19.00 €

vol. LXVIII. January-April 2020

N° 287

Four-monthly Journal
ISSN 0006-6761

BPA

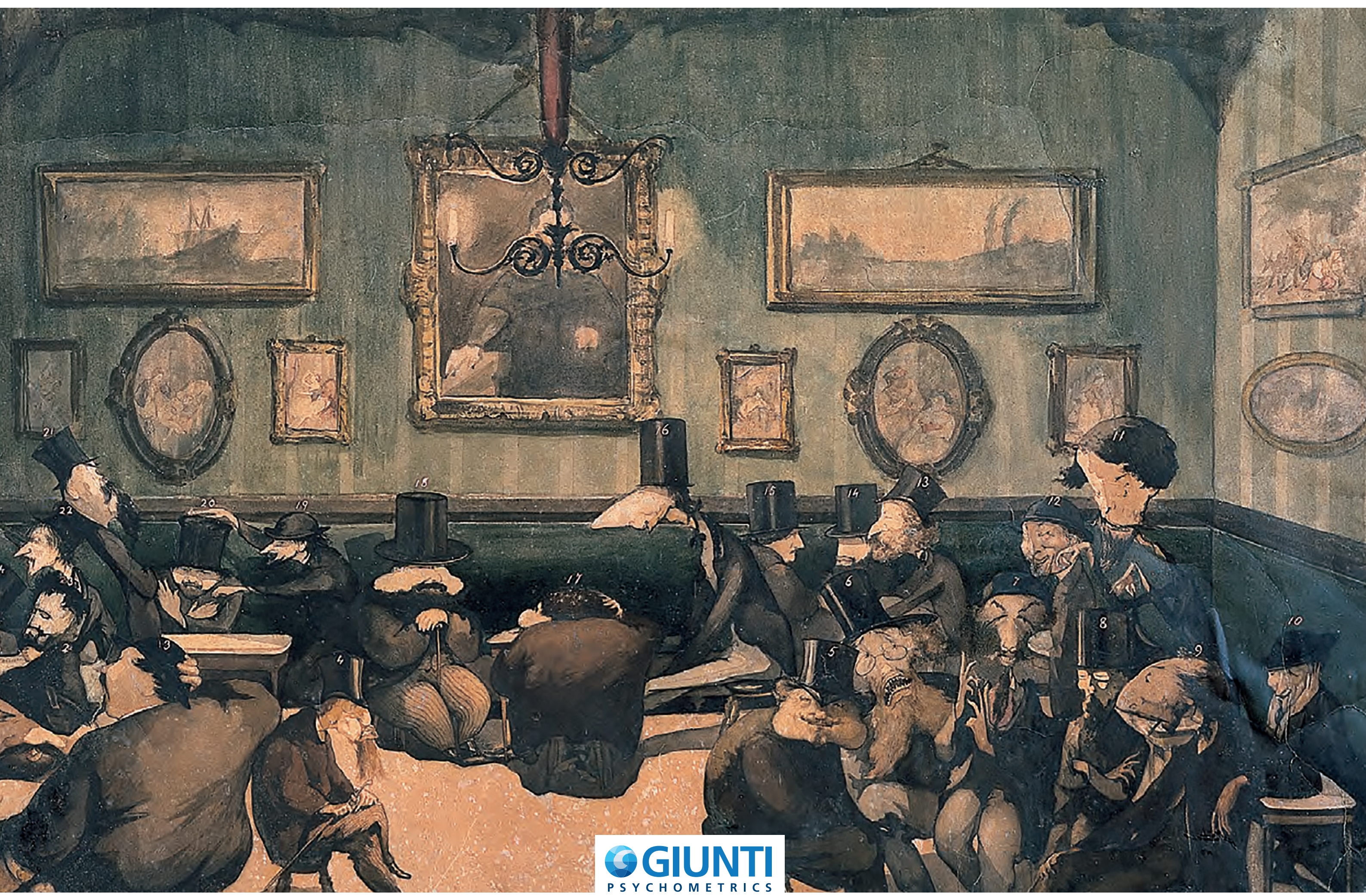
1/3

Bollettino di Psicologia Applicata
APPLIED PSYCHOLOGY BULLETIN

Indexed in PsycINFO® – Scopus Bibliographic Database

Research

Experiences & Tools



GIUNTI PSYCHOMETRICS

Scientific Director Alessandro Zennaro

CONTENTS

Research

Experiences & Tools

DOI: 10.26387/bpa.287.2

Alternative models of estimating the Stop-Signal Reaction Time in the Stop-Signal Paradigm and their differential associations with self-reports of impulsivity domains

Giulia Gialdi, Antonella Somma, Claudia Virginia Manara, Andrea Fossati

School of Psychology, Vita-Salute San Raffaele University, Milan

somma.antonella@hsr.it

ABSTRACT. Lo Stop-Signal Reaction Time (SSRT) come misura di comportamenti impulsivi è stato oggetto di discussione. L'obiettivo del presente studio era di valutare la relazione tra le misure autosomministrate di risk-taking e impulsività con diversi metodi di stima dello SSRT. A questo studio hanno partecipato cinquanta studenti universitari italiani (partecipanti di sesso maschile = 15, 30.0%; partecipanti di sesso femminile = 35, 70.0%; età media = 22.64 anni, *DS* = 2.63 anni). Si è stimato che fossero necessari circa 49 partecipanti per ottenere una potenza di .80 per rilevare un valore *r* di Spearman di .40 con *p*<.05. I partecipanti hanno completato lo SST attraverso un computer portatile in sessioni individuali e hanno completato le versioni italiane di *UPPS-P Impulsivity Scale*, *Barratt Impulsiveness Scale-11*, e *Impulsive-Unsocialized Sensation-Seeking Scale*. I valori *r* di Spearman hanno suggerito che tutti i metodi di stima dello SSRT erano significativamente associati con le misure autosomministrate di sensation-seeking/risk taking. Tuttavia, solo le stime BEEST dello SSRT hanno mostrato associazioni significative anche con le misure delle caratteristiche nucleari dell'impulsività (cioè, mancanza di premeditazione).I nostri risultati sembrano suggerire che utilizzando una prospettiva bayesiana per la stima dello SSRT si possano ottenere misure sperimentali per comportamenti impulsivi e di risk-taking.

SUMMARY. The Stop-Signal Reaction Time (SSRT) as a measure of impulsive behavior has been called into question. The aim of the present study was to assess the relationship between self-report measure of risk-taking and impulsivity and different SSRT estimation methods. Fifty Italian university students (male participants = 15, 30.0%, female participants = 35, 70.0%; mean age = 22.64 years, SD = 2.63 years) agreed to participate in the study. Roughly 49 participants were required to allow .80 power for detecting a Spearman r value of .40 with p<.05. Participants were administered the SST using a laptop computer in individual sessions and completed the Italian versions of the UPPS-P Impulsivity Scale, Barratt Impulsiveness Scale-11, and Impulsive-Unsocialized Sensation-Seeking Scale. Spearman r values suggested that all SSRT models were significantly associated with self-report measures of sensation-seeking/risk taking behaviors. However, only BEEST estimates were non-trivially associated also with measures of core features of impulsivity (i.e., lack of premeditation). Our findings seemed to suggest that adopting a Bayesian perspective on SSRT estimation may allow to obtain experimental measures of both risk-taking and impulsive behaviors.

Keywords: *Stop-Signal Paradigm, Stop-Signal Reaction Time, Impulsivity, Self-reports*

INTRODUCTION

Human beings' success in adapting to an ever-changing environment implies at least partly an ability to control impulses and suppress inappropriate responses. This ability to cancel prepotent responses when they are contextually inappropriate is known as response inhibition (RI; Skippen et al., 2019). RI represents a core component of executive functioning (Miyake et al., 2000), which has been theoretically linked to impulse control (Bari & Robbins, 2013). Notwithstanding its theoretical relevance, empirical studies have yet to provide compelling findings for a relationship between an individual's RI ability and the extent to which they act on impulse (for a review, see Sharma, Markon & Clark, 2014). This failure to find empirical support for the association between RI and impulsivity has been partly attributed to method issues.

Prominent scholars (see for a review, Sharma et al. 2014) have argued that low correlations between self-reports and laboratory tasks result from inconsistent definition of impulsivity across different methods, although all instruments were hypothesized to assess similar underlying mechanisms of behavioral dyscontrol (Sharma et al., 2014). Notably, Sharma and colleagues' meta-analytic findings (2014) confirmed the generally low relations found between self-report and behavioral tasks, but also found that both self-reports and behavioral tasks are useful to predict external criteria. Thus, Sharma and colleagues' meta-analytic findings (2014) showed that the use of multiple assessment strategies based on different methods has validity in assessing impulsive-related constructs (Sharma et al., 2014). According to these results, it seems very important to study the convergence between different methods used to assess RI both from a clinical and a research perspective. Against this background, the present study focused on the relationship between self-reported measures and behavioral tasks of RI.

RI is frequently investigated with Logan's Stop-Signal Paradigm (SSP; Logan & Cowan, 1984). Over the past 35 years, SSP has facilitated the interpretation of numerous developmental, experimental, and neuropsychological studies (e.g., Matzke, Verbruggen & Logan, 2018), and has been applied to examine the nature of inhibition deficits in clinical conditions, such as schizophrenia (Matzke, Hughes, Badcock, Michie & Heathcote, 2017) and attention deficit hyperactivity disorder (e.g., ADHD; Matzke, Curley, Gong & Heathcote, 2019). In the SS paradigm (Logan & Cowan, 1984), participants are asked to perform a two-choice visual response time task, such as responding to the color or the shape of the stimuli. This primary task is occasionally interrupted by a stop-signal that instructs participants not to respond on that trial. Response inhibition can be conceptualized as a race between two independent processes: a go process that is initialized by the primary (choice-task) stimulus and a stop process that is triggered by the stop-signal (Matzke, Love & Heatcote, 2017). The goal is to estimate the latency of the unobservable stop response (Stop-Signal Reaction Time; SSRT).

The independent race model gave rise to several methods to estimate SSRTs (e.g., Verbruggen et al., 2019); the mean SSRT method and the integration method represent the two most widely used approaches to SSRT estimation (Verbruggen et al., 2019), although the mean method was found to be biased in simulation studies (Verbruggen et al., 2019). Although non-parametric estimation methods have been developed for evaluating SSRT, parametric estimation methods are less biased than even the best non-parametric methods and avoid other problems that can be set them, although they may be more computationally intensive (Verbruggen et al., 2019). Matzke, Dolan, Logan, Brown & Wagenmakers (2013) nicely pointed out that the adequate analysis of RT data should not only focus on mean RT, but should take into account the shape of the entire RT distribution; for instance, the shape of SSRT distributions may differ between different clinical populations, without necessary differences in mean SSRT (Matzke et al., 2019). These considerations led Matzke, Dolan and colleagues (2013) to develop a Bayesian parametric approach that enables researchers to estimate the entire distribution of SSRT, under the assumption that SSRTs follow an ex-Gaussian distribution. Bayesian parameter estimation is used to obtain posterior distributions for the model parameters (Matzke et al., 2019). From this point of view, successful response inhibition not only requires relatively fast stop, but the stop process must also be successfully triggered before it can begin the race against the go process (Matzke et al., 2019). Trigger failures pose well-known theoretical and methodological challenges to the interpretation of stop-signal data (Band, Van der Molen & Logan, 2003), mostly because they bias the estimation of entire SSRT distributions resulting in in a dramatic overestimation of SSRTs (Matzke, Love & Heathcote, 2017).

In order to facilitate the application of the Bayesian approach to SSRT estimation, Matzke, Love and colleagues (2013) developed a relatively fast, user-friendly software that allows for the estimation of entire SSRT distributions (BEESTS, Bayesian Ex-Gaussian Estimation of Stop-Signal RT distributions). BEESTS can be applied to individual and hierarchical stop-signal data and comes with an easy-to-use graphical user interface. BEESTS provides users with summary statistics of the posterior distribution of the parameters as well as various diagnostic tools to assess the quality of the parameter estimates (Matzke, Love et al., 2013). Recently, Matzke and colleagues (2019) proposed a parametric framework that extends the standard 2-runner race model to account for go errors, and hence expand the scope of the stop-signal paradigm to the study of response inhibition in the context of difficult choices (Heatcote et al., 2019). This approach is based on Bayesian approach based on the ex-Gaussian distribution – the EXG3 model (Heathcote et al., 2019; Matzke et al., 2019). Interestingly, Matzke and colleagues (2019) showed that the EXG3 approach can be successfully applied to stop-signal tasks with high error rates; however, this model requires novel stop-signal data with high error rates and a manipulation of task difficulty to enable researchers to study difficult-choice inhibition (Heathcote et al., 2019).

Even keeping these issues in mind, extant research indicates that response inhibition may have important implications for both typical and atypical developmental trajectories. For instance, developmental studies documented that the SSRT manifests an inverted U-shape across the lifespan, accelerating during childhood and slowing down again in old age (e.g., van de Laar, van den Wildenberg, van Boxtel & van der Molen, 2011). Moreover, reduced SSRT during adolescence has been proposed as a major factor contributing to greater impulse control in adulthood (Shulman et al., 2016). Finally, studies on clinical populations showed that response inhibition may have relevant implications for the treatment outcome of people with several mental disorder and problem behaviors (ADHD, obsesssive-compulsive disorder, pathological gambling, substance use disorders, etc.; e.g., Nederkoorn, Jansen, Mulkens & Jansen, 2007).

Based on these findings, a link between response inhibition and impulse control was explicitly hypothesized in personality literature, where SSRT was often used as an experimental measure of impulsivity (Skippen et al., 2019). However, prominent authors (e.g., Stahl et al., 2014) have called into question the direct correspondence between the construct of response inhibition and constructs such as delay aversion (i.e., a preference for smaller immediate rather than larger later rewards), impulsivity (i.e., acting without thought of consequence or adequate information), and sensation-seeking/risk-taking (e.g., Dalley & Robbins, 2017). Extant literature indicates a clear distinction between self-report and experimental measures of impulsivity, suggesting that measures from both domains should be used to obtain an accurate description of impulsivity (e.g., Sharma et al., 2014; Stahl et al., 2014). Indeed, self-report measures operationalize impulsivity as a stable trait, asking questions about propensity towards urgency, sensation seeking, lack of premeditation, and lack of perseverance (Whiteside & Lynam, 2001). On the other hand, behavioral impulsivity measures are characterized by substantial variability (Sharma et al., 2014; Stahl et al., 2014).

Notwithstanding the adequate reliability of the SSRT (Wöstmann et al., 2013) and of several self-report measures of impulsivity (Sharma et al., 2014), meta-analytic studies suggest that the relationship between SSRT and self-report measures of impulsivity measures is weak ($r \approx .1$). Interestingly, the associations between SSRT and self-reported impulsivity have been shown to be unaffected by methodological differences across versions of the Stop-Signal Task (Skippen et al., 2019). Recently, it has been proposed that traditional way in which SSRT is measured may not provide a pure measure of response inhibition latency (e.g., Skippen et al., 2019), suggesting that improved estimation of the SSRT may lead to improved ability to identify relationships between measures of response inhibition and impulsivity self-reports.

Despite the relevance of these considerations, there is still a dearth of studies trying to provide data on how different methods to estimate SSRT provide are differently related to self-report measures of impulsive behaviors. To the best of our knowledge, Matzke and colleagues (Matzke, Hughes et al., 2017) applied the BEESTS that accounts for trigger failure to stop-signal data from a clinical sample of schizophrenia patients and matched controls. However, no direct comparison between different SSRT estimation methods was carried out in this seminal study, which indicated that attentional factors need to be taken into account when interpreting results from the stop-signal paradigm (Matzke, Hughes et al., 2017). Moreover, Skippen and colleagues (2019) evaluated if the integration method (Verbruggen et al., 2019) and the EXG3 method (Matzke et al., 2019) of SSRT estimation were characterized by different relationships with self-reports of impulsivity and sensation-seeking in a sample of 174 healthy adolescents and young adults. Skippen and colleagues' (2019) findings suggested that the integration method estimate of SSRT was significantly and modestly correlated with self-report impulsivity measures and moderately correlated with other experimental measures of impulsivity; rather, the mean SSRT derived using the EXG3 model was not reliably correlated with any impulsivity or outcome measure. However, Skippen and colleagues (2019) relied on a 700-trial stop-signal paradigm with a number parity go task which is optimal for incorporating both trigger failure and go failure (i.e., EXG3; Matzke et al., 2019). This approach can be successfully applied to relatively difficult go task with high error rates, extending the applicability of the stop-signal procedure to research areas in experimental psychology, such as recognition memory, that often rely on difficult choice tasks and manipulations that affect error rates (e.g., Kim, Potter, Craigmile, Peruggia & van Zandt, 2017). Nevertheless, no formal comparison among different nonparametric and parametric methods was carried out; moreover, these advanced stop-signal paradigms are not generally administered to assess inhibition in applied contexts.

Starting from these considerations, we tried to provide preliminary evidence on how different methods for estimating the SSRT could yield different relationships with self-reports of impulsivity dimensions in Italian community-dwelling adults. We relied on an open-source stop-signal paradigm (i.e., the Stop-Signal Task; Verbruggen et al., 2019) in order to improve the replicability of our findings (Miłkowski, Hensel & Hohol, 2018). As some scholars argue, this approach should be used whenever possible to generate publishable results (Easterbrook 2014; Gleeson, Davison, Silver & Ascoli, 2017). Indeed, the inability to reproduce the findings of many published studies has been recently highlighted (Baker, 2016; Open Science Collaboration, 2015), and there is general agreement that this is a problem that needs to be tackled. In particular, the following methods for estimating the SSRT were considered: a) the mean method; b) the integration method; c) the Bayesian estimation of ex-Gaussian SSRT (BEESTS method); and d) BEESTS with trigger failure. Moreover, a comprehensive set of measures to assess impulsivity dimensions, which included the Italian translations of the *Barratt Impulsiveness Scale-11* (Patton, Stanford & Barratt, 1995), *UPPS-P Impulsivity Scales* (Cyders & Smith, 2007; Whiteside & Lynam, 2001), and the *Impulsive Sensation-Seeking Scale* of *Zuckerman-Kuhlman Personality Questionnaire* (Zuckerman, Kuhlman, Thornquist & Kiers, 1991) was used. To be included in the set, measures should have been provided with sound psychometric properties in Italian samples. In order to control for the effect of participants' educational level on responses to self-report measures, in the present study only on adult university students were recruited.

Based on previous findings (Skippen et al., 2019) based on the EXG3 method, we hypothesized that the traditional (i.e., mean method) estimate of SSRT was weakly associated with self-report measures of impulsivity, whereas BEEST estimates were expected to be more consistently associated with different measures assessing different aspects of impulsive behavior.

METHODS

Participants

Fifty-three adult university students originally agreed to participate in the present study. However, based on stop-signal quality control (e.g., Skippen et al., 2019), three participants were not included in the final sample. In particular, three participants' mean go RT were faster than their mean RT on failed stop trials, violating the context independence assumption of the horse-race model. The reduced number of participants with poor quality of Stop-Signal Task prevented us from conducting formal missing values analyses. The final sample was composed of 50 participants; 35 (70%) participants were female, and 15 (30%) participants were male. Participants' mean age was mean age = 22.64 years, *SD* = 2.63 years. On average, participants received 16.84 years of education, *SD* = 2.58 years. The majority of the participants were unmarried, *n* = 46, 92%. In order to participate in the study, participants had to sign a written informed consent form. In the present study, we adhered to the Italian Association of Psychology (2015) ethical code of conduct for psychological research on human participants.

Measures

- *Stop-Signal Task* (Verbruggen et al., 2019). In the present study, an open-source software was used for administering a simple two-choice task that complies with the recommendations described in Verbruggen and colleagues (2019). The primary go task is two-choice task in which participants have to discriminate between an arrow pointing to the left and an arrow pointing to the right. On go trials (75% of the trials), participants have to respond as fast and accurate as possible to these arrows. On stop-signal trials (25% of the trials), the arrows are replaced by XX (i.e. a visual stop-signal) after a variable delay, instructing participants to cancel their response. The default go stimuli are two green arrows; the fixation sign and arrows are presented in the center of the screen on a white background (Verbruggen et al., 2019). As recommended by Band and colleagues' (2003), an adaptive staircase was used to adjust SSD on a trial-by-trial basis to optimize the estimation of SSRT, targeting a 50% failure rate on stop trials. The SSD increased or decreased by 50 ms after every successful or failed stop trial, respectively (Verbruggen et al., 2019). This version of STOP-IT is platform-independent and was used offline (Verbruggen et al., 2019).

– *UPPS-P Impulsive Behavior Scale* (Cyders & Smith, 2007). The UPPS-P is 59-item, Likert-type, self-report questionnaire, which was designed to measure five dimensions of impulsive behavior, namely, *Negative urgency* (12 items), (lack of) *Premeditation* (11 items), (lack of) *Perseverance* (10 items), *Sensation seeking* (12 items), and *Positive urgency* (14 items). The five scales were designed to assess the tendency to commit rash actions as a result of intense negative affect (*Negative urgency*), the tendency to think and reflect on the consequences of an act before engaging in that act (*Premeditation*), the ability to remain with a task until completion and avoid boredom (*Perseverance*), the tendency to seek excitement and adventure (*Sensation seeking*), and tendency to act rashly in response to a positive mood (*Positive urgency*). Items are assessed from 1 (*agree strongly*) to 4 (*disagree strongly*). The *UPPS-P Impulsive Behavior Scales* showed adequate psychometric properties (Cyders & Smith, 2007; Whiteside & Lynam 2001) also in their Italian translation (Fossati, Di Ceglie, Acquarini & Barratt, 2016). For ease of presentation, in the present study the *Premeditation* and *Perseverance* scales were reverse scored to reflect lack of premeditation and lack of perseverance, respectively.
– *Barratt Impulsiveness Scale-11* (*BIS-11*; Patton et al., 1995). The BIS-11 is a 30 item Likert-type self-report questionnaire that measures three facets of impulsivity: motor impulsivity, attention impulsivity, and non-planning impulsivity. The three facets scores are summed to produce a total impulsivity score. The psychometric properties of the Italian translation of the BIS-11 were previously assessed (Fossati et al., 2001).
– *Zuckerman-Kuhlman Personality Questionnaire Impulsive Unsocialized Sensation Seeking Scale* (*ImpSS*; Zuckerman et al., 1991). The ImpSS is a 19 items self-report measure assessing lack of planning and the tendency to act impulsively without thinking and the seeking of excitement, novel experiences, and the willingness to take risks for these types of experiences. The ImpSS items are general in content and do not describe specific activities such as drinking or sex. The reliability and validity of the Italian translation of the ImpSS have been previously assessed (e.g., Carlotta, Borroni, Maffei & Fossati, 2011; De Pascalis & Russo, 2003).

Procedures

Participants were administered the Stop-Signal Task using a laptop computer in individual session with the assistance of trained psychologists who were kept blind to the aims of the present study. After completing the Stop-Signal Task, participants were asked to complete the self-report questionnaires; self-report measures were administered in random order and scored blind to Stop-Signal Task results. Before gathering data, we carried out power analyses, considering that we were interested in detecting at least moderate effect size (i.e., Spearman $r \geq |.30|$; Cohen, 1988). Power analysis results indicated that roughly 49 participants were required to allow .80 power for detecting a Spearman r value of .40 with $p<.05$. However, it should be observed that the minimum Spearman r value for $p<.05$ significance level for 50 subjects was $|.28|$.

Data analysis

In the present study, both parametric and nonparametric methods to estimate SSRT were used. Although the mean method is known to be strongly influenced by the skewness of the go RT distribution and by go omissions errors, it is still the most popular nonparametric estimation method when the tracking procedure is used due to its easiness (Verbruggen et al., 2019). According to the mean method, SSRT can be estimated easily by subtracting mean SSD from mean RT on go trials (Verbruggen et al., 2019) As an alternative nonparametric estimation method, in the present study, we relied on the version of the integration method which has been shown to produce the most reliable and least biased non-parametric SSRT estimates in Verbruggen and colleagues' (2019) simulation study (i.e., the integration method with replacement of go omissions). According to this method, SSRT can then be estimated by subtracting mean SSD from the nth RT. To determine the nth RT, all go trials with a response are included (including go trials with a choice error and go trials with a premature response). Importantly, go omissions (i.e. go trials on which the participant did not respond before the response deadline) are assigned the maximum RT in order to compensate for the lacking response. Premature responses on unsuccessful stop trials (i.e. responses executed before the stop signal is presented) should also be included when calculating *p* (respond|signal) and mean SSD (Verbruggen et al., 2019).

Different from non-parametric methods, parametric methods allow for the estimation of the entire distribution of SSRTs (Matzke, Dolan et al., 2013). In particular, in the presents study we relied on two different BEESTS models, namely, the "traditional" BEESTS method (Matzke, Dolan et al., 2013), and the BEESTS method with trigger failure (Matzke, Love & Heathcote, 2017). The BEEST methods relied on a Bayesian parametric approach that allows for the estimation of the entire distribution of SSRTs. SSRTs are assumed to follow an ex-Gaussian distribution and Markov chain Monte Carlo sampling are used to estimate the parameters of the SSRT distribution (e.g., Matzke, Dolan et al., 2013). The BEESTS method with trigger failure enables researchers to simultaneously estimate the probability of trigger failures (i.e., deficiencies in triggering the stop process) and the entire distribution of stopping latencies (Matzke, Love & Heathcote, 2017); the resulting SSRT estimates are corrected for the bias that results from deficiencies in triggering the stop process (Matzke, Love & Heathcote, 2017). In the present study, we relied on an hierarchical estimation (e.g., Matzke, Dolan et al., 2013; Matzke, Love & Heathcote, 2017), so that the estimation of each individual's model parameters is informed by data from the entire sample, resulting in more precise and, on average, more accurate estimates of the true parameters (e.g., Farrell & Ludwig, 2008).

In the present study, we relied on the software developed by Verbruggen and colleagues (2019) in order to compute the SSRT based on the integration method; rather, SSRT estimates based on parametric methods were based on the BEESTS software developed by Matzke and colleagues (Matzke, Love et al., 2013; Matzke, Love & Heathcote, 2017).

Cronbach's alpha coefficient was used to estimate the internal consistency reliability of the self-report measures of impulsivity. The limited size of the sample strongly suggested to rely on nonparametric statistics for hypothesis testing. Spearman *r* coefficients with Bootstrap bias-corrected and accelerated (BCa; Efron & Tibshirani, 1998) 95% confidence intervals were computed to evaluate the strength and significance of the associations between SSRT estimates and impulsivity self-report questionnaire scores. The basic idea of bootstrapping is that inference about a population from sample data can be modelled by resampling the sample data and performing inference about a sample from resampled data; thus, bootstrapping can be used for a number of different aims, including hypothesis testing and confidence interval estimation (Efron & Tibshirani, 1998). Although bootstrapping may not provide general finite-sample guarantees, it represents a straightforward way to derive estimates of standard errors and confidence intervals for complex estimators of complex parameters of the distribution, including Spearman r coefficient (Efron & Tibshirani, 1998). To reduce the effects of random sampling error on Bootstrap estimates, in the present study 10,000 Bootstrap replications were used to generate each 95% confidence interval (Davison & Hinkley,1997).

The presence of significant differences between male and female participants on the self-report impulsivity measures was assessed using the Mann-Whitney *U* statistic and the Vargha and Delaney's (2000) *A* effect size. Vargha and Delaney's (2000) *A* measure returns a value between 0 and 1, representing the probability that a randomly selected observation from a sample (e.g., male subsample) is bigger than a randomly selected observation from another sample (e.g., female subsample). Vargha and Delaney's (2000) *A* values of .5 indicate that the medians are the same, whereas values of 1 and 0 mean that there is no overlap. In this respect, A index is analogous to the area under the receiver operating characteristic curve (Vargha & Delaney, 2000). Vargha and Delaney (2000) provide suggested thresholds for interpreting the effect size, .5 means no difference at all; up to .56 indicates a small difference; up to .64 indicates medium; values over .71 are considered large. The same intervals apply below .5.

Finally, the repeated-measure Friedman nonparametric

ANOVA was used to evaluate if the five methods for evaluating SSRT yielded homogeneous SSRT estimates; in case of significance (i.e., $p<.05$) of the omnibus test, Wilcoxon-Bonferroni post-hoc contrasts were computed to carry out pairwise median comparisons, while protecting for the familywise error rate.

RESULTS

In this sample, all impulsivity measures were significantly and non-negligibly inter-correlated, with Spearman *r* values ranging from .32 (BIS-11 and UPPS-P *Sensation seeking* total scores) to .81 (UPPS-P *Negative urgency* and *Positive urgency* total scores), all $p<.05$, with a median *r* value of .55, $SD = .15$. Also, the SSRT estimation methods yielded SSRT estimates that were substantially inter-correlated, median Spearman *r* value = .70, $SD = .11$, min. Spearman *r* value = .63 (mean method and BEESTS with no trigger failure), max. Spearman *r* value = .89 (BEESTS with no trigger failure and BEESTS with trigger failure), all $p<.001$. With the exception of the UPPS-P *Sensation seeking* scale scores (male participants: $M = 32.80$, $SD = 8.23$; female participants: $M = 25.79$, $SD = 8.36$; Mann-Whitney $U = 127.00$, $z = 2.78$, $p<.01$, common language effect size =.73), none of the remaining impulsivity scale scores significantly differentiated male participants from female participants.

The descriptive statistics of the SSRT estimates and self-report measures of impulsive behaviors, and the Spearman *r* values for the associations between the SSRT estimates and the BIS-11, UPPS-P, and ImpSS scale scores are summarized in Table 1. The Friedman ANOVA omnibus test was highly significant, $\chi^2(3) = 34.52$, $p < .001$, $W = .23$. Median SSRT estimates with different superscripts indicate significant Wilcoxon-Bonferroni post-hoc comparisons; in Wilcoxon-Bonferroni contrasts, the nominal significance level (i.e., $p<.05$) was set at $p<.0083$.

DISCUSSION

To the best of our knowledge, the present study represents the first attempt at evaluating how different methods for estimating the SSRT could yield different relationships with highly reliable self-reports of impulsivity dimensions in a sample of community-dwelling adults. In order to improve the replicability of our findings (Easterbrook, 2014; Miłkowski et al., 2018), an open access stop-signal paradigm (Verbruggen et al., 2019) was administered in the present study. When the relationships between different SSRT estimates and self-reports of impulsive behavior dimensions were evaluated, specific patterns of significant associations emerged, at least in a sample of Italian university students. As a whole, these significant associations were at least of moderate size, according to conventional standards (Cohen, 1988), and were markedly larger than the typical average effect size (i.e., *r* coefficient) estimate reported in meta-analytic studies (e.g., Sharma et al., 2014). This finding seemed to support previous considerations (e.g., Verbruggen et al., 2019) suggesting that problems with SSRT estimation may be responsible for the poor correspondence between response inhibition tasks and self-report measures of impulsive behaviors. As a whole, these findings are consistent with the recent emphasis on developing enhanced methods for SSRT assessment as a promising approach to filling the gap between experimental models and self-report measures of impulsivity (Matzke et al., 2018; Verbruggen et al., 2019).

To overcome the possible bias of correlation estimates due to the measurement error of the self-report questionnaires of impulsivity, in the present study, we relied only on measures that were provided with adequate reliability values in their Italian translations; not surprisingly, in this study all Cronbach's a were higher than .80 (median Cronbach's $\alpha = .88$, $SD = .04$, min.–max. range: .82-.95). Although the reliability estimates of the impulsivity self-reports were quite similar in their values, the four SSRT estimates yielded different patterns of associations with the self-report questionnaires of impulsive behaviors. For instance, the mean method of SSRT estimation yielded a non-trivial association only with a single measure of sensation seeking, namely, the UPPS-P *Sensation seeking* scale. Rather, the integration method, which represents the most accurate nonparametric estimate of SSRT, and both BEESTS methods showed non-negligible relationships with *Sensation seeking* as it was operationalized in both the UPPS-P and ImpSS questionnaires. It should be observed that the BEESTS method with trigger failure estimation for computing the SSRT yielded the largest and most homogeneous correlations with both *Sensation seeking* scales.

Moreover, in our study only the BEESTS SSRT estimates were non-negligibly correlated with the UPPS-P (lack of) *Premeditation* scale scores. Confirming and extending recent findings (Afonso Jr., Machado, Carreiro & Machado-

Table 1 – Correlations (Spearman *r* values) between total scores of self-report measures of impulsivity and the Stop Signal Reaction Time estimates in the full sample (N = 50)

	Stop-signal reaction time estimation methods						
	Mean method	Integration	BEESTS	BEEST-WTF			
Self-report scale total scores	*r* (95% CI)	*r* (95% CI)	*r* (95% CI)	*r* (95% CI)	*M*	*SD*	α
Barratt Impulsivity Scale-11	.02 (–.27, .29)	–.08 (–.34, .20)	–.04 (–.39, .35)	–.14 (–.47, .23)	58.45	10.55	.85
UPPS-P Negative urgency	.09 (–.21, .38)	.02 (–.27, .31)	–.18 (–.55, .19)	–13 (–.47, .24)	26.45	7.19	.88
UPPS-P Premeditation	–.04 (–.29, .22)	–.13 (–.38, .14)	**–.35 (–.60, –.03)**	**–.39 (–.64, –.07)**	20.59	5.74	.90
UPPS-P Perseverance	–.01 (.29, .25)	–.09 (–.36, .19)	–.14 (–.49, .22)	–.24 (–.58, .12)	18.54	4.64	.82
UPPS-P Sensation seeking	**–.44 (–.63, –.18)**	**–.40 (–.59, –.16)**	**–.39 (–.61, –.10)**	**–.45 (–.65, –.16)**	27.94	8.86	.90
UPPS-P Positive urgency	–.18 (–.46, .13)	–.22 (–.50, .10)	–.24 (–.54, .11)	–.25 (–.52, .07)	25.64	9.73	.95
Impulsive Sensation seeking	–.24 (–.50, .05)	**–.30 (–.55, –.01)**	**–.34 (–.60, –.02)**	**–.43 (–.68, –.09)**	6.65	4.49	.85
Mdn (ms.)	252.71[b]	239.00[a]	243.96[b]	234.65[a]			
M (ms.)	254.68	240.18	243.44	233.74			
SD (ms.)	36.14	59.47	16.75	9.32			

Legenda. BEESTS = Bayesian Ex-Gaussian Estimation of Stop-Signal Reaction Time distributions; WTF = with trigger failure.

Note. 95% CI: bias corrected and accelerated Bootstrap 95% confidence interval. In the present study, 10,000 Bootstrap replicates were used to estimate each confidence interval. Bold highlights significant Spearman *r* values. Median scores with different superscripts were significantly different in Wilcoxon-Bonferroni post-hoc comparisons; in Wilcoxon-Bonferroni contrasts, the nominal significance level (i.e., $p<.05$) was set at $p<.0083$.

Pinheiro, 2020), our results seemed to suggest that the Stop-Signal Task may represent an experimental paradigm to assess a core component of impulsivity which is central to all major theories of impulsive behaviors, namely the subject's propensity towards acting without thinking (i.e., lack of *Premeditation*; Whiteside & Lynam, 2001). Marginally, our Friedman ANOVA results showed that the BEESTS method with trigger failure estimation yielded the lowest average SSRT value, although it was not statistically different from the average SSRT value that was provided by the integration method in the Wilcoxon-Bonferroni post hoc contrasts. This finding was at least partially consistent with Matzke, Love and Heathcote (2017) hypotheses.

Interestingly, in the face of these non-trivial, significant relationships between selected SSRT estimates, and *Sensation seeking* and *Premeditation* impulsivity traits, significant associations between any SSRT estimate, and UPPS-P *Negative urgency*, UPPS-P *Positive urgency*, UPPS-P *Perseverance*, and BIS-11 total scores wasn't observed. This finding was consistent with recent studies based on the UPPS-P (Afonso et al., 2020), as well as with meta-analytic evidence largely based on the BIS-11 (Sharma et al., 2014). Although Skippen and colleagues (2019) reported a small association between the BIS-11 total score and the SSRT estimates obtained through the application of the integration method, it should be observed that Skippen and colleagues (2019) relied on a longer (i.e., 700 trials) relatively difficult go task though to be more akin to a decision make task in order to apply a Bayesian EXG3 model to estimate SSRT (e.g., Matzke et al., 2019), and that this association was not found when the EXG3 model was applied to the same data.

Thus, our data seemed to suggest that the Stop-Signal Task is likely to represent an experimental approach to evaluating participant's propensity towards excitement and adventure (i.e., sensation seeking) and (lack of) premeditation; these relationships with reliable self-report measures of *Sensation seeking* and *Premeditation* are captured with increasing accuracy moving from SSRT estimates based on the mean method to SSRT estimates based on Bayesian models with trigger failure estimation. Thus, our findings suggest that *Sensation seeking* and (lack of) *Premeditation* may represent target constructs for Stop-Signal Task studies, particularly when they are assessed using the corresponding UPPS-P and ImpSS scales. In particular, it should be observed that the ImpSS construct includes both the subject's tendency to act impulsively without thinking and his/her willingness to take risks for the sake of excitement (Zuckerman et al., 1991).

Ignoring the specificity of these relationships and the importance of accurate assessment of different self-report traits within the realm of impulsive behaviors is likely to result in severe under-estimation of the relationship between self-reports of impulsivity and SSRT estimates. For instance, in our study the median Spearman *r* value that was computed across all SSRT estimate and all self-report impulsivity scales was as small as −.19; this value was not so different from the average *r* value reported in meta-analytic studies (Sharma et al., 2014).

Of course, the results of the present study should be considered in the light of several limitations. Our sample was limited in size and included only adult university students; this makes it more a convenient study group than a sample actually representative of the Italian university student population, and inherently limits the generalizability of our findings to samples from other populations (e.g., clinical samples, forensic samples, etc.). We relied on a frequentist approach to data analyses, although the BEESTS estimates are based on Bayesian assumptions. However, it should be observed that nonparametric methods of SSRT assessment were developed outside the Bayesian framework (Matzke et al., 2018); moreover, the development of the BEESTS approaches within the Bayesian framework does not prevent from using different data analysis approaches (e.g., Matzke, Love & Heathcote, 2017). We relied on a set of measures of impulsive behaviors that were shown to be provided with adequate psychometric properties also in their Italian translations; however, using different measures of impulsivity or directly assessing behaviors that are known to be related to poor impulse control (e.g., substance abuse, pathological gambling, etc.) as outcome variables in SSRT studies may yield different results.

Although different stop-signal paradigms are available, in our study we relied on an open source software that can be used to execute a Stop-Signal Task that complies with the recommendations described in the stop-signal consensus guide (Verbruggen et al., 2019). Moreover, in the present study, despite we computed SSRT based on different estimation techniques, we relied on a single integration method. This method choice is due to the fact that in their simulation study, Verbruggen and colleagues (2019) nicely showed that the integration method with replacement of go omissions was the least biased and most reliable parametric method for estimating SSRT. It could be argued that in the same study Verbruggen and colleagues (2019) discouraged the use of the mean method. Nevertheless, SSRT estimation method was included because, although biased, it is still popular (e.g., Verbruggen et al., 2019). Finally, it should be observed that in our study we did not compute SSRT based on the EXG3 model because it was meant to extends the scope and applicability of the stop-signal paradigm to the study of response inhibition in the context of difficult choices (Matzke et al., 2019), which is not consistent with common stop-signal paradigm (Verbruggen et al., 2019).

Even keeping these limitations in mind, these findings may prove useful in providing support to the use of advanced (i.e., Bayesian) SSRT estimation methods in order to evaluate the associations between SSRT and self-reports of impulsivity. Indeed, BEESTS estimation methods may be helpful in overcoming methodological problems resulting in lack of relations between self-report scales commonly used to measure impulsivity traits and laboratory impulsive-behavior tasks (e.g., Sharma et al., 2014).

References

AFONSO, A.S. Jr., MACHADO, A.V., CARREIRO, L.R.R. & MACHADO-PINHEIRO, W. (2020). Interaction between inhibitory mechanisms involved in stroop-matching and stop-signal tasks, and their association with impulsivity levels. *Psychology* & *Neuroscience*. Advance online publication. doi: 10.1037/pne0000203

BAKER, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature News*, *533* (7604), 452.

BAND, G., VAN DER MOLEN, M. & LOGAN, G. (2003). Horse-race model simulations of the stop-signal procedure. *Acta Psychologica*, *112* (2), 105-142.

BARI, A. & ROBBINS, T.W. (2013). Inhibition and impulsivity: Behavioral and neural basis of response control. *Progress in Neurobiology*, *108*, 44-79.

CARLOTTA, D., BORRONI, S., MAFFEI, C. & FOSSATI, A. (2011). The role of impulsivity, sensation seeking and aggression in the relationship between childhood AD/HD symptom and antisocial behavior in adolescence. *Neurology, Psychiatry and Brain Research, 17* (4), 89-98.

COHEN, J. (1988). *Statistical power analysis* (*2nd ed.*). Hillsdale NJ: Erlbaum.

CYDERS, M.A. & SMITH, G.T. (2007). Mood-based rash action and its components: Positive and negative urgency. *Personality and Individual Differences*, *43*, 839-850.

DALLEY, J.W. & ROBBINS, T.W. (2017). Fractionating impulsivity: Neuropsychiatric implications. *Nature Reviews Neuroscience*, *18* (3), 158-171.

DAVISON, A.C. & HINKLEY, D.V. (1997). *Bootstrap methods and their application*. Cambridge, UK: Cambridge University Press.

DE PASCALIS, V. & RUSSO, P.M. (2003). Zuckerman-Kuhlman Personality Questionnaire: Preliminary results of the Italian version. *Psychological Reports*, *92*, 965-974.

EASTERBROOK, S.M. (2014). Open code for open science? *Nature Geoscience*, *7*, 779-781.

EFRON, B. & TIBSHIRANI, R.J. (1998). *An introduction to the bootstrap*. New York: Chapman & Hall/CRC.

FARRELL, S. & LUDWIG, C.J.H. (2008). Bayesian and maximum likelihood estimation of hierarchical response time models. *Psychonomic Bulletin* & *Review*, *15* (6), 1209-1217.

FOSSATI, A., DI CEGLIE, A., ACQUARINI, E. & BARRATT, E.S. (2001). Psychometric properties of an Italian version of the Barratt Impulsiveness Scale-11 (BIS-11) in nonclinical subjects. *Journal of Clinical Psychology*, *57* (6), 815-828.

GLEESON, P., DAVISON, A.P., SILVER, R.A. & ASCOLI, G.A.

(2017). A commitment to open source in neuroscience. *Neuron*, *96* (5), 964-965.

HEATHCOTE, A., LIN, Y.S., REYNOLDS, A., STRICKLAND, L., GRETTON, M. & MATZKE, D. (2019). Dynamic models of choice. *Behavior Research Methods*, *51* (2), 961-985.

ITALIAN ASSOCIATION OF PSYCHOLOGY (2015). *Ethical Code*. Retrieved from https://aipass.org/node/11560

KIM, S., POTTER, K., CRAIGMILE, P.F., PERUGGIA, M. & VAN ZANDT, T. (2017). A Bayesian race model for recognition memory. *Journal of the American Statistical Association*, *112* (517), 77-91.

LOGAN, G.D. & COWAN, W.B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, *91* (3), 295-327.

MATZKE, D., CURLEY, S., GONG, C.Q. & HEATHCOTE, A. (2019). Inhibiting responses to difficult choices. *Journal of Experimental Psychology: General*, *148* (1), 124-142.

MATZKE, D., DOLAN, C.V., LOGAN, G.D., BROWN, S.D. & WAGENMAKERS, E.J. (2013). Bayesian parametric estimation of stop-signal reaction time distributions. *Journal of Experimental Psychology: General*, *142*, 1047-1073.

MATZKE, D., HUGHES, M., BADCOCK, J.C., MICHIE, P.T. & HEATHCOTE, A. (2017). Failures of cognitive control or attention? The case of stop-signal deficits in schizophrenia. *Attention, Perception, & Psychophysics*, *79* (4), 1078-1086.

MATZKE, D., LOVE, J. & HEATHCOTE, A. (2017). A Bayesian approach for estimating the probability of trigger failures in the stop-signal paradigm. *Behavior Research Methods*, *49*, 267-281.

MATZKE, D., LOVE, J., WIECKI, T.V., BROWN, S.D., LOGAN, G.D. & WAGENMAKERS, E.J. (2013). Release the BEESTS: Bayesian estimation of ex-Gaussian stop-signal reaction time distributions. *Frontiers in Psychology*, *4*, 918.

MATZKE, D., VERBRUGGEN, F. & LOGAN, G. (2018). The stop-signal paradigm. In E.-J. Wagenmakers & J.T. Wixted (Eds.), *Methodology: Volume* 5. John Wiley & Sons, Inc.

MIŁKOWSKI, M., HENSEL, W.M. & HOHOL, M. (2018). Replicability or reproducibility? On the replication crisis in computational neuroscience and sharing only relevant detail. *Journal of Computational Neuroscience*, *45* (3), 163-172.

MIYAKE, A., FRIEDMAN, N.P., EMERSON, M.J., WITZKI, A.H., HOWERTER, A. & WAGER, T.D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41* (1), 49-100.

NEDERKOORN, C., JANSEN, E., MULKENS, S. & JANSEN, A. (2007). Impulsivity predicts treatment outcome in obese children. *Behaviour Research and Therapy*, *45* (5), 1071-1075.

OPEN SCIENCE COLLABORATION (2015). Estimating the reproducibility of psychological science. *Science*, *349* (6251).

PATTON, J.H., STANFORD, M.S. & BARRATT, E.S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology*, *51* (6), 768-774.

SHARMA, L., MARKON, K.E. & CLARK, L.A. (2014). Toward a theory of distinct types of “impulsive” behaviors: A meta-analysis of self-report and behavioral measures. *Psychological Bulletin*, *140* (2), 374-408.

SHULMAN, E.P., SMITH, A.R., SILVA, K., ICENOGLE, G., DUELL, N., CHEIN, J. & STEINBERG, L. (2016). The dual systems model: Review, reappraisal, and reaffirmation. *Developmental Cognitive Neuroscience*, *17*, 103-117.

SKIPPEN, P., MATZKE, D., HEATHCOTE, A., FULHAM, W. R., MICHIE, P. & KARAYANIDIS, F. (2019). Reliability of triggering inhibitory process is a better predictor of impulsivity than SSRT. *Acta Psychologica*, *192*, 104-117.

STAHL, C., VOSS, A., SCHMITZ, F., NUSZBAUM, M., TÜSCHER, O., LIEB, K. & KLAUER, K.C. (2014). Behavioral components of impulsivity. *Journal of Experimental Psychology: General*, *143* (2), 850-886.

VAN DE LAAR, M.C., VAN DEN WILDENBERG, W.P., VAN BOXTEL, G. & VAN DER MOLEN, M. (2011). Lifespan changes in global and selective stopping and performance adjustments. *Frontiers in Psychology*, *2*, 357.

VARGHA, A. & DELANEY, H.D. (2000). A critique and improvement of the CL common language effect size statistics of McGraw and Wong. *Journal of Educational and Behavioral Statistics*, *25* (2), 101-132.

VERBRUGGEN, F., ARON, A.R., BAND, G.P., BESTE, C., BISSETT, P.G., BROCKETT, A.T., ... & COLZATO, L.S. (2019). A consensus guide to capturing the ability to inhibit actions and impulsive behaviors in the stop-signal task. *Elife*, *8*, e46323.

WHITESIDE, S.P. & LYNAM, D.R. (2001). The five factors model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences*, *30* (4), 669-689.

WÖSTMANN, N.M., AICHERT, D.S., COSTA, A., RUBIA, K., MÖLLER, H.J. & ETTINGER, U. (2013). Reliability and plasticity of response inhibition and interference control. *Brain and Cognition, 81* (1), 82-94.

ZUCKERMAN, M., KUHLMAN, D.M., THORNQUIST, M. & KIERS, H. (1991). Five (or three) robust questionnaire scale factors of personality without culture. *Personality and Individual Differences*, *12* (9), 929-941.