Science, available online at: http://dx.doi.org/10.1177/2167702617720716. Note that this article may not exactly replicate the final version published in Clinical Psychological Science.

Toffalini, E., Giofrè, D., & Cornoldi, C. (2017). Pros and Cons of Using Intelligence Batteries for the Study of Clinical Populations: A Response to Beaujean (2017). *Clinical Psychological Science*, *5*, 878–879. doi:10.1177/2167702617720716

Pros and cons of using intelligence batteries for the study of clinical populations:

A response to Beaujean (2017)

Enrico Toffalini[1], David Giofrè[2], & Cesare Cornoldi[1]

1 Department of General Psychology, University of Padova, Padova, Italy.

2 Natural Science and Psychology, Liverpool John Moores University, Liverpool, UK.

Paper accepted 6/20/17

Correspondence concerning this article should be addressed to:

Dr. David Giofrè

Liverpool John Moores University

Natural Sciences and Psychology

Tom Reilly Building Byrom Street, Liverpool, L3 3AF

Tel. +44 151 904 6336

Fax. +44 151 904 6302

E-mail: david.giofre@gmail.com

Word count: 537

Pros and cons of using intelligence batteries for the study of clinical populations: A response to Beaujean (2017)

In his commentary to our study (Toffalini, Giofrè, & Cornoldi, 2017), Beaujean (2017) raised some important issues. In particular, he suggested that the reliability of the WISC-IV indexes may be inadequate for clinical interpretation, that abnormal differences should be discussed rather than statistical significance, and that evidence of the treatment utility of WISC-IV indexes is limited. We agree with Beaujean that using cognitive strengths and weaknesses in isolation for individual clinical decisions is dangerous. However, investigation of groups is somewhat different; comparing group performances under different subtests or conditions, without definitive conclusions, can enhance clinical psychological research.

Concerning technical aspects of Beaujean's commentary, it must be noted that he calculated the "unique" reliability with the omega-s (ω_s), considering a bi-factor model of intelligence. Not all researchers agree with this model, suggesting that the perceived superiority of the bi-factor over the higher-order model could be biased (Murray & Johnson, 2013; but see Gignac, 2016, and Molenaar, 2016 for a different argument). We agree with Beaujean, however, that the bi-factor model may provide relevant information. For this reason, we applied the bi-factor model to our data on 1383 children with SLD (Giofrè, Toffalini, & Cornoldi, 2017). The ω_s was .46 for VCI, .20 for PRI, .37 for WMI, and .57 for PSI, which is in line with recent studies of children with SLD in the USA (Styck & Watkins, 2016). Notably, ω_s measures the effect of reliability of each index after controlling for the g-factor, and criteria (e.g. cut-offs) for evaluating ω_s are not clear. However, some criteria have been proposed based on a simulation study (Gignac & Kretzschmar, 2017). In fact, following these criteria, the unique reliability of indexes in general is relatively large ($\omega_s >$ .30), with the exception of PRI. Demonstrably, the indexes can provide some useful information.

Concerning abnormal differences vs. clinical significance, we agree that statistical significance is often arbitrary. However, abnormal differences can also be arbitrary. For example,

abnormal differences are defined by fixed criteria (e.g., 5% or 1%) in the standardization sample (Silvertein, 1989). Both statistical significance and abnormal differences raise problems, and the "clinical significance" should be considered instead (Kazdin, 1999). We recently showed that differences within the intellectual profile can be clinically relevant in discriminating between typically developing children vs. children with SLD (with adequate sensitivity and specificity; Giofrè et al., 2017). We agree, however, that distinguishing between different subtypes of SLD on the sole basis of the intellectual profile is problematic, because similarities tend to prevail over differences. In our study (Toffalini et al., 2017) we mentioned this fact, but we noted that even small differences can be relevant from a theoretical point of view and should not be neglected.

Regarding the treatment utility, we agree that evidence is insufficient to support the exclusive use of cognitive ability tests, particularly WISC-IV indexes, for designing intervention. Intervention techniques are beyond the goals of our study and have not been sufficiently examined in the literature. Our data merely offer some evidence suggesting that different indexes may be differently linked to success in different aspects of achievement, like reading, calculation and writing.

All in all, we would like to thank Beaujean for his insightful comments, which should be considered in future research.

References

Beaujean, A. A. (2017). *Commentary on Strengths and Weaknesses in the Intellectual Profile of Different Subtypes of Specific Learning Disorder.* Clinical Psychological Science. Advanced online publication. doi:10.1177/0734282917690302

Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence, 41*, 407–422. doi:10.1016/j.intell.2013.06.004

Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence, 55,* 57–68. doi:10.1016/j.intell.2016.01.006

Gignac, G. E., & Kretzschmar, A. (2017). *Evaluating dimensional distinctness with correlated-factor models: Limitations and suggestions*. Intelligence. Advanced online publication. doi:10.1016/j.intell.2017.04.001

Giofrè, D., Toffalini, E., Altoè, G., & Cornoldi, C. (2017). Intelligence measures as diagnostic tools for children with specific learning disabilities. *Intelligence, 61,* 140–145. doi:10.1016/j.intell.2017.01.014

Kazdin, A. E. (1999). The meanings and measurement of clinical significance. *Journal of Consulting and Clinical Psychology, 67*, 332–339. doi:10.1037/0022-006X.67.3.332

Molenaar, D. (2016). On the distortion of model fit in comparing the bifactor model and the higher-order factor model. *Intelligence, 57*, 60–63. doi:10.1016/j.intell.2016.03.007

Silverstein, A. B. (1989). Reliability and abnormality of scale-score ranges. *Journal of Clinical Psychology*, 45, 926–929.

Styck, K. M., & Watkins, M. W. (2016). Structural validity of the WISC-IV for students with learning disabilities. *Journal of Learning Disabilities, 49,* 216–224. doi:10.1177/0022219414539565

Toffalini, E., Giofrè, D., & Cornoldi, C. (2017). Strengths and Weaknesses in the Intellectual Profile of Different Subtypes of Specific Learning Disorder. *Clinical Psychological Science, 5*, 402–409. doi:10.1177/2167702616672038