19.00 €

vol. LXVIII. January-April 2020

N° 287

1/3

Four-monthly Journal
ISSN 0006-6761

BPA

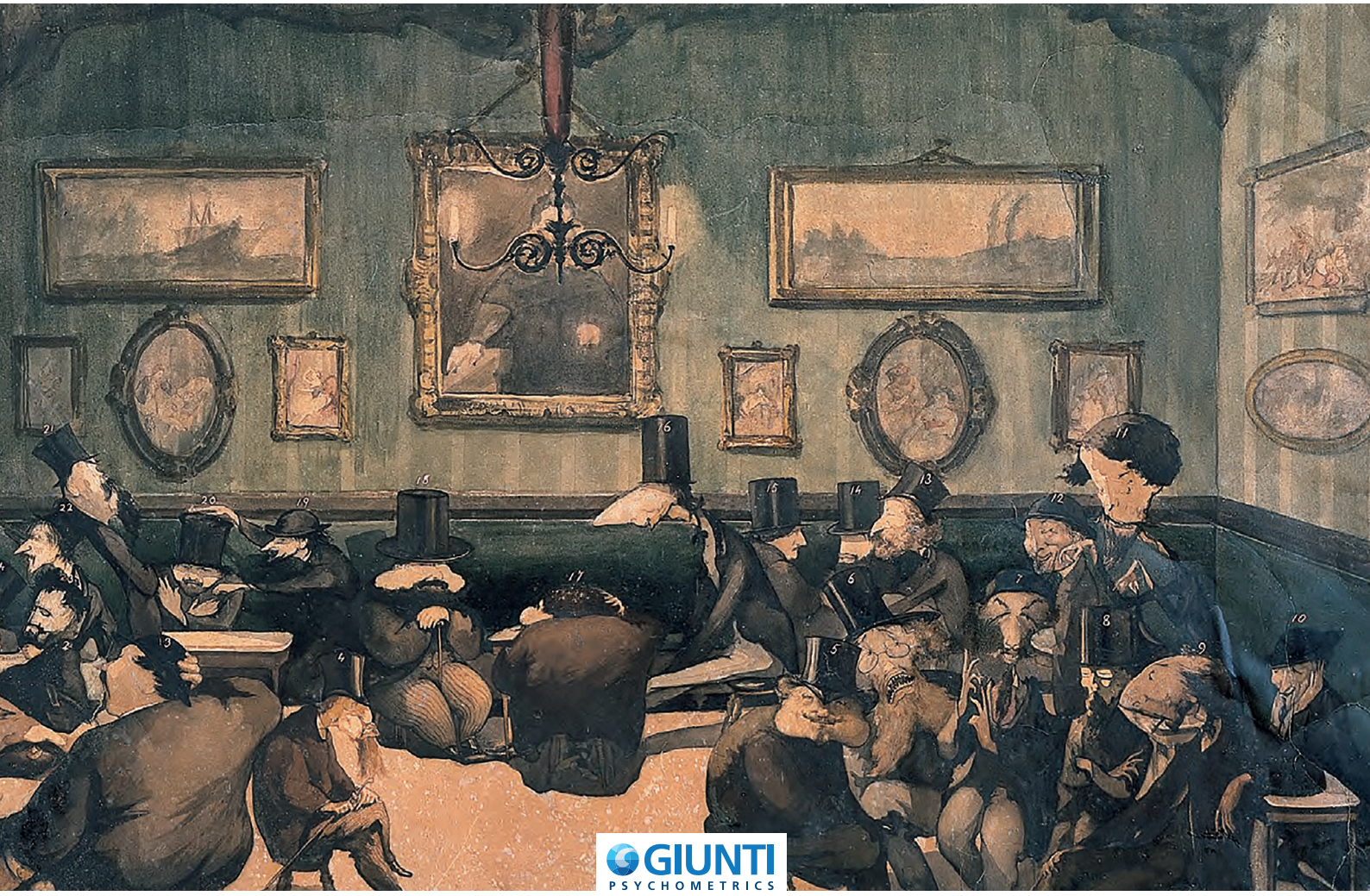
Bollettino di Psicologia Applicata

APPLIED PSYCHOLOGY BULLETIN

Indexed in PsycINFO® – Scopus Bibliographic Database

Research

Experiences & Tools



Scientific Director Alessandro Zennaro

CONTENTS

⬢ Research

▲ Experiences & Tools

DOI: 10.26387/bpa.287.1

Predictors of psychotherapy outcome in adolescents and young adults treated in routine clinical care

Paola Monaci[1], Monica Littizzetto[1], Serena Zucchi[1], Angelo d'Errico[2]

[1] Developmental Psychology Service, Local Health Unit TO3, Piedmont Region, Collegno, Turin
[2] Epidemiology Department, Local Health Unit TO3, Piedmont Region, Grugliasco, Turin

angelo.derrico@epi.piemonte.it

ABSTRACT. Una percentuale sostanziale di adolescenti trattati con psicoterapia non mostra miglioramenti nei sintomi mentali, per ragioni che ancora non sono ben comprese. Gli obiettivi dello studio erano di valutare la mancanza di guarigione, definita in base ai valori di riferimento esterni, e il cambiamento nella sintomatologia di un gruppo di adolescenti e giovani adulti trattati con psicoterapia in un ambito ambulatoriale ordinario, nonché di identificare caratteristiche sociodemografiche, cliniche e terapeutiche che potessero influenzare negativamente l'esito del trattamento. La popolazione dello studio era composta da 148 soggetti, trattati presso i Centri di Psicoterapia degli adolescenti che avevano compilato un questionario per la valutazione dei sintomi psicologici (CBA-VE) prima dell'inizio della psicoterapia (T0), subito dopo (T1) e, limitatamente a 94 pazienti, sei mesi dopo la sua fine (T2). Età, sesso, gruppo diagnostico, gravità dei sintomi, terapia farmacologica, numero e frequenza delle sedute di psicoterapia, sono stati esaminati come possibili predittori della mancanza di recupero e del miglioramento dei sintomi mentali tra pre- e post-trattamento, rispettivamente attraverso modelli multivariati di regressione robusta di Poisson e di regressione multipla. Un miglioramento significativo dei sintomi mentali si era verificato sia a T1 e T2. Tuttavia, una maggiore gravità dei sintomi a T0 incrementava la probabilità di mancata guarigione, mentre era associata a un maggiore miglioramento dei sintomi, indicando che l'impatto della psicoterapia, in termini di riduzione dei sintomi, sarebbe più forte quando la gravità è più alta.

SUMMARY. A substantial proportion of adolescents treated with psychotherapy does not show improvements in mental symptoms, for reasons which are still not well understood. Aims of the study were to evaluate lack of recovery, defined according to external reference values, and change in symptomatology of a group of adolescents and young adults treated with psychotherapy in routine clinical care, as well as to identify sociodemographic, clinical and therapeutic characteristics that can negatively influence the treatment outcome. The study population was composed of 148 subjects, treated at a Psychotherapy Center, who filled in a questionnaire for the assessment of psychological symptoms (CBA-VE) before start of psychotherapy (T0), right after (T1), and, limited to 94 patients, six months after its end (T2). Age, gender, diagnostic group, symptom severity, drug therapy, number and frequency of psychotherapy sessions, were examined as possible predictors of lack of recovery and of improvement in mental symptoms between pre- and post-treatment, through multivariate Poisson robust regression and multiple regression models, respectively. A significant improvement of mental symptoms occurred both at T1 and T2. However, higher severity of symptoms at T0 increased lack of recovery, whereas it was associated with greater improvement of symptoms, indicating that the impact of psychotherapy, in terms of reduction of symptoms, would be stronger when severity is higher.

Keywords: *Psychotherapy, Adolescent psychopathology, Evaluation, Effectiveness, Predictors*

INTRODUCTION

In the last decades, interest in adolescent mental health has been fostered by findings showing its high prevalence, which has been reported to be around 10-15% in most developed countries (Costello, Mustillo, Erkanli, Keeler & Angold, 2003; Polanczyk, Salum, Sugaya, Caye & Rohde, 2015; Roberts, Attkisson & Rosenblatt, 1998). Mental disorders are nowadays among the health problems carrying the greatest burden in adolescents and young adults, in terms of disability-adjusted life years (DALYs), especially in high-income countries (Erskine et al., 2015). Worldwide, the most common mental disorders in children and adolescents are anxiety disorders, followed by attention-deficit hyperactivity disorders (ADHD) and depressive disorders (Polanczyk et al., 2015).

Based on the results of several studies on adolescents' neurological development, they would be less "competent" in monitoring and controlling their behaviours, and in general in planning and regulating their activities, because of incomplete maturation of the pre-frontal cerebral cortex (Casey, Getz & Galvan, 2008; Jadhav & Boutrel, 2019). The greater predisposition of adolescents to higher emotional instability and in engaging in risky behaviours makes necessary that psychological public services diagnose and treat as early as possible adolescents and young adults, in order to avoid the occurrence of emotional breakdown and to reduce the likelihood of developing overt mental disorders or other health consequences, such as substance use, suicides attempts, intentional and unintentional injuries (Costello, 2016). Also, a high proportion of cases arisen in adolescence will still have a chronic mental disorder in adulthood (Ramsaw, Weisberg, Dyck, Stout & Keller, 2011), which highlights the importance of preventive or therapeutic interventions in the affected adolescents, as mental disorders are strongly associated with low social functioning and low educational attainment (de Lijster et al., 2018; Essau, Conradt & Petermann, 2000), as well as with future disability (Costello & Maughan, 2015). It is therefore worrying that only a relatively small proportion of adolescents with psychological problems receives adequate treatment, even in developed countries (Tick, van der Ende & Verhulst, 2008), and that among those treated almost 20% drop out therapy (Linardon, Fitzsimmons-Craft, Brennan, Barillaro & Wilfley, 2018).

The available therapeutic interventions include drug therapy and psychotherapy, although it has been commented that the former should be limited to more severe cases, as safety of some drugs in adolescents has been questioned, in particular that of antidepressants, in the light of an increased risk of suicide among users (Shain & Committee on Adolescence, 2016). Regarding psychotherapy, there are a number of trials documenting improvement of mental symptoms in adolescents affected by different types of disorders, including anxiety, depression, obsessive-compulsive and behavioural disorders (see for reviews: Cartwright-Hatton, Roberts, Chitsabesan, Fothergill & Harrington, 2004; Higa-McMillan, Francis, Rith-Najarian & Chorpita, 2016; Sigurvinsdóttir, Jensínudóttir, Baldvinsdóttir, Smárason & Skarphedinsson, 2019; Weersing, Jeffreys, Do, Schwartz & Bolano, 2017; Zhou et al., 2015). Several guidelines and systematic reviews on psychotherapy efficacy in adolescents have been published in the last decade, recommending specific forms of therapy for the different disorders (Bandelow, Michaelis & Wedekind, 2017; Higa-McMillan et al., 2016; NICE, 2017a, 2017b).

A major issue is that a considerable proportion of adolescents treated with psychotherapy does not show any improvement of mental symptoms, and the reasons why this occurs are still not well understood. Furthermore, in studies where psychotherapy was administered to adolescents within routine clinical care the level of improvement was lower, compared to that obtained in randomized clinical trials (RCTs), which makes even more urgent to better characterize those subjects for whom psychotherapy administered within usual care was not effective (Smith & Jensen-Doss, 2017).

The identification of predictors and moderators of treatment outcome may be limited by the selection of patients enrolled in RCTs, which apply several exclusion criteria to limit heterogeneity of the study population and are characterized by more specific and structured procedures of treatment delivery, in order to reduce dropouts and allow a more straightforward interpretation of findings. Instead, patients who turn to outdoor psychological care in routine clinical practice are generally unselected and more heterogeneous in terms of familiar, social, clinical, and psychological characteristics, which may help identifying predictors of outcome because of wider variability in these covariates. Furthermore, studies conducted within routine clinical practice could help identifying, better than RCTs, organizational peculiarities and constraints, typical of the public mental health care, that may negatively influence treatment outcome, such as a more varying theoretical foundation of therapists, or a reduced accessibility of patients to treatment, due to limited financial and human resources (long waiting lists, tight schedules, less frequent sessions).

A gradual diffusion of the Evidence-based practice in psychology (EBPP) has occurred in the last decades, defined as the "integration of the best research available with clinical expertise in the context of patient characteristics, culture, and preferences" (APA Presidential Task Force on Evidence-Based Practice, 2006, p. 273). The EBPP promotes the application of research in routine practice and in public health psychology, based on the empirical observation of outcomes and on the impact of routine practices at the local level, which may lead to improvements of therapeutic quality.

Many studies have investigated the role of sociodemographic, clinical and therapeutic characteristics as predictors or moderators of response to treatment, such as age, gender, family composition and socioeconomic status, parental psychopathology, severity of symptoms, comorbidities, type of treatment, length and frequency of the psychotherapy. However, results appear inconsistent for most of these characteristics and for the different types of disorders, including depression (Nilsen, Eisemann & Kvernmo, 2013; Weersing et al., 2017), anxiety disorders (Higa-McMillan et al., 2016; Lundkvist-Houndoumadi, Hougaard &Thastum, 2014; Nilsen et al., 2013), and obsessive-compulsive disorders (OCD) (Ginsburg, Kingery, Drake & Grados, 2008; McGuire et al., 2015; Turner, O'Gorman, Nair & O'Kearney, 2018). Such inconsistencies may be, at least in part, related to the way the outcome was operationalized in the different studies, in particular if the degree of change in symptoms between pre- and post-treatment was examined, or if lack of recovery was assessed, based on reference values from the general population. Potential predictors may be divergent for these two types of outcomes, as shown, for example, by the results of studies on predictors of anxiety disorders included in the review by Lundkvist-Houndoumadi et al. (2014), where higher severity of symptoms at pre-treatment was associated with lack of recovery in the opposite direction than with change in symptoms level.

Main objective of the present study was to assess, through a standardized methodology, effectiveness of psychotherapy in a group of adolescents and young adults treated for psychological disorders in an outpatient setting, in order to identify predictors of poor therapy response. Given the discordant results obtained in studies examining predictors of change in the level of symptoms and those evaluating post-treatment lack of recovery, we conducted separate analyses for each outcome, with the aim of comparing the set of predictors of the two types of outcomes.

METHODS

Data collection

The study was conducted on subjects who attended a psychological consultation at the Adolescents and Young Adults Psychotherapy Center (AYAP Center) of the Local Health Unit TO3 (Turin, Italy), and afterward underwent psychotherapy. The AYAP Center, which started its activities in 2010, is dedicated to people 16-26 years old. In the Center operate six therapists, who perform individual psychotherapy based on different theoretical background, including individual systemic, psychoanalytic and cognitive behavioural therapy.

Psychotherapists in the AYAP Center treat patients using a structured protocol, divided in three phases (preliminary clinical assessment, psychotherapy supply, outcome evaluation), which guarantees homogeneity of the treatment process and comparability of the outcome, in spite of different theoretical approaches of the therapists. This homogeneity derives from several factors, including: 1) random assignment of patients to therapists, which is carried out on the basis of waiting lists and other organizational constraints; 2) treatment protocols that foresee short/medium or long-term cycles, in relation to the severity of the psychodiagnosis; 3) standardized evaluation of the outcome of the treatment.

The first phase aims on one hand, at evaluating the appropriateness of the therapeutic intervention in relation to the mental diagnosis of the subject, because psychotic and severe affective disorders are not treated in the AYAP Center, but are referred to psychiatric services; on the other hand, at verifying patients' condition, in terms of awareness of the problem, stability of the symptomatology, motivation, and ability of therapeutic alliance.

The treatment phase includes short/medium-term (8-24 sessions) or long-term (25-40 sessions) cycles of individual psychotherapy, with the latter offered to patients characterized by more severe symptoms and greater clinical complexity, assessed through the presence of comorbidities, critical life events, social situation, use of drugs, etc. Length of psychotherapy is planned, agreed upon and communicated to the patients at the start of the treatment. However, based on clinical experience, part of the patients with anxiety disorders tend to prolong their relationship with the psychotherapist, independently of their clinical improvement. Some patients, especially those with more severe symptoms, are also given

drug therapy, in collaboration with psychiatric services, in particular anxiolytics, antidepressants and mood stabilizers.

At the end of the psychotherapy, patients' symptomatology is re-evaluated using the same protocol as for the pre-treatment evaluation, and, subsequently, all subjects are contacted and invited to participate to a further assessment six months after the end of the psychotherapy, also conducted using the same protocol.

Participants

From 2010 to 2017, 486 patients were treated at the AYAP Center, 88 of which dropped out before the end of the therapy for different reasons (18.1%), including life events or unmet expectations regarding the treatment.

Of the 398 remaining cases, 148 subjects (49 males and 99 females) concluded the psychotherapy and filled in the *Cognitive Behavioural Assessment – Outcome evaluation (CBA-VE)* questionnaire before the beginning of the psychotherapy (T0) and after its end (T1), while for 94 of them was also available an interview at six months after the end of the psychotherapy (T2). Of the 148 patients included in the study, 115 had been treated using an individual systemic approach, 27 using Cognitive Behavioural Therapy (CBT), and six through psychoanalytic therapy. Assignment of the patients to a specific type of therapy was determined, as discussed above, by the schedule of the therapists working at the AYAP center, rather than by a matching between the type of psychopathology affecting each patient and a specific therapy approach.

It is worth underlining that the relatively small proportion of subjects with 2- or 3-time assessment was due to the fact that the outcome evaluation through CBA-VE was introduced gradually during the period 2010-2017 and was performed systematically only in the last three years.

Patients' disorders were classified, according to the ICD-10 classification, in four groups:

1) anxiety disorders, including anxious, phobic, obsessive-compulsive and somatoform syndromes;
2) personality and behavioural disorders;
3) acute stress reactions;
4) other syndromes.

The first two groups are generally characterized by recurrent psychological disturbances, whereas acute stress reactions are more commonly single episodes consequent to the exposure to psychological strain or adverse live events.

Pure depressive syndromes have been observed rarely in our patient series, while they have been found more frequently associated with anxiety, and such cases have been classified as affected by anxiety disorders.

Outcome assessment

Outcome evaluation is based on a questionnaire developed at University of Padua, the CBA-VE (Michielin et al., 2008), which is administered before, at the end, and after 6 months from the end of the psychotherapy. The CBA-VE is a self-administered questionnaire on the perceived psychological condition during the previous 15 days, which is composed of 80 items with answers on a 5-point Likert scale (from "not at all" to "very much"), and scores going from 1 to 5. Scores are aggregated in five scales, including one assessing well-being (14 items), one awareness of positive change (11 items), and three evaluating symptoms severity, including the dimensions of Anxiety (14 items), Depression (19 items), and Distress (21 items) (the latter also defined as "reduced impulse control"). Therefore, the range of possible scores is 14-70 for Well-being and for Anxiety, 11-55 for Perceived change, 19-95 for Depression and 21-105 for Distress. Scores in the scales of Well-being and Perceived change are reversed, compared to scores in the other scales.

The tool has been validated in previous studies through analyses aimed at evaluating discriminant and concurrent validity, sensitivity to change, as well as at verifying structure and internal consistency of the five composing dimensions (Michielin et al., 2008; Michielin, Bertolotti, Sanavio, Vidotto & Zotti, 2009). Regarding discriminant validity, for all five dimensions significant differences in average scores between normal and clinical subjects at pre-treatment have been observed. The analyses aimed at evaluating concurrent validity demonstrated strong correlations of both the Anxiety and the Depression scale with the State-Trait Anxiety Inventory Form Y2 (STAI-Y2) and the CES-D scores, respectively. Test-retest reliability, assessed at 1-month distance, was satisfactory (all Pearson's correlation coefficients $\geq .75$), as it was internal consistency of the scales, with all Cronbach alphas around .90 among both patients and normal subjects, except for the Perceived change scale (alpha = .80 among patients and .74 among controls).

Cut-off values of the different scales, which allow to identify subjects with abnormal scores, have been set from reference values of the Italian population of similar age: Well-being score<21; Perceived change score<20; Anxiety score>24; Depression score>31; Distress score>25 (Michielin et al., 2009).

Data analysis

Differences in scores of the scales between T0 and T1, and between T0 and T2, were assessed by means of paired t-tests, in order to evaluate statistical significance of the changes occurred in the different dimensions after psychotherapy in the overall study population.

Lack of recovery in the different dimensions after the end of the psychotherapy was defined using the cut-offs established for the Italian population (see above). Poisson multivariable regression models with robust variance were used to estimate relative risks of lack of recovery associated with each demographic and clinical characteristic of the patients at pre-treatment, adjusting the analyses for the other available covariates. Prevalence Ratios (PRs) obtained through Poisson robust regression models, which use the Huber-White sandwich estimator of variance, have been demonstrated to provide correct estimates of relative risks in studies where a high prevalence of the outcome (i.e. above 10%), as was the case for some of the dimensions analysed, limits the possibility of using logistic regression models (Barros & Hirakata, 2003).

In separate analyses, predictors of changes in the scores of the scales between pre- (T0) and post-treatment (T1), and between pre-treatment (T0) and follow-up (T2) were also assessed, through multiple regression models, also adjusted for the other covariates, in which the dependent variable was the difference in scores of each scale between T0 and T1, and between T0 and T2, and the independent variables were the same as in the analyses on lack of recovery.

The characteristics evaluated as predictors in all these analyses included age class (≤18, >18 years), gender, type of psychological disorder (four groups: 1. anxiety disorders; 2. acute stress reactions; 3. personality and behavioural disorders; 4. other syndromes), use of drugs for mental symptoms (yes/no), number (≤24, 25-35, ≥36) and frequency (weekly, bi-monthly, monthly) of the psychotherapy sessions, and symptoms score of the corresponding scale at T0, treated as a continuous measure (Michielin et al., 2009). Multicollinearity between independent variables in the regression models was assessed through the Variance Inflation Factor test (VIF).

RESULTS

In Table 1, the frequency distribution at T0 of the characteristics of the study population is presented, in terms of age class, gender, type of psychological disorder, use of drugs for mental symptoms, number and frequency of psychotherapy sessions. Two-third of patients were females, 68% were older than 18 years, and almost 20% were prescribed drugs for mental symptoms. More than half of the patients had been diagnosed with anxiety disorders, 20% with acute stress reactions, 7% with personality and behavioural disorders and 15% with other syndromes. Regarding psychotherapy, 59% of the patients received less than 25 sessions, 22% between 25 and 35 sessions, and 19% more than 35 sessions, while the frequency of the psychotherapy was weekly or bi-monthly for 95% of the patients.

At all three times (T0, T1, T2), scores in the different scales were strongly correlated, with Pearson's correlation coefficients ranging from −.39 for the relation between scores of Anxiety and Perceived change at T0, to .86 for that between Depression and Distress scores at T2. Also, the Anxiety scale was strongly correlated with those of Depression and Distress (r = .74 for both).

The proportion of subjects who did not recover, i.e. whose scores were still above reference values after the psychotherapy (or below for Well-being and Perceived change, whose scores were reversed), was quite variable across dimensions, going from 6.8% for the Depression scale to 20.7% for that of Perceived change at T1, and from 8.5% for the Well-being scale to 23.4% for that of Perceived change at T2 (see Table 2).

For all five dimensions, a significant improvement was found between T0 and T1, as assessed through the paired t-test, with the strongest decrease in the level of symptoms observed for Depression, followed by Distress and Anxiety (see Table 3). Similar improvements in the scales were also found between T0 and T2 interviews.

Table 4 shows Prevalence Ratios (PRs) of lack of recovery at T1 for each dimension, in relation to patients' characteristics, obtained through models adjusted for the other covariates. Older age class (>18 years) was associated with a more than 50% reduced risk of lack of recovery for all

Table 1 – Frequency distribution of the study population characteristics at pre-treatment (T0)

Covariates	N	%
Gender		
Males	49	33.1
Females	99	66.9
Age		
≤18 years	47	31.8
>18 years	101	68.2
Type of psychological disorder		
Anxiety disorders	85	57.4
Acute stress reactions	29	19.6
Personality and behavioural disorders	11	7.4
Others	23	15.5
Drug use		
Yes	26	17.6
No	121	81.8
Missing	1	.7
Number of sessions		
≤24	87	58.8
25-35	33	22.3
≥36	28	18.9
Frequency of sessions		
Weekly	53	35.8
Bi-monthly	88	59.5
Monthly	7	4.7

Table 2 – Proportion of subjects with abnormal scores at pre-treatment (T0), at the end of psychotherapy (T1) and at 6-month follow-up (T2) for each dimension

	T0 (N = 148)	**T1 (N = 148)**	**T2 (N = 94)**
Psychological dimension	N (%)	N (%)	N (%)
Anxiety	84 (56.8)	19 (12.8)	17 (18.1)
Well-being	58 (39.2)	12 (8.1)	8 (8.5)
Perceived change	88 (59.5)	30 (20.7)	22 (23.4)
Depression	61 (41.2)	10 (6.8)	9 (9.6)
Distress	75 (50.7)	19 (12.8)	13 (13.8)

Table 3 – Change in average scores for each psychological dimension between T0 and T1 (148 subjects), and between T0 and T2 (94 subjects)

Psychological dimension	**T0**	**T1**	**Difference T0-T1[a]**		**T0**	**T2**	**Difference T0-T2[a]**	
	M (SD)	*M (SD)*	*M (SD)*	*t-values*	*M (SD)*	*M (SD)*	*M (SD)*	*t-values*
Anxiety	26.7 (11.8)	14.9 (9.1)	11.8 (12.3)	11.6	26.0 (11.5)	14.4 (9.7)	−11.7 (13.4)	8.5
Well-being	24.2 (9.7)	34.2 (9.8)	10.0 (10.2)	−11.9	24.4 (9.6)	34.6 (10.5)	10.1 (12.4)	−7.9
Perceived change	18.4 (7.1)	24.9 (6.7)	6.51 (8.1)	−9.8	18.1 (6.7)	25.0 (7.1)	6.9 (8.6)	−7.7
Depression	29.2 (14.3)	15.0 (9.8)	−14.1 (13.8)	12.5	28.3 (13.5)	14.8 (10.3)	−13.5 (15.2)	8.6
Distress	28.0 (15.3)	15.2 (9.7)	−12.8 (14.2)	11.0	27.1 (14.5)	14.4 (11.6)	−12.7 (16.1)	7.7

Note. [a] $p<.0001$ for all differences in scores between T0-T1 and T0-T2 (paired *t*-test).

Table 4 – Prevalence Ratios (PR) of lack of recovery at T1 associated with patients' and therapy characteristics. Robust Poisson regression models adjusted for age class, gender, type of disorder, drug use, scale score at T0, number and frequency of psychotherapy sessions (148 subjects)

Covariates	Anxiety PR (95% CI)	Well-being PR (95% CI)	Perceived change PR (95% CI)	Depression PR (95% CI)	Distress PR (95% CI)
Gender (ref: Males)	1	1	1	1	1
Females	1.84 (.54-6.29)	4.73 (.63-35.7)	1.50 (.64-3.55)	1.95 (.53-7.18)	1.41 (.42-4.69)
Age (ref: ≤18 years)	1	1	1	1	1
>18 years	.40 (.17-.98)	.06 (.02-.20)	.46 (.23-.92)	.14 (.04-.58)	.42 (.15-1.19)
Type of psychological disorder (ref: Anxiety disorders)	1	1	1	1	1
Acute stress reaction	.18 (.04-.83)	.35 (.10-1.20)	.93 (.35-2.46)	.06 (.01-.45)	.21 (.03-1.40)
Personality and behavioural disorders	1.40 (.22-9.05)	–	.96 (.29-3.13)	1.90 (.18-19.9)	1.14 (.23-5.78)
Others	.97 (.39-2.40)	1.21 (.37-3.90)	.97 (.41-2.32)	.76 (.21-2.71)	.84 (.30-2.35)
Drug use (ref: No)	1	1	1	1	1
Yes	.92 (.29-2.89)	2.60 (.48-14.1)	1.66 (.81-3.39)	.70 (.20-2.44)	.44 (.08-2.46)
Number of sessions (ref: ≤24)	1	1	1	1	1
25-35	1.92 (.83-4.43)	3.54 (.83-15.0)	2.43 (1.21-4.88)	4.33 (1.18-15.9)	2.18 (.82-5.79)
≥36	.71 (.12-4.35)	15.6 (2.48-98.7)	1.26 (.44-3.62)	–	.58 (.09-3.86)
Frequency of sessions (ref: Weekly)	1	1	1	1	1
Bi-Monthly	1.31 (.57-3.05)	1.71 (.48-6.00)	1.78 (.84-3.78)	.93 (.31-2.77)	1.09 (.42-2.79)
Monthly	1.32 (.20-8.61)	7.36 (.93-58.4)	1.01 (.12-8.69)	–	–
Scale score at T0					
1-point increase[1]	1.05 (1.01-1.09)	1.21 (1.09-1.34)	1.04 (.99-1.09)	1.10 (1.04-1.16)	1.06 (1.03-1.09)

Note. [1]1-point decrease for the scores of Well-being and Perceived change scales.

the dimensions, although not significant for the Distress scale. Women were at higher risk of lack of recovery, compared to men, for all the dimensions, but the difference never reached statistical significance. Regarding clinical diagnoses, subjects affected by "acute stress reactions" showed PRs<.5 for all scales, except for the Perceived change dimension (PR = .93), with risks significantly reduced for Anxiety and Depression.

Unexpectedly, a higher number of psychotherapy sessions was significantly associated with an increased risk of lack of recovery at T1 for the Well-being (PR = 15.6 for the highest category), the Perceived change (PR = 2.43 for the intermediate category), and the Depression scales (PR = 4.33 for the intermediate category); in contrast, no association was found with psychotherapy frequency.

Greater severity of symptoms in each scale at T0 was associated with a higher risk of lack of recovery at T1 in the corresponding scale, especially for Well-being (PR = 1.21 for 1-point decrease in Well-being at T0) and Depression (PR = 1.10 for 1-point increase in Anxiety at T0), although the association with Perceived change was not statistically significant (p = .14).

Regarding lack of recovery at T2 (see Table 5), the pre-treatment characteristics associated were mostly similar to those identified for lack of recovery at T1, but the smaller number of subjects did not allow observing as much significant associations as for lack of recovery at T1. Like for the analysis on lack of recovery at T1, older age, male gender and a diagnosis of "acute stress reactions" showed a positive effect on recovery for most of the dimensions assessed, and a negative one for having received more than 24 psychotherapy sessions. Severity of symptoms in the corresponding dimension at T0 was a less strong predictor of lack of recovery at T2, especially for the Well-being and the Depression scales.

The results of the analyses examining differences in the continuous scores of the five dimensions between T0 and T1, and between T0 and T2 (see Table 6 and Table 7, respectively), were mostly consistent with those obtained dichotomizing the outcome at cut-offs based on normative values. However, a striking difference was found for severity of symptoms at T0, which was significantly associated with changes in the scores of all scales in the opposite direction than that observed in the analysis on lack of recovery, i.e. worse scores at T0 were predictive of larger improvement in the scale scores at post-treatment and follow-up.

No multicollinearity was found in any regression model, with all VIF values <2.

DISCUSSION

The present study showed good effectiveness of psychotherapy in adolescents and young adults, with only a minor proportion of subjects with abnormal scores in the five dimensions examined at the end of the treatment, ranging from 7% for depressive scores to 21% for scores of the Perceived change dimension. The rather low proportion of lack of recovery observed, compared to other studies (Bodden et al., 2008; Cobham, Dadds, Spence & McDermott, 2010; Kendall et al., 1997; Legerstee et al., 2009; Southam-Gerow, Kendall & Weersing, 2001), likely depends on the fact that access to the AYAP Center is restricted to adolescents and young adults affected by relatively mild symptoms, while more severe cases are referred to a psychiatric outpatient clinic.

Among the dimensions examined through the CBA-VE questionnaire, at both T1 and T2 the scale that showed the highest proportion of lack of recovery, as well as the smallest improvement in scores between T0-T1 and T0-T2, was that of Perceived change, which derives from the psychological construct of awareness of functioning modification (Michielin et al., 2008). According to the theoretical model by Howard, Lueger, Maling and Martinovich (1993), during the course of therapy psychological change develops in three phases: at first, improvement would occur in well-being, then in symptomatology, and, last, in behavioural functioning. Therefore, as patients' perception of better functioning would occur as the last step of psychological recovery, Perceived change would be also the dimension less likely to be modified by the therapy, which would possibly necessitate of a longer, more personalized and intense treatment, to obtain further improvement.

Regarding predictors, several characteristics were found significantly associated with lack of recovery at the end of the treatment (T1) in at least one of the dimensions examined, including higher severity of symptoms and a higher number of psychotherapy sessions, while being older than 18 years and having a diagnosis of acute stress disorder were protective factors. Females were also at higher risk of lack of recovery in all scales, but never significantly. Results from the follow-up assessment (T2) revealed similar findings, with associations between a higher number of sessions and lack of recovery even more consistent across dimensions.

Concerning severity of symptoms at pre-treatment, most studies that examined it in relation to post-treatment symptomatology, as in the present study, found higher

Table 5 – Prevalence Ratios (PR) of lack of recovery at T2 associated with patients' and therapy characteristics. Robust Poisson regression models adjusted for age class, gender, type of disorder, drug use, scale score at T0, number and frequency of psychotherapy sessions (94 subjects)

Covariates	Anxiety PR (95% CI)	Well-being PR (95% CI)	Perceived change PR (95% CI)	Depression PR (95% CI)	Distress PR (95% CI)
Gender (ref: Males)	1	1	1	1	1
Females	1.07 (.35-3.33)	1.83 (.42-8.00)	1.96 (.66-5.81)	∞[1]	3.44 (.96-12.4)
Age (ref: ≤18 years)	1	1	1	1	1
>18 years	.50 (.17-1.49)	.23 (.04-1.36)	.65 (.24-1.71)	.24 (.05-1.13)	.28 (.09-.89)
Type of psychological disorder (ref: Anxiety disorders)	1	1	1	1	1
Acute stress reaction	.62 (.12-3.26)	.21 (.02-2.23)	.64 (.18-1.19)	–	–
Personality and behavioural disorders	1.37 (.26-7.16)	–	.37 (.06-2.16)	–	.49 (.08-3.01)
Others	.59 (.16-2.17)	.31 (.07-1.32)	.39 (.12-1.27)	.66 (.20-2.15)	.32 (.11-.96)
Drug use (ref: No)	1	1	1	1	1
Yes	1.09 (.49-2.42)	1.54 (.30-7.91)	1.38 (.61-3.13)	.34 (.08-1.47)	1.03 (.35-3.03)
Number of sessions (ref: ≤24)	1	1	1	1	1
25-35	3.38 (1.31-8.71)	6.02 (.83-43.9)	1.73 (.66-4.55)	3.42 (.67-17.4)	2.68 (.87-8.25)
≥36	1.55 (.41-5.90)	6.96 (.93-51.9)	1.90 (.69-5.19)	2.44 (.38-15.7)	3.07 (.75-12.5)
Frequency of sessions (ref: Weekly)	1	1	1	1	1
Bi-monthly	.82 (.31-2.17)	1.89 (.54-6.63)	1.27 (.51-3.17)	1.37 (.38-4.96)	.77 (.27-2.22)
Monthly	–	–	3.59 (.37-35.0)	–	–
Scale score at T0					
1-point increase[2]	1.04 (1.00-1.08)	1.04 (.95-1.13)	1.04 (0.99-1.10)	1.05 (1.00-1.10)	1.06 (1.01-1.12)

Note. [1] Gender not included in the regression model on depression, as nine females and no males did not recover at T2, Fisher exact test: $p = .05$.

[2] 1-point decrease for the scores of the Well-being and Perceived change scales.

Table 6 – Changes in the scales scores between T0 and T1 associated with patients' and therapy characteristics. Multiple regression models adjusted for age class, gender, type of disorder, drug use, scale score at T0, number and frequency of psychotherapy sessions (148 subjects)

Covariates	Anxiety beta (p-value)	Well being beta (p-value)	Perceived change beta (p-value)	Depression beta (p-value)	Distress beta (p-value)
Gender (ref: Males)					
Females	1.08 (.45)	−2.90 (.07)	−.34 (.77)	2.15 (.15)	.48 (.73)
Age (ref: ≤18 years)					
>18 years	−4.09 (.02)	2.69 (.11)	2.00 (.09)	−4.21 (.01)	−1.97 (.24)
Type of psychological disorder (ref: Anxiety disorders)					
Acute stress reaction	−3.17 (.11)	.64 (.73)	−.19 (.89)	−2.81 (.15)	−3.18 (.10)
Personality and behavioural disorders	−1.06 (.75)	−1.57 (.53)	.38 (.88)	.86 (.76)	.86 (.74)
Others	1.24 (.56)	−.35 (.90)	.59 (.68)	.48 (.84)	.77 (.74)
Drug use (ref: No)					
Yes	−.37 (.85)	.21 (.92)	−1.57 (.33)	−1.02 (.60)	−1.63 (.34)
Number of sessions (ref: ≤24)					
25-35	5.06 (.02)	−4.53 (.01)	−3.10 (.02)	5.54 (.01)	3.06 (.15)
≥36	2.63 (.12)	−2.98 (.14)	−1.04 (.49)	2.08 (.27)	−1.52 (.41)
Frequency of sessions (ref: Weekly)					
Bi-monthly	1.54 (.32)	−.93 (.55)	−1.12 (.30)	.32 (.84)	1.01 (.53)
Monthly	.77 (.79)	−4.73 (.13)	−3.95 (.08)	−.87 (.78)	.21 (.94)
Scale score at T0					
1-point increase[1]	−.79 (<.001)	.57 (<.001)	.71 (<.001)	−.75 (<.001)	−.73 (<.001)

Note. [1] 1-point decrease for the scores of the Well being and Perceived change scales.

Table 7 – Changes in the scales between T0 and T2 associated with patients' and therapy characteristics. Multiple regression models adjusted for age class, gender, type of disorder, drug use, scale score at T0, number and frequency of psychotherapy sessions (94 subjects)

Covariates	Anxiety beta (p-value)	Well being beta (p-value)	Perceived change beta (p-value)	Depression beta (p-value)	Dstress beta (p-value)
Gender (ref: Males)					
Females	1.72 (.48)	−3.91 (.10)	.82 (.58)	3.33 (.13)	1.00 (.70)
Age (ref: ≤18 years)					
>18 years	−4.87 (.03)	6.83 (.01)	1.10 (.54)	5.56 (.02)	−3.96 (.11)
Type of psychological disorder (ref: Anxiety disorders)					
Acute stress reaction	−6.63 (.02)	4.04 (.16)	.16 (.94)	−7.12 (<.01)	−6.40 (.03)
Personality and behavioural disorders	−3.93 (.39)	5.56 (.10)	1.05 (.76)	−4.50 (.29)	.96 (.87)
Others	−2.27 (.42)	4.63 (.12)	2.05 (.35)	1.78 (.62)	−3.17 (.35)
Drug use (ref: No)					
Yes	1.63 (.48)	−3.28 (.28)	−.27 (.88)	1.42 (.57)	1.90 (.45)
Number of sessions (ref: ≤24)					
≥25-35	4.86 (.10)	−6.42 (.02)	−2.48 (.26)	4.88 (.13)	3.34 (.35)
≥36	5.80 (.02)	−4.48 (.08)	−1.71 (.34)	5.56 (.02)	3.89 (.23)
Frequency of sessions (ref: Weekly)					
Bi-monthly	−.75 (.76)	−.51 (.81)	−1.17 (.46)	.93 (.70)	.32 (.91)
Monthly	1.54 (.58)	−6.35 (.04)	−5.39 (.24)	−.29 (.94)	−2.38 (.52)
Scale score at T0					
1-point increase[1]	−.90 (<.001)	.81 (<.001)	.77 (<.001)	−.87 (<.001)	−.82 (<.001)

Note. [1] 1-point decrease for the scores of the Well being and Perceived change scale.

proportions of lack of recovery among subjects with higher pre-treatment levels of symptoms, consistently with our results (Berman, Weems, Silverman & Kurtines, 2000; Clarke et al., 1992; Curry et al., 2006; Festen et al., 2013; Liber et al., 2010; Southam-Gerow et al., 2001; Torp et al., 2015). In contrast, in studies examining the change in mental scores between pre- and post-treatment, results are mixed, with studies showing non-significant change (Liber et al., 2010), smaller (Piacentini, Bergman, Jacobs, McCracken & Kretchman, 2002; Rudy, Lewin, Geffken, Murphy & Storch, 2014), or even larger improvement in scores among subjects with higher severity of symptoms at pre-treatment (Berman et al., 2000; Kerns, Read, Klugman & Kendall, 2013; Kley, Heinrichs, Bender & Tuschen-Caffier, 2012). It is worth considering that in two studies on anxiety disorders examining both types of outcomes, discordant results were observed: in one of them, no association between severity of symptoms at pre-treatment and lack of recovery at post-treatment was found, whereas higher severity was associated with a significantly larger decrease of symptoms between pre- and post-treatment (Berman et al., 2000); in the other one, higher severity at pre-treatment was associated with a higher likelihood of lack of recovery at post-treatment, whereas no association was present with change in scores (Liber et al., 2010). Although among subjects with higher severity of symptoms at pre-treatment there is potentially larger space for improvement, the discrepancy observed in our results between the two types of outcomes may be in part the consequence of a regression to the mean phenomenon, as commented by others (Lundkvist-Houndoumadi et al., 2014), due to imprecision in the measurement of symptoms at a single occasion (Barnett, van der Plos & Dobson, 2005).

Therefore, the two types of outcomes examined, i.e. change in scores between pre- and post-treatment, and lack of recovery, seem to indicate different phenomena, providing complementary information on the course of treatment of the patients, which would be both important to decide about the eventual continuation/discontinuation of the psychotherapy or the switch to a different therapy method. Our results indicate that the presumed impact of the psychotherapy, in terms of reduction of symptoms, would be stronger when the severity of the patients is greater, which may represent an important benefit for their social functioning, even if they would have a lower likelihood of full recovery.

The association observed between younger age and a poorer outcome is not in agreement with the literature, as most studies did not find an association with age, and in a few studies a significantly worse outcome was found among older subjects (Bodden et al., 2008; Jayson, Wood, Kroll, Fraser & Harrington, 1998; Southam-Gerow et al., 2001). However, the comparability with other studies is limited by the fact that most studies were conducted on populations of patients mainly composed by children and adolescents up to 18 years, while our study population did not include children and was composed for more than two-thirds of youngsters older than 18 years.

Patients diagnosed with acute stress disorders were more likely to recover at both post-treatment and follow-up, compared to those affected by anxiety disorders (who were set as the reference category, being the majority of patients), which supports a more favourable psychotherapy outcome in these patients, who are unlikely to become chronically ill (Gillies, Taylor, Gray, O'Brien & D'Abrew, 2013).

We did not find a significant association with gender, like most studies on adolescents affected by various types of disorders (Lundkvist-Houndoumadi et al., 2014; Nilsen et al., 2013; Turner et al., 2018; Weersing et al., 2017), although a worse outcome was consistently observed among women.

The finding of an association for most dimensions between a greater number of psychotherapy sessions and a lower success rate, at both post-treatment and follow-up, seems to indicate that extending the number of therapy sessions does not provide benefits to patients. In a large observational cohort of patients treated with psychotherapy in clinical care in UK, a similar association was observed, which was interpreted as attributable to higher patient's severity at pre-treatment, because of a higher likelihood of therapy discontinuation among patients who recovered (Stiles, Barkham & Wheeler, 2015). Although we adjusted the analysis for severity of symptoms in the corresponding dimension at baseline, it seems possible that the poorer outcome likelihood observed among patients treated with a higher number of sessions could have arisen from residual confounding by the level of severity at pre-treatment in other scales, which was not adjusted for in the analysis.

In any case, these results suggest the need of a reassessment of the psychological condition of the patients after a psychotherapy cycle of medium duration (such as the 24 sessions we used as a cut-off), to evaluate the opportunity of prolonging the psychotherapy, in order to avoid possible negative consequences on patients' recovery.

Furthermore, psychotherapy treatment in public care is generally constrained by the necessity to deliver it rapidly to patients, often without having the available resources to match symptoms' patients to a specific type of therapy which demonstrated greater efficacy in the literature. Because of this, it seems important to identify patients who do not recover after a first cycle of psychotherapy, also to evaluate the opportunity of changing type of therapeutic approach.

Among the strengths of the study, the availability of two sets of results computed using different measures of the outcome gave us the possibility of assessing consistency of the predictors obtained for lack of recovery with those identified for changes in scores between pre- and post-treatment.

The results of this study should be interpreted with caution, in particular because of the relative small size of the cohort and the low number of patients with abnormal scores at the end of treatment, and in particular at follow-up, which may have limited the identification of covariates significantly associated with the outcome. Furthermore, given that several regression models were built, in order to assess the association of potential predictors with lack of recovery or change in each dimension at both T1 and T2, it seems difficult to exclude that some statistically significant findings may have been the result of multiple testing. Nonetheless, given the strong correlations among the different scales at pre-treatment, especially those of Anxiety, Depression and Distress, a support to the robustness of our results comes from the quite high consistency of the predictors' estimates observed among the different scales, in particular for severity of symptoms at pre-treatment.

Another limitation is the relative small proportion of the patients assessed through the CBA-VE on the total number of patients treated after excluding dropouts (37.2%), which was however determined only by the progressive introduction of this method of assessment during the years of activity of the AYAP Center, without any apparent selection of the patients. In contrast, the assessment at follow-up was completed by less than two-third of the patients assessed at T1, so that it seems possible that the results of the analyses on recovery and scores change at follow-up (T2) have been distorted by self-selection of the patients, which could explain, at least in part, some of the differences observed between the analyses on predictors of lack of recovery and change in scores at T1 and T2.

Last, a further limitation is that most patients (115 out of 148) were treated using the individual systemic approach, so that our results on the psychotherapy outcome are expected to reflect mainly the impact of this specific type of psychotherapy, with problems of generalizability to other kinds of therapies. We did not include type of psychotherapy in the regression models, as the number of patients treated with the two other types of therapy was small, in particular those treated using psychoanalytic therapy (n = 6), with the consequence of instability of the risk estimates of other covariates in the models.

CONCLUSIONS

In conclusion, the present study found good effectiveness of psychotherapy in a population of adolescents and young adults affected by psychological symptoms. Several demographic and clinical characteristics of adolescents and young adults treated with the psychotherapy were found to predict lack of recovery, including higher severity of symptoms at pre-treatment, younger age, type of diagnosis, and higher number of sessions. The results of the analyses on differences in scores between pre- and post-treatment identified the same set of predictors, but symptoms severity at pre-treatment was associated with amount of change in the opposite direction, with greater improvement in scores among patients with higher symptoms severity. This finding of such discrepant results suggest that these two types of outcome provide complementary information on the course of treatment of the patients.

Ethical considerations

Informed consent was obtained from all subjects involved in the study. For all minors the informed consent was obtained from their parents.

Ethical approval was not sought because the present study was conducted on subjects undergoing psychotherapy within usual outpatient care, using only information collected for routine outcome evaluation.

References

APA Presidential Task Force on Evidence-Based Practice (2006). Evidence-based practice in psychology. *American Psychologist*, *61*, 271-285. PMID: 16719673.

BANDELOW, B., MICHAELIS, S. & WEDEKIND, D. (2017). Treatment of anxiety disorders. *Dialogues in Clinical Neuroscience*, *19*, 93-107. PMCID: PMC5573566

BARNETT, A.G., van der PLOS, J.C. & DOBSON, A.J. (2005). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology*, *34*, 215-220. doi:10.1093/ije/dyh299

BARROS, A.J. & HIRAKATA, V.N. (2003). Alternatives for logistic regression in cross-sectional studies: An empirical comparison of models that directly estimate the prevalence ratio. *BMC Medical Research Methodology*, *3*, 21. doi:10.1186/1471-2288-3-21

BERMAN, S.L., WEEMS, C.F., SILVERMAN, W.K. & KURTINES, W.M. (2000). Predictors of outcome in exposure-based cognitive and behavioral treatments for phobic and anxiety disorders in children. *Behavioral Therapy*, *31*, 713-731. doi:10.1016/s0005-7894(00)80040-4

BODDEN, D.H.M., BÖGELS, S.M., NAUTA, M.H., DE HAAN, E., RINGROSE, J. & APPELBOOM-GEERTS, K.C. (2008). Child versus family cognitive – behavioural therapy in clinically anxious youth: An efficacy and partial effectiveness study. *Journal of the American Academy of Child and Adolescent Psychiatry*, *47*, 1384-1394. doi:10.1097/chi.0b013e318189148e

CARTWRIGHT-HATTON, S., ROBERTS, C., CHITSABESAN, P., FOTHERGILL, C. & HARRINGTON, R. (2004). Systematic review of the efficacy of cognitive behaviour therapies for childhood and adolescent anxiety disorders. *British Journal of Clinical Psychology*, *43,* 421-436. doi:10.1348/0144665042388928

CASEY, B.J., GETZ, S. & GALVAN, A. (2008). The adolescent brain. *Developmental Review*, *28*, 62-77. doi:10.1016/j.dr.2007.08.003

CLARKE, G.N., HOPS, H., LEWINSOHN, P., ANDREWS, J., SEELEY, J. & WILLIAMS, J. (1992). Cognitive-behavioral group treatment of adolescent depression: Prediction of outcome. *Behavioral Therapy*, *23*, 341-354. doi:10.1016/S0005-7894(05)80162-5

COBHAM, V.E., DADDS, M.R., SPENCE, S.H. & McDERMOTT, B. (2010). Parental anxiety in the treatment of childhood anxiety: A different story three years later. *Journal of Clinical Child and Adolescent Psychology*, *39*, 410-420. doi:10.1080/15374411003691719

COSTELLO, E.J. (2016). Early detection and prevention of mental health problems: Developmental epidemiology and systems of support. *Journal of Clinical Child and Adolescent Psychology*, *45*, 710-717. doi:10.1080/15374416.2016.1236728

COSTELLO, E.J. & MAUGHAN, B. (2015). Annual research review: Optimal outcomes of child and adolescent mental illness. *Journal of Child Psychology and Psychiatry*, *56*, 324-341. doi:10.1111/jcpp.12371

COSTELLO, E.J., MUSTILLO, S., ERKANLI, A., KEELER, G. & ANGOLD, A. (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of General Psychiatry*, *60*, 837-844. doi:10.1001/archpsyc.60.8.837

CURRY, J., ROHDE, P., SIMONS, A., SILVA, S., VITIELLO, B., KRATOCHVIL, C., … & TADS TEAM (2006). Predictors and moderators of acute outcome in the treatment for adolescents with depression study (TADS). *Journal of the American Academy of Child and Adolescent Psychiatry*, *45*, 1427-1439. doi:10.1097/01.chi.0000240838.78984.e2

de LIJSTER, J.M., DIELEMAN, G.C., UTENS, E.M.W.J., DIERCKX, B., WIERENGA, M., VERHULST, F.C. & LEGERSTEE, J.S. (2018). Social and academic functioning in adolescents with anxiety disorders: A systematic review. *Journal of Affective Disorders*, *230*, 108-117. doi:10.1016/j.jad.2018.01.008

ERSKINE, H.E., MOFFITT, T.E., COPELAND, W.E., COSTELLO, E.J., FERRARI, A.J., PATTON, G., … & SCOTT, J.G. (2015). A heavy burden on young minds: The global burden of mental and substance use disorders in children and youth. *Psychological Medicine*, *45*, 1551-1563. doi:10.1017/s0033291714002888

ESSAU, C.A., CONRADT, J. & PETERMANN, F. (2000). Frequency, comorbidity, and psychosocial impairment of anxiety disorders in German adolescents. *Journal of Anxiety Disorders*, *14*, 263-279. PMID: 10868984

FESTEN, H., HARTMAN, C.A., HOGENDOORN, S., de HAAN, E., PRINS, P.J., REICHART, C.G., … & NAUTA, M.H. (2013). Temperament and parenting predicting anxiety change in cognitive behavioral therapy: The role of mothers, fathers, and children. *Journal of Anxiety Disorders*, *27*, 289-297. doi:10.1016/j.janxdis.2013.03.001

GILLIES, D., TAYLOR, F., GRAY, C., O'BRIEN, L. & D'ABREW, N. (2013). Psychological therapies for the treatment of post-traumatic stress disorder in children and adolescents. *Evidence-Based Child Health*, *8*, 1004-1116. doi:10.1002/ebch.1916

GINSBURG, G.S., KINGERY, J.N., DRAKE, K.L. & GRADOS, M.A. (2008). Predictors of treatment response in pediatric obsessive-compulsive disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *47,* 868-878. doi:10.1097/chi.0b013e3181799ebd

HIGA-McMILLAN, C.K., FRANCIS, S.E., RITH-NAJARIAN, L. & CHORPITA, B.F. (2016). Evidence base update: 50 years of research on treatment for child and adolescent anxiety. *Journal of Clinical Child and Adolescent Psychology*, *45*, 91-113. doi:10.1080/15374416.2015.1046177

HOWARD, K.I., LUEGER, R.J., MALING, M.S. & MARTINOVICH, Z. (1993). A phase model of psychotherapy outcome: Causal mediation of change. *Journal of Consulting and Clinical Psychology*, 61, 678-685. PMID: 8370864

JADHAV, K.S. & BOUTREL, B. (2019). Prefrontal cortex development and emergence of self-regulatory competence: The two cardinal features of adolescence disrupted in context of alcohol abuse. *European Journal of Neuroscience*, *50* (3), 2274-2281. doi:10.1111/ejn.14316

JAYSON, D., WOOD, A., KROLL, L., FRASER, J. & HARRINGTON, R. (1998). Which depressed patients respond to cognitive-behavioral treatment? *Journal of the American Academy of Child & Adolescent Psychiatry*, *37*, 35-39. doi:10.1097/00004583-199801000-00014

KENDALL, P.C., FLANNERY-SCHROEDER, E., PANICHELLI-MINDEL, S.M., SOUTHAM-GEROW, M., HENIN, A. & WARMAN, M. (1997). Therapy for youths with anxiety disorders: A second randomized clinical trial. *Journal of Consulting and Clinical Psychology*, *65*, 366-380. PMID: 9170760

KERNS, C.M., READ, K.L., KLUGMAN, J. & KENDALL, P.C. (2013). Cognitive behavioral therapy for youth with social anxiety: Differential short and long-term treatment outcomes. *Journal of Anxiety Disorders*, *27*, 210-215. doi:10.1016/j.janxdis.2013.01.009

KLEY, H., HEINRICHS, N., BENDER, C. & TUSCHEN-CAFFIER, B. (2012). Predictors of outcome in a cognitive-behavioral group program for children and adolescents with social anxiety disorder. *Journal of Anxiety Disorders*, *26*, 79-87. doi:10.1016/j.janxdis.2011.09.002

LEGERSTEE, J.S., TULEN, J.H.M., KALLEN, V.L., DIELEMAN, G.C., TREFFERS, P.D.A., VERHULST F.C. & UTENS, E.M. (2009). Threat-related selective attention predicts treatment success in childhood anxiety disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, *48*, 196-205. doi:10.1097/chi.0b013e31819176e4

LIBER, J.M., van WIDENFELT, B.M., van der LEEDEN, A.J., GOEDHART, A.W., UTENS, E.M. & TREFFERS, P.D. (2010). The relation of severity and comorbidity to treatment outcome with Cognitive Behavioral Therapy for childhood anxiety disorders. *Journal of Abnormal Child Psychology*, *38*, 683-694. doi:10.1007/s10802-010-9394-1

LINARDON, J., FITZSIMMONS-CRAFT, E.E., BRENNAN, L., BARILLARO, M. & WILFLEY, D.E. (2018). Dropout from interpersonal psychotherapy for mental health disorders: A systematic review and meta-analysis. *Psychotherapy Research*, *13*, 1-12. doi:10.1080/10503307.2018.1497215

LUNDKVIST-HOUNDOUMADI, I., HOUGAARD, E. & THASTUM, M. (2014). Pre-treatment child and family characteristics as predictors of outcome in cognitive behavioural therapy for youth anxiety disorders. *Nordic Journal of Psychiatry*, *68*, 524-535. doi:10.3109/08039488.2014.903295

McGUIRE, J.F., PIACENTINI, J., LEWIN, A.B., BRENNAN, E.A., MURPHY, T.K. & STORCH, E.A. (2015). A meta-analysis of cognitive behavior therapy and medication for child obsessive-compulsive disorder: Moderators of treatment efficacy, response, and remission. *Depression and Anxiety*, *32*, 580-593. doi:10.1002/da.22389

MICHIELIN, P., BERTOLOTTI, G., SANAVIO, E., VIDOTTO, G. & ZOTTI, A.M. (2009). CBA-VE – Un test per valutare l'efficacia nella pratica dei trattamenti psicologici e psicoterapeutici. [CBA-VE – A test to evaluate effectiveness in psychological and psychotherapeutic treatments]. *La professione di psicologo*, *3*, 8-11.

MICHIELIN, P., VIDOTTO, G., ALTOÈ, G., COLOMBARI, M., SARTORI, L., BERTOLOTTI, G., … & ZOTTI, A.M. (2008). Proposta di un nuovo strumento per la verifica dell'efficacia nella pratica dei trattamenti psicologici e psicoterapeutici. [New questionnaire to evaluate psychological treatment effects]. *Giornale Italiano di Medicina del Lavoro ed Ergonomia*, *30*, A98-104. PMID:18700484

NICE GUIDELINE 28 (2017a). *Depression in children and young people: Identification and management*. Available at: https://www.nice.org.uk/guidance/CG28© NICE 2017

NICE GUIDELINE 69 (2017b). *Eating disorders: Recognition and treatment*. Available at: https://www.nice.org.uk/guidance/ng69/resources/eating-disorders-recognition-and-treatment-pdf-1837582159813

NILSEN, T.S., EISEMANN, M. & KVERNMO, S. (2013). Predictors and moderators of outcome in child and adolescent anxiety and depression: A systematic review of psychological treatment studies. *European Child & Adolescent Psychiatry*, *22*, 69-87. doi:10.1007/s00787-012-0316-3

PIACENTINI, J., BERGMAN, R., JACOBS, C., McCRACKEN, J. & KRETCHMAN, J. (2002). Open trial of cognitive behavior therapy for childhood obsessive-compulsive disorder. *Journal of Anxiety Disorders*, *16*, 207-219. PMID: 12194545

POLANCZYK, G.V., SALUM, G.A., SUGAYA, L.S., CAYE, A. &

ROHDE, L.A. (2015). Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of Child Psychology and Psychiatry*, *56*, 345-365. doi:10.1111/jcpp.12381

RAMSAWH, H.J., WEISBERG, R.B., DYCK, I., STOUT, R. & KELLER, M.B. (2011). Age of onset, clinical characteristics, and 15-year course of anxiety disorders in a prospective, longitudinal, observational study. *Journal of Affective Disorders*, *132*, 260-264. doi:10.1016/j.jad.2011.01.006

ROBERTS, R.E., ATTKISSON, C.C. & ROSENBLATT, A. (1998). Prevalence of psychopathology among children and adolescents. *American Journal of Psychiatry*, 155, 715-725. doi:10.1176/ajp.155.6.715

RUDY, B.M., LEWIN, A.B., GEFFKEN, G.R., MURPHY, T.K. & STORCH, E.A. (2014). Predictors of treatment response to intensive cognitive-behavioral therapy for pediatric obsessive-compulsive disorder. *Psychiatry Research*, *220*, 433-440. doi:10.1016/j.psychres.2014.08.002

SHAIN, B. & COMMITTEE ON ADOLESCENCE (2016). Suicide and suicide attempts in adolescents. *Pediatrics*, *138*, e20161420. doi: 10.1542/peds.2016-1420

SIGURVINSDÓTTIR, A.L., JENSÍNUDÓTTIR, K.B., BALDVINSDÓTTIR, K.D., SMÁRASON, O. & SKARPHEDINSSON, G. (2019). Effectiveness of cognitive behavioral therapy (CBT) for child and adolescent anxiety disorders across different CBT modalities and comparisons: A systematic review and meta-analysis. *Nordic Journal of Psychiatry*, *18*, 1-13. doi:10.1080/08039488.2019.1686653.

SMITH, A.M. & JENSEN-DOSS, A. (2017). Youth psychotherapy outcomes in usual care and predictors of outcome group membership. *Psychological Services*, *14*, 66-76. doi:10.1037/ser0000115

SOUTHAM-GEROW, M.A., KENDALL, P.C. & WEERSING, V.R. (2001). Examining outcome variability: Correlates of treatment response in a child and adolescent anxiety clinic. *Journal of Clinical Child Psychology*, *30*, 422-436. doi:10.1207/s15374424jccp3003_13

STILES, W.B., BARKHAM, M. & WHEELER, S. (2015). Duration of psychological therapy: Relation to recovery and improvement rates in UK routine practice. *British Journal of Psychiatry*, *207*, 115-122. doi:10.1192/bjp.bp.114.145565

TICK, N.T., van der ENDE, J. & VERHULST, F.C. (2008). Ten-year trends in self-reported emotional and behavioral problems of Dutch adolescents. *Social Psychiatry and Psychiatric Epidemiology*, *43*, 349-355. doi:10.1007/s00127-008-0315-3

TORP, N.C., DAHL, K., SKARPHEDINSSON, G., COMPTON, S., THOMSEN, P.H., WEIDLE, B., … & IVARSSON, T. (2015). Predictors associated with improved cognitive-behavioral therapy outcome in pediatric obsessive-compulsive disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *54*, 200-207.e1. doi:10.1064/j.aac.2014.12.007

TURNER, C., O'GORMAN, B., NAIR, A. & O'KEARNEY, R. (2018). Moderators and predictors of response to cognitive behaviour therapy for pediatric obsessive-compulsive disorder: A systematic review. *Psychiatry Research*, *261*, 50-60. doi:10.1016/j.psychres.2017.12.034

VIJAYAKUMAR, N., OP de MACKS, Z., SHIRTCLIFF, E.A. & PFEIFER, J.H. (2018). Puberty and the human brain: Insights into adolescent development. *Neuroscience and Biobehavioral Reviews*, *92*, 417-436. doi:10.1016/j.neubiorev.2018.06.004

WEERSING, V.R., JEFFREYS, M., DO, M.T., SCHWARTZ, K.T. & BOLANO, C. (2017). Evidence base update of psychosocial treatments for child and adolescent depression. *Journal of Clinical Child and Adolescent Psychology*, *46*, 11-43. doi:10.1080/15374416.2016.1220310

ZHOU, X., HETRICK, S.E., CUIJPERS, P., QIN, B., BARTH, J., WHITTINGTON, C.J., … & XIE, P. (2015). Comparative efficacy and acceptability of psychotherapies for depression in children and adolescents: A systematic review and network meta-analysis. *World Psychiatry*, *14*, 207-222. doi:10.1002/wps.2021

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.

DOI: 10.26387/bpa.287.3

Boosting human and social capital in the university: The role of seeking challenges on academic performance

Emanuela Ingusci[1], Fulvio Signore[1], Paola Pasca[1], Alessia Catalano[1], Silvia De Masi[1], Giovanni Mattia Gioia[1], Amelia Manuti[2], Monica Molino[3], Paola Spagnoli[4], Claudio Giovanni Cortese[3]

[1] History, Society and Human Studies Department, University of Salento, Lecce
[2] Department of Education, Psychology, Communication, University of Bari
[3] Psychology Department, University of Turin
[4] Department of Psychology, University of Campania "Luigi Vanvitelli", Caserta

claudio.cortese@unito.it

ABSTRACT. Il presente lavoro, inserito nella cornice teorica del modello Job Demands-Resources (JD-R) adattato al contesto universitario, si propone di esplorare in che modo gli studenti gestiscono le richieste accademiche, attraverso la ricerca di sfide, ed il loro impatto sul capitale umano (ad es. meta-competenze) e sociale (ad es. networking), in modo da migliorare la propria performance accademica. Il campione oggetto di indagine è composto da 152 studenti di psicologia che hanno compilato un questionario online composto da domande riguardanti caratteristiche personali ed il percorso accademico. Le analisi sono state condotte ipotizzando quattro variabili latenti (ricerca di sfide, meta-competenze, networking e performance accademica) ed eseguite tramite il PLS-SEM, un approccio non parametrico di equazioni strutturali, mentre il ricampionamento Bootstrap pari a 200 ha consentito di incrementare la validità del modello. I risultati hanno evidenziato una relazione significativa, positiva e diretta tra le variabili ricerca di sfide, meta-competenze e networking degli studenti; nello specifico, la ricerca di sfide influenza in maniera indiretta la performance accademica.

SUMMARY. Drawing and adapting the Job Demands-Resources Model to the academic context, the study aimed at exploring how students manage academic study demands by seeking challenges and adopting their human (e.g. meta-competencies) and social capital (e.g. networking) to improve their academic performance. 152 undergraduate psychology students were asked to fill out an online anonymous questionnaire, whose reliability was assessed. Analyses through four latent variables (seeking challenges, meta-competencies, networking and academic performance) were carried out through PLS-SEM, a non-parametrical approach to Structural Equation Modelling, while Bootstrap validation with n = 200 strengthened model validity. Results showed a direct, positive, significant structural relationship between students' seeking challenges, meta-competencies and networking; moreover, the additional, indirect, presence of human capital arises in the relationship between seeking challenges and academic performance.

Keywords: *Academic crafting, Human and social capital, PLS-SEM*

INTRODUCTION

The relationship between motivational processes and academic performance in college students received growing attention within the last decades (Bailey & Phillips, 2016; Robins, Roberts & Sarris, 2015; Secundo et al., 2019; Signore, Catalano, De Carlo, Madaro & Ingusci, 2019). Universities are a very important educational context for students' personal and professional development, as well as for their future career. During academic experience, students are called to practise their personal resources to become proactive, autonomous and to develop sense of initiative and self-management skills (Geertshuis, Jung & Cooper-Thomas, 2014). By this, their motivation increases leading to positive outcomes, such as academic performance, engagement and retention (Bailey & Phillips, 2016; Ingusci, Palma, De Giuseppe & Iacca, 2016). According to Ryan and Deci (2000), the ambition to seek new challenges, to learn and explore, features students' motivation. When motivated, in fact, students are led to craft and develop more effective learning strategies, to select challenging tasks, to appreciate their classes and are generally more prone to start new projects besides ordinary university assignments (Bailey & Phillips, 2016; Trigueros et al., 2019). Accordingly, university contributes to prepare students to future turbulent ever-changing professional situations: it requires students to engage in achieving academic and professional success, thus becoming capable, well-informed graduates (Ng, Choong, Kuar, Tan & Teoh, 2019).

Several theoretical approaches support the relationship between students' motivation, transversal skills and positive academic outcomes, Self-Determination Theory (SDT) and Social Learning Theory being some of the most relevant ones.

In the organizational as well as in the academic context, these approaches are used as lens to explore and understand both motivational processes and positive outcomes (Bakker, Vergel & Kuntze, 2015; Cilliers, Mostert & Nel, 2018). In view of the above, within the last decades, the Job Demands-Resources (JD-R) model has been traditionally used to describe the balance between job demands and resources, its effects on well-being at work (both positive and negative effects), and its individual and organizational outcomes (Bakker & Demerouti, 2017). In the JD-R model, there are two main elements: job demands and job resources. In particular, job demands are viewed as heterogeneous: some job demands appear to be hindering motivational processes, thus leading to negative outcomes such as burn out and disengagement; on the contrary, other job demands are known as "job challenges": even though they demand individuals to put effort and energy in their tasks in order to achieve goals and satisfy needs, they can generate growth and development opportunities (Van den Broeck, De Cuyper, De Witte & Vansteenkiste, 2010). Therefore, job demands include all aspects that require energy and vigour: they include workload, pressure to hurry, emotionally and cognitively challenging interactions with others, high responsibility, new projects and other challenging demands. Job resources instead are all those physical, psychological, social or organizational work aspects functional to achieve goals and reduce the requests-related psychological costs (e.g. work autonomy, performance-related feedback, social support, supervision, coaching and time management).

The JD-R model allows for a parallel between the academic activities to be carried out and employees work. In other words, students' well-being and efficiency can be thought of as the result of two conditions: study demands (e.g. studying for tests, starting new projects, carrying out training, completing assignments, attending classes, managing the study load) and study resources (meta-competencies, networking, social feedback, relationships with professors). Specifically, social and personal resources (e.g. proactivity and networking) as well as technical and structural ones (e.g. technical skills and knowledge) allow to handle the demands. Taken together, demands and resources trigger two opposite processes related to two opposite individual outcomes: on one hand, emotional exhaustion and burnout, on the other hand engagement and a better job performance.

Key construct of the JD-R model is job crafting, a proactive strategy including all employees' behavioural changes aimed at balancing job demands and resources. Job crafting behaviours foster employees to develop and increase their own skills, to make job true to their own needs and produce positive outcomes (such as person-job fit, job meaningfulness, job satisfaction, work engagement) (Bakker & Demerouti, 2014). As a construct, job crafting is made up by four dimensions one of which, known as "increasing challenging job demands" (that is, situations workers must overcome in order to learn and achieve goals), particularly reflects seeking challenges. Tims, Bakker and Derks (2015) examined the relationship between job

crafting and job performance and showed that both work engagement and job performance can be increased through job crafting. Although the JD-R model has already been adapted to the academic context (Cilliers et al., 2018; Ouweneel, LeBlanc & Schaufeli, 2011), there is still a lack of literature about job crafting behaviours adapted to the educational contexts. Therefore, in line with the JD-R model and the job crafting construct, we proposed to redefine the latter as "academic crafting". Within this theoretical framework, the aim of this research is to explore the role of the academic crafting, that is, seeking challenges in first-level undergraduate psychology students, and its relation with both study resources (human and social capital) and academic performance.

Study demands: the role of seeking challenges

In this study, we adapted the JD-R model, a strong and flexible theoretical model, to the academic context, proposing a distinction between study demands and study resources. Study demands can be defined as the study conditions where effort, energy and engagement are needed to achieve a goal (e.g. study load - or overload - commitment to new projects, time pressure for deadlines, homework and exams) and which need to be balanced by resources. As already mentioned, study demands are not always negative, as they can be divided into hindrance demands and challenging demands (Bakker & Demerouti, 2017; Van den Broeck et al., 2010). When students seek challenges, they are intrinsically and extrinsically motivated to improve their skills for goal achievement and success (Strauser, O'Sullivan & Wong, 2012). Seeking challenges, as a positive study demand, is a form of proactive behaviour that can last in time and provides a good balance between demands and resources (Ingusci, Spagnoli, Zito, Colombo & Cortese, 2019). It refers to all those behaviours (starting new projects, either asking for or accepting more responsibility, etc.) aimed at developing knowledge and skills to manage hard goals and improve satisfaction and motivation, thus positively influencing the actual academic performance (Ingusci et al., 2019).

Often, individuals spontaneously and proactively seek challenging situations that can encourage learning and high motivation levels (Bakker & Demerouti, 2017). In the organizational context, seeking challenges is associated with positive outcomes such as performance, career success, stress tolerance and proactive participation to organizational initiatives. Geertshuis et al. (2014) underlined that proactive behaviour is important for achieving academic success, especially at the end of semester compared to the beginning. Additionally, proactive behaviours have been positively related to learning motivation and the resulting acquisition of new knowledge and skills (Major, Turner & Fletcher, 2006). Tymon and Batistic (2016) found that proactive students are more prone to scan the environment, anticipate possible future problems and engage in behaviours to overcome them. Proactivity, as well as confidence to perform, triggers a self-assessed learning which relies on a mastery approach, thus affecting grades (Geertshuis et al., 2014; Tymon & Batistic, 2016). Based on these premises, we hypothesized that study demands (seeking challenges) could be positively related to personal resources (meta-competencies) (H_1).

Study resources and social resources

Networking can be considered an important social resource. It plays a key role in the educational and professional life of individuals. In fact, developing, maintaining, and increasing relationships can help individuals move towards promising employment opportunities, access necessary information, get useful resources, obtain sponsorship and overall social support. Relationships with others can stimulate new ideas, timely information, career promotion, influence over peers as well as social support (Baker, 2000). According to Lo Presti, Ingusci, Magrin, Manuti and Scrima (2019), networking is a key human capital skill linked to the whole set of attitudes and actual behaviours aimed at improving one's own social capital. Forret and Dougherty (2001) suggested that networking can be considered as the proactive effort to develop and maintain personal and professional relationships useful for mutual improvement. Factor analysis helped them to identify five types of networking behaviours: 1) maintaining contacts; 2) socializing; 3) engaging in professional activities; 4) participating in community activities; 5) increasing internal visibility. These factors shape networking as a proactive skill that leads to a valuable and resourceful career. Networking measures typically assess how often individuals show networking behaviours (e.g. using contacts to get confidential advice or taking the opportunity to meet new people). In the

academic setting, networking became paramount as it shapes and reinforces students' interpersonal relationships, as well as their educational and career outcomes (Havnes, 2008; Wolff & Moser, 2010). Networking skills are beneficial in terms of students' adjustment (Buote et al., 2007), academic attainment (Havnes, 2008), perceived psychological wellbeing and physical health (Cassidy, 2004). Recently, scholars called on for more initiatives to encourage students to develop networking skills (Friar & Eddleston, 2007) and for educators to offer appropriate training (Villar & Albertin, 2010). Although networks development may not always be easy, social capital is a crucial advantage for individuals, whatever the field of application is. Building a contact's chain translates into a strategic advantage. In view of the above, the study argued that study demands (seeking challenges) could be positively related to social resources (networking) (H_2).

Personal resources

According to the JD-R model, resources include personal, social and job conditions useful for the individual to balance and manage the demands. Also, meta-competencies could be conceptualised as personal resources. The term was first used in the educational sciences to denote individual's ability to adapt to environmental changes by transforming his/her own models of knowledge and action (Brown & McCartney, 1995). Differently from technical skills, which are task and context-specific, meta-competencies are non-specific (Brown, 1994). Therefore, they represent a higher order soft skill concerning the ability to learn, anticipate and create (Brown & McCartney, 1995). Soft skills can be considered as socio-emotional abilities and represent a crucial strategy to promote personal development, social participation, academic and work achievement (Ricchiardi & Emanuel, 2018). Within this framework, meta-competencies encompass cognitive skills (e.g. analytical thinking, the use of abstract concepts, technical expertise, pattern recognition), organizational and interpersonal skills (e.g. interpersonal sensitivity, building relationships, organizational awareness, understanding power relationship within the organization), management skills (e.g. group management, guiding others, using unilateral authority) and personal skills (e.g. initiative, proactivity, determination, flexibility, self-confidence and self-control) (Brown, 1994). On this basis, the third hypothesis of the study was that personal resources (meta-competencies) could be positively related to academic performance (H_3).

HYPOTHESES AND AIMS

Moving from the theoretical background drawn above, the study aimed at exploring the relationship between seeking challenges (as a form of academic demand), meta-competencies, networking (as forms of personal and social academic resources, respectively) and academic performance in a group of psychology undergraduate students.

More specifically, the study assumed that:

H_1: seeking challenges would show a positive relationship with meta-competencies;

H_2: seeking challenges would show a positive relationship with networking;

H_3: meta-competencies would positively relate to academic performance.

METHODS

Participants

Participants, 152 psychology undergraduate students from an Italian university, were selected through a convenience sampling. Most of them were females (81.6%), while only 18.4% were males. The mean age was 22 years, while the mode and the median were 21 years. Globally, age ranged from 19 to 51 years. Participants had an average of 1 failed exam.

Procedure and measures

Participants were asked to fill in an online questionnaire. Anonymity was guaranteed according to the General Data Protection Regulation (EU) 2016/679 (GDPR). The questionnaires assessed the following psychological constructs:

– *Seeking challenges*. Seeking challenges are defined as behaviours that increase motivation, promote autonomy, and facilitate learning. This variable was measured through 3 items adapted from the *Job Crafting Scale* ($\omega_{McDonald}$ = .68; Petrou, Demerouti, Peeters, Schaufeli & Hetland,

2012). Participants were asked to express their agreement/disagreement with each item using a 6-point scale (from 1 = *completely disagree*, to 6 = *completely agree*). An example of the items is "In University, when an interesting project comes along, I offer myself proactively to participate in" ($CHAL_1$).

– *Networking*. This variable identifies a set of behaviours addressed to use relationships inside the academic context as a precious capital to achieve goals. It was assessed by 3 items selected from the *Social Capital and Networking* subdimension of *The Human Capital and Professional Development Scale* ($\omega_{McDonald}$ = .79; Lo Presti et al., 2019). Participants assessed the personal occurrence of each behaviour using a 6-point scale (from 1 = *never*, to 6 = *always*). Example of item is "In University, meeting new people is an opportunity that I rarely miss" (NET_2).
– *Meta-competencies*. This variable, connected to the whole of transversal/soft skills complementing students' personal resources, was assessed through 3 items selected from the *Human Capital and Professional Development* subdimension of *The Human Capital and Professional Development Scale* ($\omega_{McDonald}$ = .76 - Lo Presti et al., 2019). Participants were invited to express their agreement/disagreement with each item using a 6-point scale (from 1 = *completely disagree*, to 6 = *completely agree*). Example of item is "Developing new knowledge about my university career is easy to me" ($META_2$).
– *Academic performance*: this variable was used as an objective measure of students' academic performance. It was assessed by asking participant to declare the average grade of their exams. According to University Italian evaluation system, the scale range for passing an exam is from a minimum of 18 to a maximum of 30 with laude, which we considered as a 31 in our analysis. All the items were adapted to the academic context.

Data analysis

Analyses were performed through Jamovi, version 1.2.2.0, and R Studio software (Version 1.2.5033), using the Plspm package (Sanchez, 2013).

Since missing values represented 1.31% of the dataset, they were imputed to the median. We performed KMO measure of sample adequacy and Bartlett's test of sphericity for each dimension, then we assessed variables skewness and kurtosis. Research hypotheses were investigated via PLS-SEM modelling, where four latent variables were measured by their reflective indicators (Cheah et al., 2019). Reliability analysis based on Cronbach's alpha and McDonald's Omega allowed to assess the internal coherence of latent variables, while discriminant validity was assessed by examining the cross-loadings. Bootstrap validation (n = 200) was carried out on both the inner and the outer model loadings.

RESULTS

KMO test showed a mediocre adequacy for all factors (CHAL = .61, META = .62, NET = .63) while Bartlett's test was statistically significant confirming items appropriateness to be factorized (Cerny & Kaiser, 1977). Table 1 shows the principal descriptive statistics of aggregated measures (item means).

For what concerns skewness and kurtosis, values ranging between –1 and +1 indicate a negligible deviation from normal distribution (Hair, Sarstedt, Ringle & Gudergan, 2017). Among the constructs, only networking shows a consistent non-normality. However, due to the sampling strategy (convenience sampling), sample size (152 individuals) and the Shapiro-Wilk test, further and mostly confirming non-normality, the non-parametrical approach was preferred. The latent dimensions included seeking challenges (academic crafting), meta-competencies, networking, and academic performance. Table 2 shows the indicators for each of the latent variables.

Table 2 reports Cronbach's alpha and McDonald's Omega coefficients. As academic performance was a single-indicator variable, it was not included as a latent variable. Networking and meta-competencies showed a good reliability (values higher than .70 are considered acceptable), while seeking challenges showed a lower reliability. More specifically, the item-scale correlation confirmed $CHAL_3$ to be a problematic item. In fact, without $CHAL_3$ reliability increased to .71 (Cronbach) and .71 (McDonald). As said before, increasing challenging demands is a subdimension of job crafting, a construct never adapted before to university students. In the academic context, results showed that removing the item $CHAL_3$ from the scale would be preferable. For this reason, only $CHAL_1$ and $CHAL_2$ were retained for further analysis.

The overall model considered seeking challenges as an academic crafting behaviour (more specifically, as a job

Table 1 – Descriptive statistics of variables of the questionnaire

	PERF	CHAL	META	NET
Mean	25.70	3.01	4.42	4.01
Standard deviation	2.03	1.03	.83	.79
Minimum	20.0	1.00	1.00	1.00
Maximum	31.0	6.00	6.00	5.00
Skewness	−.11	.34	−.64	−.90
Std. error skewness	.20	.20	.20	.20
Kurtosis	−.20	.02	.41	1.23
Std. error Kurtosis	.39	.39	.39	.39
Shapiro-Wilk p	.094	.020	<.001	<.001

Legenda. PERF = Academic performance; CHAL = Seeking challenges; META = Meta-competencies; NET = Networking.

Table 2 – Latent variables, number of indicators and reliability measures of the first explorative model

Latent variables	Indicators	Label	Cronbach's alpha	McDonald's omega
Seeking challenges	3	CHAL	.65	.68
Networking	3	NET	.76	.78
Meta-competencies	3	META	.73	.76

crafting dimension named increasing challenging demands), adapted it to the university context, and investigated its role in implementing some transversal skills, such as meta-competencies and networking. The second hypothesis tested the relationship between meta-competencies and academic performance measured by the average grade, as represented in Figure 1.

With an explorative aim, we first ran a PLS-SEM model including all indicators for each of the latent variables. According to the rule of thumb presented in Hair et al. (2017), loadings lower than .70 reveal poor indicators of the latent variable, which in turn explain less than the 50% of items variance (also called communality). In this first model, we found that $CHAL_3$ (.60) and NET1 (.57) did not reach that cut-off, so we deleted them from the measurement structure and then ran our final model (see later Figure 2).

Figure 1 – First explorative overall PLS-SEM model with outer and inner structure

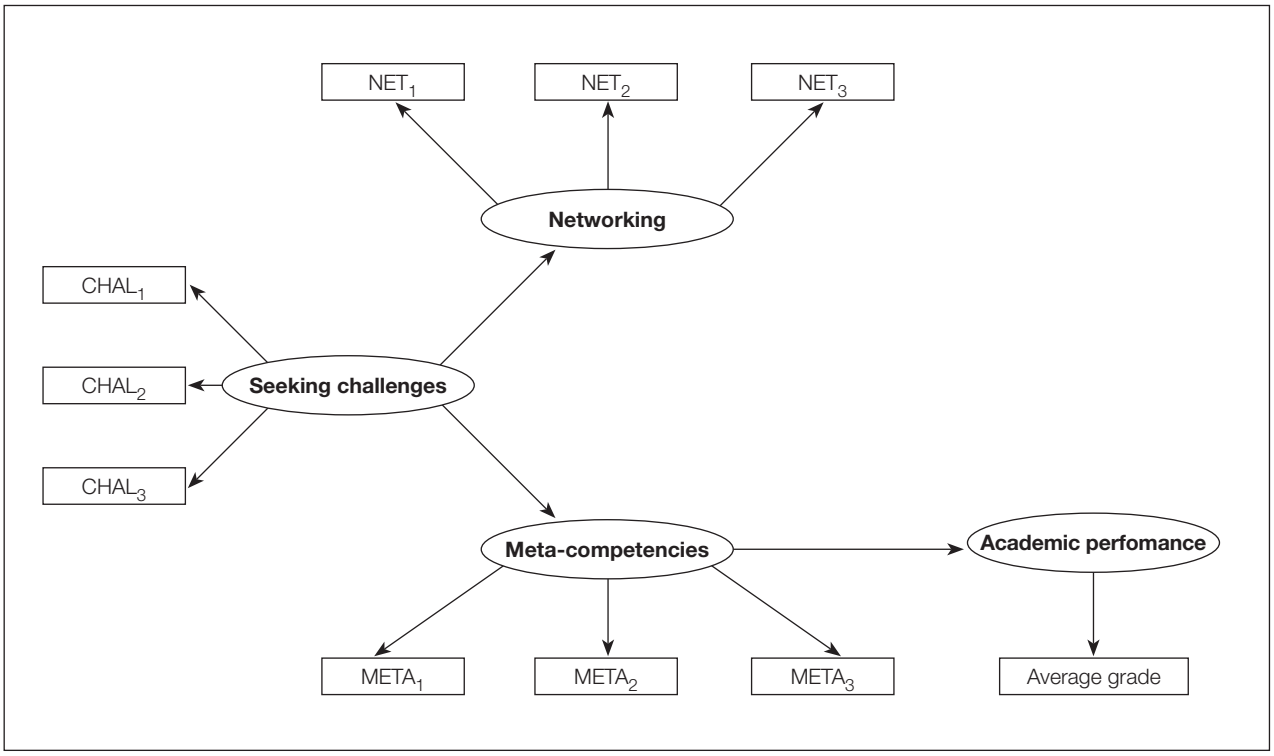
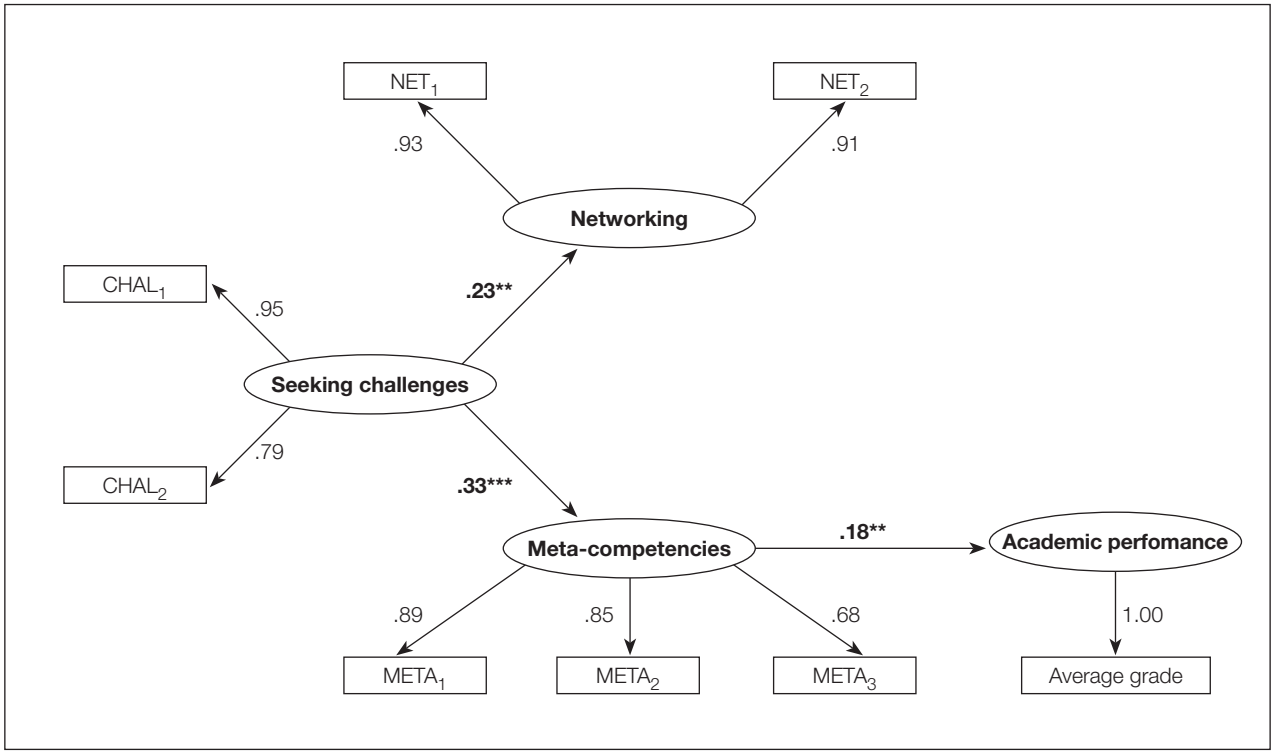


Figure 2 – Final model with path coefficients and loadings



Cronbach's alpha and Dillon-Goldstein's rho coefficients allowed to assess reliability of new latent variables. As academic performance was measured by only one indicator, it showed the highest value for reliability (1.00). All latent variables showed Cronbach's alpha and Dillon-Goldstein's rho higher than .70 (see Table 3). Also, as traditionally recommended, the first eigenvalues are larger than 1, while the second are lower. Therefore, we confirm the good internal reliability of our constructs.

According to PLS-SEM modelling guidelines about convergent validity, items should display loadings higher than .70 in order to be considered acceptable indicators. Loadings >.70 correspond to more than 50% of variance of each item explained by the latent variable, also called communality. As shown in Table 4, loadings and communalities in our model mostly met these criteria: even though $META_3$ showed a slightly lower loading, we still included it in further analysis, due to its extreme proximity to the threshold, as well as the explorative aim of our study (Hair et al., 2017). Finally, we calculated Average Variance Extracted (AVE), a measure of global convergent validity indicating how much variance of the latent variable is explained by its indicators, in contrast to variance due to measurement error. AVE greater than .50 is considered acceptable. Our analysis showed that $AVE_{CHAL} = .76$, $AVE_{META} = .66$, $AVE_{MEDIA} = 1.0$ and $AVE_{NET} = .85$ and confirmed that measurement error is not problematic.

Table 3 – Reliability measures for latent variables of the final model

Latent variables	MVs	Cronbach's alpha	Dillon-Goldstein ρ	1st eigenvalue	2nd eigenvalue
Seeking challenges	2	.71	.87	1.56	.44
Meta-competencies	3	.74	.85	1.98	.71
Networking	2	.82	.92	1.70	.30

Legenda. MVs = Manifest Variables.

Table 4 – Final outer model with loadings and communalities

Indicators	Latent variables	Loadings	Communality
$CHAL_1$	Seeking challenges	.95	.90
$CHAL_2$	Seeking challenges	.79	.63
$META_1$	Meta-competencies	.89	.80
$META_2$	Meta-competencies	.85	.72
$META_3$	Meta-competencies	.68	.46
NET_1	Networking	.93	.86
NET_2	Networking	.91	.83

To assess discriminant validity, we examined the cross-loadings. As shown in Table 5 (in bold), the highest correlations of our manifest variables were with their respective latent constructs. Discriminant validity was thus confirmed. Final model is highlighted in Figure 2.

All the coefficient estimates were statistically significant. Seeking challenges, made up of two indicators of *Job Crafting Scale*, affected both human (meta-competencies) and social capital (networking). Seeking challenges positively influenced the development of meta-competencies (β_2 = .33, p = .000), as assumed in H_1, and networking (β_1 = .23, p = .004), as supposed in H_2. At the same time, the structural relationship between meta-competencies and performance resulted to be positive and significant (β_3 = .18, p = .025), confirming the hypothesis proposed in H_3. In other words, seeking challenges had a significant, indirect influence on performance (β_4 = .06). To assess the overall validity of the model and estimates accuracy, we performed non-parametric Bootstrap, with which the original dataset was resampled 200 times, allowing to get mean estimates of the PLS-SEM model. Results are displayed in Table 6 and Table 7.

Table 5 – Discriminant analysis: cross-loadings of the manifest variables

	CHAL	META	PERF	NET
$CHAL_1$	**.95**	.33	.13	.26
$CHAL_2$	**.79**	.21	.09	.10
$META_1$	.31	**.89**	.20	.09
$META_2$	.27	**.85**	.09	.09
$META_3$	.21	**.68**	.13	.10
$PERF_1$	.13	.18	**1.00**	–.07
NET_1	.22	.14	–.04	**.93**
NET_2	.20	.07	–.09	**.91**

Legenda. CHAL = Seeking challenges; META = Meta-competencies; PERF = Academic performance; NET = Networking.

Table 6 – Bootstrap validation (n = 200) of final inner model path coefficients

Relation	Original	Mean Bootstrap	Std error	Lower ci	Upper ci
CHAL → META	.33	.34	.07	.20	.47
CHAL → NET	.23	.24	.10	.08	.42
META → PERF	.18	.19	.08	.05	.33

Legenda. CHAL = Seeking challenges; META = Meta-competencies; NET = Networking; PERF = Academic performance.

Table 7 – Bootstrap validation (n = 200) of loadings between indicators and latent variables (final outer model)

	Original	Mean Bootstrap	Std error	Lower CI	Upper CI
$CHAL_1$-Seeking challenges	.95	.94	.03	.91	.99
$CHAL_2$-Seeking challenges	.79	.78	.09	.63	.90
$META_1$-Meta-competencies	.89	.89	.09	.81	.95
$META_2$-Meta-competencies	.85	.84	.06	.68	.93
$META_3$-Meta-competencies	.68	.68	.10	.47	.83
NET_2-Networking	.93	.92	.04	.84	.98
NET_3-Networking	.91	.90	.04	.83	.96

As showed in Table 6 and 7, standard errors of path coefficients and loadings are very low; moreover, confidence intervals did not contain the 0 value, thus confirming the significant nature of the statistical relationships observed.

DISCUSSION AND CONCLUSIONS

Within the current scenario featured by the need to enhance academic retention and career success, universities represent the physical and psychological environments that could provide students with opportunities to develop and enhance their human capital, that is, their soft skills (whose meta-competencies are a fundamental part), networking and proactive behaviours, that are important resources for their future career. Nowadays, a challenging task for universities is to provide students with competitive knowledge and with skills and abilities useful to manage their personal resources in line with the demands of a complex and mutable labour market (Ricchiardi & Emanuel, 2018). Therefore, as showed by prior research in the field (Robins et al., 2015), higher levels of personal and social resources could foster students' motivation toward academic goals, positively influencing their satisfaction, engagement and academic performance. In line with these results, Chamorro-Premuzic and colleagues (Chamorro-Premuzic, Arteche, Bremner, Greven & Furnham, 2010) reported that soft skills significantly correlated with students' commitment toward the goal of the degree, with their engagement in their studies and finally with their academic performance. Therefore, training programs and interventions addressed to the improvement of soft skills within the academic context could be strategic actions both for university policies in order to retain students and to prevent dropout and for students' future professional expertise development based on learning acquired in formal and informal contexts (Ricchiardi & Emanuel, 2018).

During university, students are often called to manage their personal and social resources to improve learning behaviours and performance (Bakker et al., 2015) and to buffer academic demands that could be also stressful and difficult to be managed. While hindrance demands have been widely explored, little attention has been given to challenging demands (Van den Broeck et al., 2010). Moving from the JD-R theoretical model, the present study aimed at investigating the positive side of academic demands by examining seeking challenges, a sub-dimension of the job crafting construct traditionally explored in the organizational context. The study extended the investigation of this variable to the university context by considering students as crafters of their

own skills and needs, in order to properly respond to the environmental demands.

Results showed that crafting one's own academic job (e.g. modifying, developing and reinforcing one's own personal and social resources) can positively influence academic performance by enhancing skills useful to overcome academic challenges, to make students employable and able to build up their own career path (Secundo et al., 2019; Signore et al., 2019).

This study showed that also a study demand, such as seeking challenges, can be a strategic way to reinforce resources belonging to human and social capital (e.g. meta-competencies and networking). In addition, seeking challenges indirectly positively and significantly affected one's academic performance.

However, this study is not without limits: first, the study used a cross-sectional design; in the future, longitudinal studies can be used to explore causal relationships among these variables across time. Secondly, due to the use of only self-report data, the risk of common method-bias must be considered. Furthermore, future research in the filed should consider also the contribution of more objective and structured index (number of exams, student's timeline, etc.) related to academic performance. The third limitation was the convenience sampling procedure adopted by the present study. Future studies could consider this aspect and complement the analysis by involving students enrolled in different scientific paths. Another limitation was related to the sample size: although PLS-SEM does not imply a minimum sample size, the exploration and prediction features of this model increase with a higher number of individuals. Despite the relationships considered in the present study, the above-mentioned limitations undermine any form of generalizability. Therefore, it is important to consider results cautiously, taking them as a stimulus to improve the research design and avoiding any broader inference.

Overall, the study paved the way to future research in the context of academic crafting, showing new intervention strategies to support students in their academic life. Universities and career services could develop retention programs, through networks with stakeholders, as well as potential employers and companies. Preliminary results coming from the study indicated that universities might boost academic crafting behaviours by developing interventions focused on personal and social resources, hindrances and challenging demands management, training students to exercise job crafting thorough dedicated interventions that might have a positive influence on their future working environment (Ingusci et al., 2018; Ingusci et al., 2019; van Wingerden, Derks & Bakker, 2017; Zito et al., 2019).

There is currently no extensive research on the role of academic crafting on university students' outcomes. Within this conceptual framework, this study represents a starting point in exploring the role and importance of proactive behaviours during academic career. Job crafting interventions have been mostly studied in the organizational field (Van Wingerden et al., 2017; Van Wingerden, Bakker & Derks, 2017a). However, drawing from the theoretical framework of the JD-R model, results coming from the present study contributed to pave the way to future interventions on the empowerment of active behaviours within the academic context. At its initial stage, a training process might provide subjects with information on how demands and resources can be connected to motivational and well-being processes, as well as on how to act on them in order to promote favourable outcomes (Van Wingerden, Bakker & Derks, 2017b). Crafting behaviours make possible to manage and balance demands and resources within each context. Through an intervention, participants might be supported in a reflection and mapping process of their competences, tasks, demands and resources. As a final step, such reflection about the overall framework might help in identifying how crafting behaviours impact on demands and improve well-being (Van Den Heuvel, Demerouti & Peeters, 2015). Crafting behaviours are defined as spontaneous and unaware processes in which individuals redesign their own tasks by altering the boundaries of their activities (Ingusci et al., 2018). Based on direct and indirect relationships that we found in our study, crafting interventions could encourage autonomy and foster a positive management of academic activities.

References

BAILEY, T.H. & PHILLIPS, L.J. (2016). The influence of motivation and adaptation on students' subjective well-being, meaning in life and academic performance. *Higher Education Research & Development*, *35* (2), 201-216.

BAKER, W.E. (2000). *Achieving success through social capital: Tapping the hidden resources in your personal and business networks*. San Francisco: Jossey-Bass.

BAKKER, A.B. & DEMEROUTI, E. (2014). Burnout and work engagement: The JD-R approach. *Annual Review of Organizational Psychology and Organizational Behavior*, *1*, 389-411.

BAKKER, A.B. & DEMEROUTI, E. (2017). Job demands-resources theory: Taking stock and looking forward. *Journal of Occupational Health Psychology*, *22* (3), 273-285.

BAKKER, A.B., VERGEL, A.I.S. & KUNTZE, J. (2015). Student engagement and performance: A weekly diary study on the role of openness. *Motivation and Emotion*, *39* (1), 49-62.

BROWN, R.B. (1994). Refrain the competency debate: Management knowledge and meta-competence in graduate education. *Management Learning*, *25* (2), 289-299.

BROWN, R.B. & McCARTNEY, S. (1995). Competence is not enough: Meta-competence and accounting education. *Accounting Education*, *4* (1), 43-53.

BUOTE, V.M., PANCER, S.M., PRATT, M.W., ADAMS, G., BIRNIE-LEFCOVITCH, S., POLIVY, J. & WINTRE, M.G. (2007). The importance of friends: Friendship and adjustment among 1st-year university students. *Journal of Adolescent Research*, *22* (6), 665-689.

CASSIDY, T. (2004). Mapping variables related to social identity, distress and perceived health in an undergraduate student population. *Social Psychology of Education*, *7* (3), 339-352.

CERNY, B.A. & KAISER, H.F. (1977). A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivariate Behavioral Research*, *12* (1), 43-47.

CHAMORRO-PREMUZIC, T., ARTECHE, A., BREMNER, A.J., GREVEN, C. & FURNHAM, A. (2010). Soft skills in higher education: Importance and improvement ratings as a function of individual differences and academic performance. *Educational Psychology*, *30* (2), 221-241.

CHEAH, J.H., TING, H., RAMAYAH, T., MEMON, M.A., CHAM, T.H. & CIAVOLINO, E. (2019). A comparison of five reflective-formative estimation approaches: Reconsideration and recommendations for tourism research. *Quality & Quantity*, *53* (3), 1421-1458.

CILLIERS, J.R., MOSTERT, K. & NEL, J.A. (2018). Study demands, study resources and the role of personality characteristics in predicting the engagement of first-year university students. *South African Journal of Higher Education*, *32* (1), 49-70.

FORRET, M.L. & DOUGHERTY, T.W. (2001). Correlates of networking behavior for managerial and professional employees. *Group & Organization Management*, *26* (3), 283-311.

FRIAR, J.H. & EDDLESTON, K.A. (2007). Making connections for success: A networking exercise. *Journal of Management Education*, *31* (1), 104-127.

GEERTSHUIS, S., JUNG, M. & COOPER-THOMAS, H. (2014). Preparing students for higher education: The role of proactivity. *International Journal of Teaching and Learning in Higher Education*, *26* (2), 157-169.

HAIR, J.F., SARSTEDT, M., RINGLE, C.M. & GUDERGAN, S.P. (2017). *Advanced issues in partial least squares structural equation modeling.* SaGe publications.

HAVNES, A. (2008). Peer-mediated learning beyond the curriculum. *Studies in Higher Education*, *33* (2), 193-204.

INGUSCI, E., PALMA, F., DE GIUSEPPE, M.C. & IACCA, C. (2016). Social and scholar integration and students satisfaction: The mediating role of career adaptability. *Electronic Journal of Applied Statistical Analysis*, *9* (4), 704-715.

INGUSCI, E., CALLEA, A., CORTESE, C.G., ZITO, M., BORGOGNI, L., CENCIOTTI, R., SIGNORE, F., CIAVOLINO, E. & DEMEROUTI, E. (2019). Self-efficacy and work performance: The role of job crafting in middle-age workers. *International Journal of Business and Society*, *20* (2) 533-551.

INGUSCI, E., CORTESE, C.G., ZITO, M., COLOMBO, L., COLETTA, G., TORDERA, N., SIGNORE, F. & CIAVOLINO, E. (2018). An Italian validation of the job crafting scale: A short form. *Statistica & Applicazioni*, *16* (2), 159-179.

INGUSCI, E., SPAGNOLI, P., ZITO, M., COLOMBO, L. & CORTESE, C.G. (2019). Seeking challenges, individual adaptability and career growth in the relationship between workload and contextual performance: A two-wave study. *Sustainability*, *11* (2), 422.

LO PRESTI, A., INGUSCI, E., MAGRIN, M.E., MANUTI, A. & SCRIMA, F. (2019). Employability as a compass for career success: Development and initial validation of a new multidimensional measure. *International Journal of Training and Development*, *23* (4), 253-275.

MAJOR, D.A., TURNER, J.E. & FLETCHER, T.D. (2006). Linking proactive personality and the Big Five to motivation to learn and development activity. *Journal of Applied Psychology*, *91* (4), 927-935.

NG, L.P., CHOONG, Y.O., KUAR, L.S., TAN, C.E. & TEOH, S.Y. (2019). The effects of psychological capital and proactive personality on undergraduate students'academic performance. In *Proceedings of the Second International Conference on the Future of ASEAN (ICoFA) 2017- Volume 1*. Singapore: Springer.

OUWENEEL, E., LE BLANC, P.M. & SCHAUFELI, W.B. (2011). Flourishing students: A longitudinal study on positive emotions, personal resources, and study engagement. *The Journal of Positive Psychology*, *6* (2), 142-153.

PETROU, P., DEMEROUTI, E., PEETERS, M.C., SCHAUFELI, W.B. & HETLAND, J. (2012). Crafting a job on a daily basis: Contextual correlates and the link to work engagement. *Journal of Organizational Behavior*, *33* (8), 1120-1141.

RICCHIARDI, P. & EMANUEL, F. (2018). Soft skill assessment in higher education. *Journal of Educational, Cultural and Psychological Studies (ECPS Journal)*, *18*, 21-53.

ROBINS, T.G., ROBERTS, R.M. & SARRIS, A. (2015). Burnout and engagement in health profession students: The relationships between study demands, study resources and personal resources. *The Australian Journal of Organisational Psychology*, *8*.

RYAN, R.M. & DECI, E.L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25* (1), 54-67.

SANCHEZ, G. (2013). *PLS path modeling with R*. Berkeley: Trowchez Editions.

SECUNDO, G., DE CARLO, E., MADARO, A., MARUCCIO, G., SIGNORE, F. & INGUSCI, E. (2019). The impact of career insight in the relation with social networks and career self-management: Preliminary evidences from the Italian contamination lab. *Sustainability*, *11* (21), 5996.

SIGNORE, F., CATALANO, A., DE CARLO, E., MADARO, A. & INGUSCI, E. (2019). The role of employability in students during academic experience: A preliminary study through PLS-PM technique. *Electronic Journal of Applied Statistical Analysis*, *12* (4), 720-747.

STRAUSER, D.R., O'SULLIVAN, D. & WONG, A.W. (2012). Work personality, work engagement, and academic effort in a group of college students. *Journal of Employment Counseling*, *49* (2), 50-61.

TIMS, M., BAKKER, A.B. & DERKS, D. (2015). Job crafting and job performance: A longitudinal study. *European Journal of Work and Organizational Psychology*, *24* (6), 914-928.

TRIGUEROS, R., AGUILAR-PARRA, J.M., CANGAS, A.J., BERMEJO, R., FERRANDIZ, C. & LÓPEZ-LIRIA, R. (2019). Influence of emotional intelligence, motivation and resilience on academic performance and the adoption of healthy lifestyle habits among adolescents. *International Journal of Environmental Research and Public Health*, *16* (16), 2810.

TYMON, A. & BATISTIC, S. (2016). Improved academic performance and enhanced employability? The potential double benefit of proactivity for business graduates. *Teaching in Higher Education*, *21* (8), 915-932.

VAN DEN BROECK, A., DE CUYPER, N., DE WITTE, H. & VANSTEENKISTE, M. (2010). Not all job demands are equal: Differentiating job hindrances and job challenges in the job demands-resources model. *European Journal of Work and Organizational Psychology*, *19* (6), 735-759.

VAN DEN HEUVEL, M., DEMEROUTI, E. & PEETERS, M.C.W. (2015). The job crafting intervention: Effects on job resources, self-efficacy, and affective well-being. *Journal of Occupational and Organizational Psychology*, *88* (3), 511-532.

VAN WINGERDEN, J., BAKKER, A.B. & DERKS, D. (2017a). Fostering employee well-being via a job crafting intervention. *Journal of Vocational Behavior*, *100*, 164-174.

VAN WINGERDEN, J., BAKKER, A.B. & DERKS, D. (2017b). The longitudinal impact of a job crafting intervention. *European Journal of Work and Organizational Psychology*, *26* (1), 107-119.

VAN WINGERDEN, J., DERKS, D. & BAKKER, A.B. (2017). The impact of personal resources and job crafting interventions on work engagement and performance. *Human Resource Management*, *56* (1), 51-67.

VILLAR, E. & ALBERTÍN, P. (2010). 'It is who knows you'. The positions of university students regarding intentional investment in social capital. *Studies in Higher Education*, *35* (2), 137-154.

WOLFF, H.G. & MOSER, K. (2010). Do specific types of networking predict specific mobility outcomes? A two-year prospective study. *Journal of Vocational Behavior*, *77* (2), 238-245.

ZITO, M., COLOMBO, L., BORGOGNI, L., CALLEA, A., CENCIOTTI, R., INGUSCI, E. & CORTESE, C.G. (2019). The nature of job crafting: Positive and negative relations with job satisfaction and work-family conflict. *International Journal of Environmental Research and Public Health*, *16* (7), 1176.

DOI: 10.26387/bpa.287.4

Intelligence profiles of children and adolescents with High functioning autism spectrum disorder

Riccardo Alessandrelli[1], Claudia Di Bucchianico[1], Valeria Mancini[1], Dominga Marfisi[1], Tatiana Bortolatto[1], Candida Marchione[1], Luana Pitturelli[1], Maria Elena Di Bucchianico[1], Antonietta Vassalli[1], Morena Farese[2], James Dawe[2], Lina Pezzuti[2]

[1] Department of Child and Adolescent Psychiatry, ASL2 Abruzzo Lanciano-Vasto-Chieti
[2] Department of Dynamic and Clinical Psychology, Faculty of Medicine and Psychology, Sapienza University of Rome

lina.pezzuti@uniroma1.it

ABSTRACT. Nel presente articolo identifichiamo un endofenotipo per soggetti con disturbo dello spettro autistico – livello 1 (Autism spectrum disorder, ASD-1) e normale funzionamento cognitivo, utilizzando la *Wechsler Intelligence Scale for Children – Fourth Edition* con un campione clinico di 80 bambini con diagnosi di ASD-1 senza disabilità intellettiva (con QI>70), e un gruppo di controllo di confronto (*n* = 80) appaiato per età, genere dei bambini e livello di istruzione dei genitori. Dai risultati è emerso che il gruppo clinico con alto funzionamento (High functioning autism spectrum disorder - level 1, HFASD-1) ha ottenuto risultati inferiori rispetto al gruppo di controllo appaiato all'Indice di Velocità di elaborazione e all'Indice della Memoria di lavoro, evidenziando la sensibilità di queste misure sul deterioramento cognitivo generalizzato. Questo risultato è confermato anche dall'assenza di una differenza tra il gruppo HFASD-1 e quello di controllo all'Indice di Abilità generale e dalla grande differenza all'Indice di Competenza cognitiva a favore del gruppo di controllo. Inoltre, il 36% dei bambini HFASD-1 manifestava una grande e rara differenza tra i 4 indici e quindi il QI totale poteva essere considerato non interpretabile come abilità unitaria e coesa. Possiamo sostenere che il profilo cognitivo del HFASD-1 non possa essere interpretato come un'entità unitaria rappresentata semplicemente dal QI, ma si evince che è possibile ottenere una migliore valutazione del loro livello cognitivo utilizzando separatamente l'Indice di Abilità generale e l'Indice di Competenza cognitiva.

SUMMARY. In this paper we identify an endophenotype for individuals with Autism spectrum disorder – level 1 (ASD-1) and normal cognitive functioning using the Wechsler Intelligence Scale for Children – Fourth Edition with a clinical sample of 80 diagnosed ASD-1 children without intellectual disability (with FSIQ>70), and a comparison matched-paired control group (n = 80) combined for age, gender of children and parents' level education. From results emerged that the clinical ASD-1 with High functioning group (ASD-1 HF) performed worse than the matched-paired control group on Processing Speed Index and Working Memory Index, reflecting the sensitivity of these measures to generalized cognitive impairment. This result is also confirmed by the absence of a difference between the ASD-1 HF and control groups in the General Ability Index and the large difference to the Cognitive Proficiency Index in favour of the control group. Again, 36% of ASD children had a rare and large difference between the 4 indices and then the FSIQ could be deemed uninterpretable as unitary and cohesive ability. We argue that the ASD-1's cognitive profile cannot be interpreted as a unitary entity represented from simply FSIQ, but we can obtain a better assessment of cognitive level in ASD subjects using separately the General Ability Index and the Cognitive Proficiency Index.

Keywords: *Autism spectrum disorder, Intelligence, WISC-IV, Cognitive profile, Full Scale Intelligent Quotient*

INTRODUCTION

In the *Diagnostic and Statistical Manual of Mental Disorders – Fifth Edition* (*DSM-V*; American Psychiatric Association, 2013), the autism, Asperger disorder and pervasive developmental disorder not otherwise specified, have been collapsed into a single disorder, the Autism spectrum disorder (ASD) (Kaufmann, 2012). Language abnormalities, repetitive/restricted behaviors and social impairment are the triad of characteristics shared by children with ASD (Zayat, Kalb & Wodka, 2011). As research has often also highlighted intellectual deficits, its assessment in children with ASD is of fundamental importance. In fact, the DSM-5 requires to specify whether ASD is associated with an intellectual disability (American Psychiatric Association, 2013).

Although the new DSM-V classification has unified Asperger syndrome (AS) and High functioning autism (HFA), some studies have suggested that persons with AS possess a distinct profile on tests of intelligence characterized by a high verbal IQ and a low performance IQ, whereas in persons with HFA, the pattern is often reversed (i.e., Ghaziuddin & Mountain-Kimchi, 2004; Mouga et al., 2016).

One of the most commonly used intelligence tests for children is the *Wechsler Intelligence Scale for Children – Fourth Edition* (*WISC-IV*; Wechsler, 2003), and we argue it is a helpful tool that better differentiates and eliminates confounding factors at play in the debate outlined in literature and overcoming the dichotomy of the verbal and performance. The WISC-IV, in addition to Full Scale Intellectual Quotient (FSIQ), implies a the four-factors solution (four indices), i.e. the Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI), Working Memory Index (WMI) and Processing Speed Index (PSI) and two additional indices, i.e. the General Ability Index (GAI) and the Cognitive Proficiency Index (CPI). In this way, the WISC-IV allows for better discrimination between abilities on the aggregate level compared to its previous editions. However, only a limited amount of published information is available regarding its utility when assessing clinical samples. In particular, since studies of other clinical groups (e.g., children with traumatic brain injury; children with attention deficit/hyperactivity disorder; children with High functioning autism) have shown profile differences when comparing the WISC-IV to older versions of the WISC (respectively, Donders & Jenke, 2008; Mayes & Calhoun, 2008), it is important to define the WISC-IV profile in children with Autism spectrum disorder of level 1 (according to DSM-V).

WISC-IV test score results for some special groups are included in the *WISC-IV Technical and Interpretive Manual* (Wechsler, 2008) and in the *Essentials of WISC-IV Assessment* (Flanagan & Kaufman, 2009), to help provide information about the test's specificity and its clinical utility for diagnostic assessment: the special groups studied include children with autistic disorder and with Asperger syndrome according to DSM-IV. According to the *WISC-IV Technical and Interpretive Manual*, the clinical autistic disorder sample (n = 16) scored significantly lower than the matched control group on all 4 indices and the Full Scale IQ (FSIQ), with large effects sizes. The largest effect sizes were obtained for the Verbal Comprehension Index (VCI), the Processing Speed Index (PSI) and the FSIQ. These results were consistent with other studies indicating that individuals with autistic disorder demonstrate lowered general intellectual functioning, especially on verbal and processing speed tasks, and obtain relatively higher scores on perceptual tasks (Flanagan & Kaufman, 2009; Goldstein, Minshew, Allen & Seaton, 2002; Kuriakose, 2014; Liss et al., 2001; Mayes & Calhoun, 2003, 2004; Nader, Courchesne, Dawson & Soulières, 2016; Nader, Jelenic & Soulières, 2015). In contrast, the 40 individuals in the Asperger disorder group scored significantly lower than the matched control group on the PSI, WMI and the Full Scale IQ with large effect size, and a small effect for the PRI and a negligible effect for the VCI. These results are consistent with other research with individuals with Asperger's disorder, which had lower processing speed performance and maintained verbal ability (Ambery, Russel, Perry, Morris & Murphy, 2006; Cederlund & Gillberg, 2004; Flanagan & Kaufman, 2009; Koyama, Tachimori, Osada, Taked & Kurita, 2007; Nader et al., 2015; Spek, Scholte & Van Berckelaer-Onnes, 2008).

In a study by Mayes and Cahloun (2008) 54 children 6-14 years of age with High functioning autism scored above average at WISC-IV in Perceptual Reasoning Index, Verbal Comprehension Index, and General Ability Index, and scores below average 100 at Working Memory Index and Processing Speed Index. The GAI was significantly higher than FSIQ that doesn't differ significantly from the population mean. In another paper, Oliveras-Rentas and colleagues (Oliveras-Rentas, Kenworthy, Robertson, Martin & Wallace, 2012) administered the WISC-IV to a clinical sample of 22 children with High functioning autism, 22 with Asperger syndrome

and 12 with pervasive developmental disorders. Comparing this clinical sample with the normal population the only index score that was significantly lower than the population was the Processing Speed Index (PSI). Most notably, significantly lower scores were found for the *Coding*, *Symbol Search* and *Comprehension* subtests, while the *Similarities* and *Matrix Reasoning* subtests were significantly higher. Hence, these results confirm strengths on WISC-IV structured and motor-free subtests (e.g., *Similarities* and *Matrix Reasoning*) and weakness on subtests with more complex/social language demands (e.g. *Comprehension*). However, the manuscript does not provide us any information about the differences between the three clinical sub-samples and fails to take into account the state variables that could affect their intellectual performance. It may not be methodologically correct to compare the performance of the clinical sample with the population mean when studying such small samples; variables such as the parents' educational level or different clinical diagnoses may create a bias for evaluation. In the end, little is yet known about cognitive strengths or ASD difficulties, and the size of ASD samples has always been very small, but using WISC-IV with a larger ASD sample could be useful to better differentiate and eliminate confusion factors by highlighting possible strengths in verbal abilities and weaknesses in memory, attention, graphomotor and processing speed.

In this research we have analyzed the WISC-IV scores in 80 ASD - level 1 (ASD-1) children and adolescents without verbal and intellectual disabilities to study their specific cognitive profile and to compare the results with previous research discussed above. More in particular, we wanted to study the differences in scores on the subtests, on the four core and two additional indices of the WISC-IV. In addition, we studied the difference Max-Min of four core indices as an expression of the unitarity ability of the IQ of the subject (Flanagan & Kaufman, 2009). Flanagan and Kaufman (2009, p. 143) used to define the unitary ability as "an ability (…) that is represented by a cohesive set of scaled scores, each reflecting slightly different or unique aspects of the ability". To measure the unit skill, then, Flanagan and Kaufman used the difference between the highest score (Max) and the lowest score (Min) obtained by a participant in the four indexes of the latest editions of the Wechsler scales. Therefore, the main criterion to define the non-interpretability, or rather the poor cohesion, of FSIQ is based on the relative infrequency of the Max-Min difference between the 4 indices.

Finally, it is especially important to check whether the differences between FSIQ and two additional indices (GAI and CPI), can discriminate between the clinical group and the control group. Everything is designed to identify any patterns of intellectual efficiency of the group diagnosed with High functioning ASD-level 1. We hypothesize they show relative weaknesses on the WMI and PSI indices of WISC-IV, while performing relatively well on VCI and PRI indices because the WISC-IV subtests measure verbal-language and visual reasoning variables without a confounding motor and memory components.

METHOD

Participants

The WISC-IV was administered to 80 individuals (64 males and 16 females), aged 6-16 years of age (*Mean* = 9.81, *SD* = 2.90), who were identified as Autism spectrum disorder of level 1 (ASD-1) (according the DSM-V criteria; APA, 2013), without verbal and intellective deficits. Each child of the clinical group were evaluated following all the requirements for a clinical diagnosis of ASD-1, and they received a comprehensive neuropsychological evaluation by an expert clinician and a multidisciplinary team evaluation that included a detailed medical and developmental history, an extensive diagnostic battery, as well as administration of the *Autism Diagnostic Interview-Revised* (*ADI-R*; Lord, Rutter & Le Couteur, 1994), by a trained research reliable clinician; *Autism Diagnostic Observation Schedule – Second edition* (*ADOS-2*; Lord et al., 2011). Individual diagnosed with Autism spectrum disorder were excluded from this study if they had general cognitive ability scores more than 2 *SD*s below the mean (i.e., FSIQ<70). All participants were evaluated at the Multidisciplinary Unit, Department of Prevention of Public Health ASL2 of Abruzzo, Lanciano-Vasto-Chieti. The Unit consists of a child neuropsychiatry, psychologists and a social worker, operating into the prevention of school medicine. Parents of children gave their authorization, through an informed consent. The research was approved by Ethic Committee of the Child Neuropsychiatry Units. Data were collected between 2017 and 2019. Nine individuals did not agree to participate in the evaluations, and four abandoned the research.

This clinical sample was compared with typically developing children who were part of the Italian WISC-IV

standardization sample, matched for gender and age of ASD-1 children and education of both parents. The use of the latter variable is due by results of two studies showed that, while parental influence on children's subtests, FSIQs, indices and GAI of WISC-III and WISC-IV is independent of the parent's gender, it varies as a function of the parent's level of education (Cianci, Orsini, Hulbert & Pezzuti, 2013; Pezzuti, Farese & Dawe, 2019). So, the two groups (i.e., ASD-1 and matched-paired control) included exactly the same numbers of males and females, and were perfectly matched for age in years and months, and education level of both parents.

Instruments

The WISC-IV (Orsini, Pezzuti & Picone, 2012; Wechsler, 2003) was been used. The WISC-IV, in addition to Full Scale IQ (FSIQ), expected four indices, such as the Verbal Comprehension Index (VCI; the subtests are: *Similarities*, *Vocabulary* and *Comprehension*), Perceptual Reasoning Index (PRI; the subtests are: *Block Design*, *Picture Concepts* and *Matrix Reasoning*), Working Memory Index (WMI; the subtests are: *Digit Span and Letter-Number Sequencing*) and Processing Speed Index (PSI; the subtests are: *Coding and Symbol Search*) and two additional indices as General Ability Index (GAI; the subtests are those of VCI and PRI indices) and Cognitive Proficiency Index (CPI; the subtests are those of WMI and PSI indices). The WISC-IV dependent variables studied in this paper are the 10 core subtest scores, the 4 core indices (VCI, PRI, WMI, PSI) and the Full Scale IQ (FSIQ). However, it is possible that the FSIQ to be affected by some variability of four underlying dimensions and this must always be taken into account by the clinicians. So, two optional composite indices have been proposed alongside the FSIQ and these are known as the General Ability Index (GAI) and the Cognitive Proficiency Index (CPI). The GAI, introduced by Prifitera, Weiss and Saklofske (1998), represents a composite measure of cognitive ability comprises the verbal comprehension and perceptual reasoning subtests that, in comparison with FSIQ, minimizes the impact of working memory and processing speed, and reflects reasoning abilities. The CPI, proposed by Dumont and Willis (2001), is therefore an index that summarizes the outcomes of both the working memory and processing speed subtests. The CPI, represented by a quick visual speed, an efficient memory and good mental control, helps fluid reasoning and acquisition of new information, and reduces the cognitive load required by newer or more difficult tasks (Weiss et al., 2006). So, in the present research we used these 2 optional indices (GAI and CPI); differences between indices (FSIQ vs GAI, FSIQ *vs* CPI, GAI *vs* CPI) referring data of Italian WISC-IV standardization (Orsini & Pezzuti, 2014; Orsini & Pezzuti, 2016).

Another WISC-IV dependent variable is the difference between the highest score (Max) and the lowest score (Min) in the four indices of the test as a measure of the unitary ability of FSIQ according to Flanagan and Kaufman (2009). However, as demonstrated in a paper of Orsini, Pezzuti and Hulbert (2014), the statistical method used by Flanagan and Kaufman (2009) to find the threshold 23, didn't really fit for purpose. In the Italian WISC-IV standardization sample the correct statistical method was carried out bringing out the correct threshold of 40. So, when the Max-Min difference score between four core subtests is equal to or greater than 40 scores, then it is considered very rare and it possible to conclude that the FSIQ score cannot be interpreted as one unitary ability of intelligence.

Data analysis

After confirming that assumptions of normality and homogeneity of the variances of the two groups were satisfied using a Levene's test. Various analyses of variance (ANOVA) were used to compare the ASD-1 group with the matched-paired control group. We compared the mean differences between the first and second group (clinical vs control) for each dependent variable of the WISC-IV. Though a *p*-value can determine whether an effect exists, it will not reveal the effect's size. The effect's size provides information regarding its practical significance, whereas the p-value does not assess practical significance. Knowing an effect's magnitude allows one to ascertain the practical significance of statistical significance. Statistical significance can always be reached if there is a large enough sample size, unless the effect size is 0. Even a large effect may not be statistically significant if the sample size is too small. Therefore, according to Cohen (1990, p. 1307), "The primary product of a research inquiry is one or more measures of effect size, not p values". So, we reported also eta-squared as a measure of effect size which can be interpreted using Cohen's (1988) guidelines for determining small (.01), medium (.06), and large (.14) effects.

RESULTS

By comparing the clinical group of subjects with a diagnosis of ASD-1 to the control group on WISC-IV subtests (see Table 1), the eta-squared ranges from .00 (for *Block Design*, *Matrix Reasoning* and *Comprehension*) to .27 (for *Digit Span*), then, from null to very large effect. In particular, *Digit Span*, *Letter-Number Sequencing*, *Symbol Search*, *Coding* and *Cancellation* performances tended to be lower for ASD-1 group than for control group, while for the other subtest there would not be any noteworthy differences. Such subtests are mostly dependent on verbal memory and processing speed abilities.

For what concerns the indices, Table 2 shows a large effect-size for Working Memory, Processing Speed, Full Scale IQ and Cognitive Proficiency indices: the ASD-1 group has a significantly lower mean than the control group.

Analyzing the variability of the indices, that is the difference between the highest (Max) and lowest (Min) of four core indices (VCI, PRI, WMI, and PSI), from the results showed in Table 2, a large effect-size emerges: the ASD-1 group has a significantly higher mean than the control group. In particular, the mean value of the group is next the clinical cut-off, that in the Italian standardization sample (Orsini et al., 2014) it is equal to 40, and 29 ASD-1 subjects (36.2%) have a statistically significant and a rare Max-Min difference, that is greater than cut-off 40 IQs scores (with range difference between 41 to 82); it is rare because it occurs in less than 6.7% on normal subjects. On the contrary, only 3 control subjects (3.7%: with range difference between 41 to 46) have a statistically significant and rare Max-Min difference between four indices. In conclusion, in the ASD-1 group a very wide and rare variability among the 4 indices emerged.

From the study of GAI (General Ability Index) and CPI (Cognitive Proficiency Index), (see Table 2), the ASD-1 group has a mean GAI score (sum of VCI and PRI) of almost 10 points higher than their FSIQ, in the normal group the mean GAI score is almost equal to the FSIQ. This result indicates that WMIs and/or PSIs negatively affect the expression of their general intellectual ability as measured by the FSIQ score in the ASD-1 clinical group. In contrast, the CPI (sum of WMI and PSI) is almost 12 IQ points lower than the FSIQ score in the Autism spectrum disorder group, while this difference is minimal (almost 2 IQ points) in the normal control group. Finally, the average difference between GAI and CPI is in favour of GAI in the clinical group (21.05 IQ points) and is different from the control group which is lower (1.28 IQ points).

Finally, analyses to study the cognitive profiles of the ASD-1 group were carried out. In particular, the subtest performances belonging to the four indices, and the seven indices were compared with each other. The results reported in Table 3 show that, within each of the four indices, there were no relevant discrepancies between the pairs of subtests: the ASD-1 group shows generally homogeneous cognitive profiles between subtests within each index. However, from the results on comparisons between pairs of indices it emerges that both Verbal Comprehension and Perceptual Reasoning indices differ significantly from Working Memory and Processing Speed indices with higher performances in VCI and PRI. On the contrary, there are no differences between Verbal Comprehension and Perceptual Reasoning indices and between the Working Memory and Processing Speed indices. Comparisons between the more general intelligence indices (FSIQ, GAI and CPI) are also all significant, with the highest GAI and the lowest CPI.

DISCUSSION AND CONCLUSION

The diagnostic characterization of ASD patients without intellectual disability is rather difficult because of the milder symptoms and the compensatory abilities (Frith, 2004), which allow the patients to be well adapted in the social environment. Often, they show only secondary symptoms, frequently psychiatric comorbidities with age.

In the present paper, the WISC-IV, administered to children and adolescent to ASD without intelligence deficit, captures their verbal ability and perceptual reasoning strengths, while identifying their memory, attention, grapho-motor and processing speed weaknesses. These results are consistent with some papers (i.e., Mouga et al., 2016; Nader et al., 2016).

The lowest performances of ASD children and adolescents without intellectual deficit were on subtests that make up the Processing Speed Index and Working Memory Index, reflecting the sensitivity of these measures to generalized cognitive impairment. This result is also confirmed by the absence of a difference between the ASD and control groups in the General Ability Index and the large difference to the Cognitive Proficiency Index in favour of the control group. As discussed previously, Saklofske and colleagues (Saklofske,

Table 1 – Comparisons on subtests between Autism spectrum disorder (level 1) group and matched-paired control group

	Autism spectrum disorder (level 1) group (n = 80)		Matched-paired control group (n = 80)				
	Mean	**SD**	**Mean**	**SD**	**F**	***p***	**Eta-squared**
WISC-IV core subtests							
Block Design	10.76	3.92	11.09	3.16	.33	.565	.00
Similarities	11.29	3.86	11.82	2.82	1.01	.316	.01
Digit Span	7.95	2.50	10.97	2.50	58.43	<.001	.27
Picture Concepts	10.35	3.37	10.87	2.39	1.27	.262	.00
Coding	7.26	3.33	10.30	3.25	34.09	<.001	.18
Vocabulary	10.29	3.88	11.34	3.04	3.62	.059	.02
Letter-Number Sequencing	7.71	3.13	10.84	3.05	39.81	<.001	.21
Matrix Reasoning	10.76	3.48	10.79	3.01	.00	.961	.00
Comprehension	10.35	4.65	10.57	2.95	.13	.721	.00
Symbol Search	7.975	3.18	11.20	3.02	43.22	<.001	.21
WISC-IV supplemental subtests							
Picture completion	10.61	3.38	11.60	2.89	3.95	.049	.02
Cancellation	6.86	4.13	10.05	3.19	29.82	<.001	.16
Information	9.59	3.71	11.37	2.56	12.60	<.001	.07
Arithmetic	8.32	3.62	9.81	2.74	8.64	.004	.05
Word Reasoning	9.05	3.40	10.92	3.02	13.59	<.001	.08

Note. Eta-squared values were calculated as a measure of effect size. and results were interpreted using Cohen's (1988) guidelines for determining small (.01). medium (.06). and large (.14) effects.

Table 2 – Comparisons on indices between Autism spectrum disorder group and matched-paired control group

	Autism spectrum disorder (level 1) group (n = 80)		Matched-paired control group (n = 80)				
	Mean	**SD**	**Mean**	**SD**	**F**	***p***	**Eta squared**
Indices							
1. Verbal Comprehension Index (VCI)	103.60	21.54	107.77	14.64	2.05	.154	.01
2. Perceptual Reasoning Index (PRI)	102.70	17.92	105.86	13.49	1.59	.209	.01
3. Working Memory Index (WMI)	86.85	14.52	105.17	14.56	63.55	<.001	.29
4. Processing Speed Index (PSI)	85.14	15.64	104.11	14.64	62.77	<.001	.28
Full Scale Intelligence Quotient (FSIQ)	94.89	15.45	107.84	13.30	32.26	<.001	.17
Diff. Max-Min of 4 indices	38.96	14.66	25.17	9.32	50.38	<.001	.24
Additional indices							
1. General Ability Index (GAI)	104.30	16.82	107.44	13.72	1.66	.199	.01
2. Cognitive Proficiency Index (CPI)	83.25	15.71	106.16	13.87	92.88	<.001	.38
Differences between indices							
FSIQ-GAI	–9.38	7.33	.40	5.66	88.87	<.001	.36
FSIQ-CPI	11.64	13.73	1.67	9.32	28.27	<.001	.16
GAI-CPI	20.85	18.37	1.27	14.64	54.18	<.001	.26

Note. Eta-squared values were calculated as a measure of effect size. and results were interpreted using Cohen's (1988) guidelines for determining small (.01). medium (.06). and large (.14) effects.

Table 3 – Cognitive profiles of the ASD-1 group: comparisons between pairs of subtests within 4 indices. and comparisons between the seven indices of the WISC-IV (n = 80)

Indices	Comparisons between subtests in ASD-1 group	$Mean_{[1]}$	$SD_{[1]}$	$Mean_{[2]}$	$SD_{[2]}$	F	*p*	Eta squared
VCI	$Similarities_{[1]}$ vs $Vocabulary_{[2]}$	11.29	3.86	10.29	3.88	7.82	.006	.09
	$Similarities_{[1]}$ vs $Comprehension_{[2]}$	11.29	3.86	10.35	4.65	3.77	.055	.05
	$Vocabulary_{[1]}$ vs $Comprehension_{[2]}$	10.29	3.88	10.35	4.65	.00	1.00	.00
PRI	$Block\ Design_{[1]}$ vs $Picture\ Concept_{[2]}$	10.76	3.92	10.35	3.37	.52	.474	.01
	$Block\ Design_{[1]}$ vs $Matrix\ Reasoning_{[2]}$	10.76	3.92	10.76	3.48	.00	.99	.00
	$Picture\ Concept_{[1]}$ vs $Matrix\ Reasoning_{[2]}$	10.35	3.37	10.76	3.48	.50	.479	.01
WMI	$Digit\ Span_{[1]}$ vs $Letter\text{-}Number\ Sequencing_{[2]}$	8.05	2.52	7.71	3.13	1.08	.030	.01
PSI	$Coding_{[1]}$ vs $Symbol\ Search_{[2]}$	7.26	3.33	7.97	3.18	3.59	.065	.04
	Comparisons between Indices in ASD-1 group							
	$VCI_{[1]}$ vs $PRI_{[2]}$	103.60	21.54	102.70	17.92	.10	.753	.00
	$VCI_{[1]}$ vs $WMI_{[2]}$	103.60	21.54	86.85	14.52	40.10	<.001	.34
	$VCI_{[1]}$ vs $PSI_{[2]}$	103.60	21.54	85.14	15.64	47.22	<.001	.37
	$PRI_{[1]}$ vs $WMI_{[2]}$	102.70	17.92	86.85	14.52	43.92	<.001	.36
	$PRI_{[1]}$ vs $PSI_{[2]}$	102.70	17.92	85.14	15.64	75.02	<.001	.49
	$WMI_{[1]}$ vs $PSI_{[2]}$	86.85	14.52	85.14	15.64	.60	.440	.01
	$FSIQ_{[1]}$ vs $GAI[_{[2]}$	94.89	15.45	104.30	16.82	129.43	<.001	.58
	$FSIQ_{[1]}$ vs $CPI_{[2]}$	94.89	15.45	83.25	15.71	53.91	<.001	.42
	$GAI_{[1]}$ vs $CPI_{[2]}$	104.30	16.82	83.25	15.71	96.67	<.001	.57

Note. Eta-squared values were calculated as a measure of effect size. and results were interpreted using Cohen's (1988) guidelines for determining small (.01). medium (.06). and large (.14) effects.

Gorsuch, Weiss, Rolfhus & Zhu, 2005), indicated that more than 60% of the children with diagnoses of ASD in the WISC-IV showed GAIs five or more points greater than their FSIQs. In our result, the clinical group shows GAI mean almost 10 IQ points greater than their FSIQ; at the same time, the FSIQ mean is almost 12 points IQ greater than CPI, and they present a significant difference between GAI and CPI (about 21 points of IQ). These results indicate that Working Memory and Processing Speed indices negatively affect the expression of general intellectual abilities measured by Full Scale IQ.

So, in the ASD children the weaknesses in the WMI and PSI indices and the strengths in VCI and PRI demonstrate an adequate performance in verbal language and visual reasoning, since these indices do not have a confounding motor component and executive processing verbal short-term memory. Motor clumsiness is considered an endophenotype of ASD (Dziuk et al., 2007; Rourke, 2009). In the previous versions of the Wechsler scales, it was related to impaired performance index in both patients with only motor dyspraxia and those with ASD (Rourke, 1989). The WISC-IV profile in the ASD sample provides more elements and is further evidence that clumsiness is caused by a spatial working memory deficit, part of a more general impairment in nonverbal abilities (Klin, Volkmar, Sparrow, Cicchetti & Rourke, 1995), and by the deficit in planning and executing movement, rather than only by a motor skills deficit (Blake, Turner, Smoski, Pozdol & Stone, 2003). Motor planning is directly related to working memory and, therefore, we corroborate the results of Rinehart and colleagues (Rinehart, Bradshaw, Moss, Brereton & Tonge, 2001), who demonstrated a difficulty in maintaining attention and maintaining the appropriate preparatory set in working memory for ASD.

Finally, since Flanagan and Kaufman (2009) stated that a wide and rare variability between indices can be an expression of an IQ that cannot be interpreted as a unitary ability, we studied this variability in the ASD group. Using cut-off value extracted from the Italian standardization sample of WISC-IV (see Orsini et al., 2014), the 36% of ASD children had a rare and large difference between the 4 indices and then the FSIQ could be deemed uninterpretable as unitary and cohesive ability.

There is research that has highlighted how the cognitive profiles of ASD children can be characterized by a fall in verbal tests and therefore a fall in the total IQ of a Wechsler scales (i.e. Flanagan & Kaufman, 2009; Goldstein et al, 2002; Liss et a1., 2001; Kuriakose, 2014; Wechlser, 2008). As well as there are authors concluding that such scales (the WISC-III and WISC-IV) when compared to the *Raven Matrices* or the *Leiter International Performance Scale – 3*, may underestimate the overall level of intelligence of these clinical subjects (i.e., Dawson, Soulieres, Gernsbacher & Mottron, 2007; Giofrè et al., 2019; Nader et al., 2016). However, these conclusions probably have limitations: 1) they are formulated on samples that are almost very small; 2) clinical samples are heterogeneous, in the sense that ASD subjects are not distinguished in with and without intellectual deficit; 3) control samples are rarely perfectly matched. Although in the paper of Giofrè et al. (2019) 31 children ASD with IQ>70 and 19 children with IQ<70 were compared, the first group had a very low mean FSQI score compared to the results of the present research on a group of 80 children ASD with IQ>70. On the other hand, already other research have highlighted as children ASD with higher cognitive abilities and children ASD with lower cognitive abilities, present different cognitive profiles with important differences in the strengths and weaknesses (Mayes & Calhoun, 2003; Mouga et al., 2016). And in part the results of the present paper confirm it, because if children ASD with IQ>70 perform on mean in ICV and IRP that do not differ from those of the paired control group, the profile could be likely to be different from what could emerge with children ASD with IQ<70. This hypothesis has in fact already been confirmed in the research of Mouga et al. (2016) administering the WISC-III to ASD group with normal/high IQ and lower IQ. Therefore, the use of a Wechsler scale with the 4 core indices and the two supplemental ones, will better define the intellectual profile of the ASD clinical group.

In summary, in this paper we aimed to identify an endophenotype for ASD with normal cognitive functioning at the WISC-IV evaluation. Results suggest that ASD's with high functioning profile cannot be interpreted as a unitary and cohesive ability, represented from simply FSIQ, but we can obtain a better assessment of cognitive level in ASD subjects using separately GAI and CPI indices. The high discrepancy among the 4 core indices could shape up a characteristic endophenotype of ASD and be used not only for general cognitive assessment, but even as a contribution to differential diagnostic assessment of ASD. We should always ask ourselves what is the advantage of using FSIQ alone with children and adolescent with ASD that does not establish suitability to receive additional services or provide the most useful information for educational planning. In other words, the labelling of intellectual disability to an ASD children with moderate to severe disability can be

inaccurate and has no educational function but adds another stigma. An effective assessment for this population should use a functional approach with a differential diagnosis that analyses intellectual profiles and allows psychologists and educators to define the skills that need to be developed and the educational methods most likely to be successful. An assessment of the overall intellectual level of ASD children is less likely to provide life-enhancing progress than an assessment that identifies strengths, weaknesses and ways to address deficits. As such, a cognitive assessment is necessary to measure the level of functioning of an individual in the various domains, specifying whether ASD is present with or without accompanying intellectual or linguistic disabilities, in order to identify appropriate interventions and supports.

The strengths of the present research that distinguish it from previous research are: the use of performance about all 15 WISC-IV subtests (core and supplemental subtests) to get more information on cognitive profiles of ASD group; the greater size, compared to other research, of the clinical sample of 80 children and adolescents diagnosed with Autism spectrum disorder – level 1 (ASD-1) without intellectual disability (with FSIQ>70); and the use of a perfectly matched-paired control sample for age, gender of children and parents' level education, sample very often absent or untreated in pairing for important status variables in other research.

However, this research is not without limits, surely the most important is that we do not have a sample of comparison of ASD children with intellectual disabilities (i.e. with IQ<70). Another limitation, we have not used data on adaptive behaviour (i.e., ADOS-2), although we have already seen how WISC-IV indices can predict the adaptive functioning of children with ASD with high functioning (e.g. Oliveras-Rentas et al., 2012).

Despite these limitations, we believe that this study provides clinicians and researchers with important insights into the intellectual functioning of ASD children with high functioning, demonstrating that the way intelligence is assessed in these children is important and requires careful analysis of cognitive profiles rather than focusing on FSIQ.

However, the future study should be supported by: a) to use other measurements of the average in verbal and visuospatial competences, as the fall in Working Memory and Processing Speed, for example with neuropsychological measurements; b) to know if there are subgroups of ASD children with various levels of functioning, given the variability of cognitive patterns that often emerge in research (e.g. as defined by the DSM-5 severity levels and specifiers; American Psychiatric Association, 2013); c) to compare the profile of ASD children without intellectual deficit and ASD children with intellectual deficit at the WISC-IV, to further information the question if the strengths and deficits are the same in high and low functioning ASD; d) to exam cognitive patterns in ASD children with and without language (receptive and/or expressive) impairment or disorder; e) to deepen how cognition is related to the main characteristics of ASD and adaptive behavior, as well as to the associated psychopathology; f) to compare the ASD-1 profile at the WISC-IV with other diagnoses, e.g. patients with (only) developmental coordination disorder; g) to search for any relationships between motor coordination disorders and CPI indices, also discriminating further between WMI and PSI indices; h) to compare the ASD-1 profile at WISC-IV with than WISC-V not yet available in Italy Regarding this last point, a paper was recently published in which Kuehnel and colleagues (Kuehnel, Castro & Furey, 2019) compared the performance at the Verbal Comprehension Index (VCI) of WISC-IV and WISC-V of ASD and ADHD children. From results the changes in VCI (from WISC-IV to WISC-V) subtest scores were minimal although a statistically significant increase of 5 IQ scores in VCI score occurred. More in particular, for both WISC-IV and WISC-V, the authors found significant differences between pairs of verbal subtests (*Similarities*, *Vocabulary* and *Comprehension*) with *Similarities* subtest was a relative strength and *Comprehension* subtest is weakness. The distint pattern performance (*Similarities* > *Vocabulary* > *Comprehension*) emerged confirming data in literature (e.g., Mayes & Calhoun, 2008; Zayat et al., 2011). These results are partially superimposable to those that emerged in the present research in which our ASD high functioning children and adolescents showed higher performance at the subtest of *Similarities*, while the performance at *Vocabulary* and *Comprehension* subtests are lower but almost similar. We agree with the conclusions of Kuehnel and colleagues (2019) that verbal intellect measurements are particularly important for ASD sufferers, since language disorders are quite common for many individuals across the autistic spectrum and, consequently, their performance on verbal intellect measurements (especially when there is an intellectual disability) is often lower.

Conflict of interest. The Authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

AMBERY, F.Z., RUSSEL, A.J., PERRY, K., MORRIS, R. & MURPHY, D.G.M. (2006). Neuropsychological functioning in adults with Asperger syndrome. *Autism*, 10 (6), 551-564.

AMERICAN PSYCHIATRIC ASSOCIATION (2013). DSM-5. *Diagnostic and statistical manual of mental disorders (5th ed.)*. Arlington, VA: American Psychiatric Publishing (It. ed.: *DSM-5. Manuale diagnostico e statistico dei disturbi mentali, quinta edizione*. 2014, Milano: Raffaello Cortina).

BLAKE, R., TURNER, L.M., SMOSKI, M.J., POZDOL, S.L. & STONE, W.L. (2003). Visual recognition of biological motion is impaired in children with autism. *Psychological Science*, *14*, 151-157.

CEDERLUND, M. & GILLBERG, C. (2004). One hundred males with Asperger syndrome: A clinical study of background and associated factors. *Developmental Medicine & Child Neurology*, *46*, 652-660.

CIANCI, L., ORSINI, A., HULBERT, S. & PEZZUTI, L. (2013). The influence of parents' education in the Italian standardization sample of the WISC-III. *Learning and Individual Differences*, *28*, 47-53.

COHEN, J. (1988). *Statistical power analysis for the behavioral sciences* (*2nd* ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

COHEN, J. (1990). Things I have learned (so far). *American Psychologist*, *45*, 1304-1312.

DAWSON, M., SOULIERES, I., GERNSBACHER, M.A. & MOTTRON, L. (2007). The level and nature of autistic intelligence. *Psychological Science*, *18* (8), 657-662.

DONDERS, J. & JENKE, K. (2008). Criterion validity of the Wechsler Intelligence Scale for Children-Fourth Edition after pediatric traumatic brain injury. *Journal of the International Neuropsychological Society*, *14*, 651-655.

DUMONT, R. & WILLIS, J. (2001). *Use of the Tellegen & Briggs formula to determine the Dumont-Willis Indexes (DWI-1 & DWI-2) for the WISC-IV*. Retrieved from http://alpha .fdu.edu/psychology/

DZIUK, M.A., GIDLEY LARSON, J.C., APOSTU, A., MAHONE, E.M., DENCKLA, M.B. & MOSTOFSKY, S.H. (2007). Dyspraxia in autism: Association with motor, social, and communicative deficits. *Developmental Medicine and Child Neurology*, *49*, 734-739.

FLANAGAN, D.P. & KAUFMAN, A.S. (2009). *Essentials of WISC-IV Assessment, 2nd edition*. Hoboken: John Wiley & Sons, Inc.

FRITH, U. (2004). Emanuel Miller lecture: Confusions and controversies about Asperger syndrome. *Journal of Child Psychology and Psychiatry*, *45* (4), 672-686.

GHAZIUDDIN, M. & MOUNTAIN-KIMCHI, K. (2004). Defining the intellectual profile of Asperger syndrome: Comparison with high-functioning autism. *Journal of Autism and Developmental Disorders*, *34* (3), 279-284.

GIOFRÈ, D., PROVAZZA, S., ANGIONE, D., CINI, A., MENAZZA, C., OPPI, F. & CORNOLDI, C. (2019). The intellectual profile of children with autism spectrum disorders may be underestimated: A comparison between two different batteries in an Italian sample. *Research in Developmental Disabilities*, *90*, 72-79.

GOLDSTEIN, G., MINSHEW, N.J., ALLEN, D.N. & SEATON, B.E. (2002). High-functioning autism and schizophrenia: A comparison of an early and late onset neurodevelopmental disorder. *Archives of Clinical Neuropsychology*, 17, 461-475.

HWANG, E.J. & SHADMEHR, R. (2005). Internal models of limb dynamics and the encoding of limb state. *Journal of Neural Engineering*, *2* (3), S266-S278.

KAUFMANN, W.E. (2012). *DSM-5: The new diagnostic criteria for Autism Spctrum Disorders*. Presented at Autism Consortium Research Symposium, Boston, October 24, 2012.

KLIN, A., VOLKMAR, F.R., SPARROW, S.S., CICCHETTI, D.V. & ROURKE, B.P. (1995). Validity and neuropsychological characterization of Asperger syndrome: Convergence with nonverbal learning disabilities syndrome. *Journal of Child Psychology and Psychiatry*, *36*, 1127-1140.

KOYAMA, T., TACHIMORI, H., OSADA, H., TAKED, T. & KURITA, H. (2007). Cognitive and symptom profiles in Asperger's syndrome and high-functioning autism. *Psychiatry and Clinical Neurosciences*, *61*, 99-104.

KUEHNEL, C.A., CASTRO, R. & FUREY, W.M. (2019). A comparison of WISC-IV and WISC-V verbal comprehension index scores for children with autism spectrum disorder. *The Clinical Neuropsychologist*, *33* (6), 1127-1137.

KURIAKOSE, S. (2014). Concurrent validity of the WISC-IV and DAS-II in children with autism spectrum disorder. *Journal of Psychoeducational Assessment, 32* (4), 283-294.

LISS, M., FEIN, D., ALLEN, D., DUNN, M., FEINSTEIN, C. & MORRIS, R. (2001). Executive functioning in high-functioning children with autism. *Journal of Child Psychology and Psychiatry*, *42* (2), 261-270.

LORD, C., RUTTER, M., DiLAVORE, P.C., RISI, S., GOTHAM, K. & BISHOP, S.L. (2011). *Autism Observation Schedule – Second Edition (ADOS-2)*. (It. ed.: C. Colombi, R. Tancredi, A. Persico & R. Faggioli. 2013, Hogrefe).

LORD, C., RUTTER, M. & LE COUTEUR, A. (1994). Autism

Diagnostic Interview-Revised: A revised version of a Diagnostic Interview for Caregivers of Individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *24* (5), 659-685.

MAYES, S.D. & CALHOUN, S.L. (2003). Analysis of WISC-III, Stanford-Binet IV, and academic achievement test scores in children with autism. *Journal of Autism and Developmental Disorders*, *33* (3), 329-341.

MAYES, S.D. & CALHOUN, S.L. (2004). Similarities and differences in Wechsler Intelligence Scale for Children – Third Edition (WISC-III) profiles: Support for subtest analysis in clinical referrals. *Clinical Neuropsychologist*, *18*, 559-572.

MAYES, S.D. & CALHOUN, S.L. (2008). WISC-IV and WIAT-II profiles in children with high-functioning autism. *Journal of Autism and Developmental Disorders*, *38*, 428-439.

NADER, A.M., COURCHESNE, V., DAWSON, M. & SOULIÈRES, I. (2016). Does WISC-IV underestimate the intelligence of autistic children? *Journal of Autism and Developmental Disorders*, *46* (5), 1582-1589.

NADER, A.M., JELENIC, P. & SOULIÈRES, I. (2015). Discrepancy between WISC-III and WISC-IV cognitive profile in autism spectrum: What does it reveal about autistic cognition? *PLoS One*, *10* (12).

OLIVERAS-RENTAS, R.E., KENWORTHY, L., ROBERTSON, R.B., MARTIN, A. & WALLACE, G.L. (2012). WISC-IV profile in high-functioning autism spectrum disorders: impaired processing speed is associated with increased autism communication symptoms and decreased adaptive communication abilities. *Journal Autism Developmental Disorder*, 42 (5), 655-664.

ORSINI, A. & PEZZUTI, L. (2014). L'indice di abilità generale della scala WISC-IV. *Psicologia Clinica dello Sviluppo*, *18* (2), 301-310.

ORSINI, A. & PEZZUTI, L. (2016). *L'interpretazione clinica della WISC-IV alla luce della taratura italiana. Con CD Rom allegato*. Firenze: Giunti OS Organizzazioni Speciali.

ORSINI, A., PEZZUTI, L. & HULBERT, S. (2014). The unitary ability of IQ in the WISC-IV and its computation. *Personality and Individual Differences*, *69*, 173-175.

ORSINI, A., PEZZUTI, L. & PICONE, L. (2012). *WISC-IV. Contributo alla taratura italiana*. Firenze: Giunti OS Organizzazioni Speciali.

PEZZUTI, L., FARESE, M. & DAWE, J. (2019). Relation between parents' education and sons' intellectual profile on Wechsler Intelligence Scale for Children. *BPA-Applied Psychology Bulletin (Bollettino di Psicologia Applicata)*, *67* (284), 2-9.

PRIFITERA, A., WEISS, L.G. & SAKLOFSKE, D.H. (1998). The WISC-III in context. In A. Prifitera & D.H. Saklofske (Eds.), *WISC-III clinical use and interpretation: Scientist-practitioner perspectives*. San Diego, CA: Academic Press.

RINEHART, N.J., BRADSHAW, J.L., MOSS, S.A., BRERETON, A.V. & TONGE, B.J. (2001). A deficit in shifting attention present in high functioning autism but not Asperger's disorder. *Autism*, *5* (1) 67-80.

ROURKE, B. (1989). *Non verbal learning disabilities: The syndrome and the model*. New York: Guilford Press.

SAKLOFSKE, D.H., GORSUCH, R.L., WEISS, L.G., ROLFHUS, E. & ZHU, J. (2005). Clinical interpretation of the WISC-IV FSIQ and GAI. In A. Prifitera, D.H. Saklofske & L.G. Weiss (Eds.), *WISC-IV clinical use and interpretation: Scientist-practitioner perspectives*. New York: Academic Press.

SPEK, A.A., SCHOLTE, E.M. & VAN BERCKELAER-ONNES, I.A. (2008). Brief report: The use of WAIS-III in adults with HFA and Asperger syndrome. *Journal of Autism and Developmental Disorders*, *38* (4), 782-787.

WECHSLER, D. (2003). *Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV). Administration and scoring manual*. San Antonio, TX: The Psychological Corporation (It. ed.: A. Orsini, L. Pezzuti & L. Picone. 2012, Firenze: Giunti OS Organizzazioni Speciali).

WECHSLER, D. (2008). *WISC-IV technical and interpretive manual*. San Antonio, TX: The Psychological Association.

ZAYAT, M., KALB, L. & WODKA, E.L. (2011). Brief report: Performance pattern differences between children with autism spectrum disorders and attention deficit-hyperactivity disorder on measures of verbal intelligence. *Journal of Autism and Developmental Disorders*, *41* (12), 1743-1747.

DOI: 10.26387/bpa.287.5

Mindful compassion for perfectionism in personality disorders: A pilot acceptability and feasibility study

Simone Cheli[1,2], Veronica Cavalletti[2], Gordon L. Flett[3], Paul L. Hewitt[4]

[1] School of Human Health Sciences, University of Florence
[2] Center for Psychology & Health, Tages Charity, Florence
[3] LaMarsh Centre for Child and Youth Research, York University, Toronto
[4] Perfectionism and Psychopathology Lab, University of British Columbia, Vancouver

simone.cheli@tagesonlus.org

ABSTRACT. Il perfezionismo è considerato uno stile di personalità multidimensionale e transdiagnostico che può presentarsi all'interno di forme gravi di disturbi di personalità. Lo scopo di questo studio è quello di esplorare l'accettabilità e la fattibilità di un intervento integrato di gruppo di mindful compassion per pazienti a cui sono stati diagnosticati dei disturbi della personalità e presentano prominenti tratti perfezionistici. Abbiamo testato per la prima volta un intervento di gruppo di 8 settimane in un campione di pazienti (n = 5) che avevano completato una psicoterapia individuale. L'intervento ha integrato pratiche di mindful compassion con il modello di disconnessione sociale del perfezionismo. Gli outcome primari riguardavano l'accessibilità e la fattibilità dell'intervento, valutato con misure quantitative e qualitative. Gli outcome secondari corrispondevano alle differenze tra pre- e post-assessment nei tratti perfezionistici e autrocritici. Non si sono verificati né eventi avversi né drop-out. Tutti i partecipanti hanno evidenziata un'elevata accettabilità dell'intervento e risultati positivi in termini di sviluppo di nuove strategie adattative. Nel post-assessment, le dimensioni perfezionistiche che presentavano i punteggi più alti nel pre-assessment hanno mostrato un cambiamento significativo al *Reliable Change Index*. Lo studio mette in evidenza l'accettabilità e la fattibilità dell'intervento proposto. Sono necessarie ulteriori ricerche per confermare l'efficacia dell'intervento.

SUMMARY. Perfectionism is considered a multidimensional and transdiagnostic personality style that can occur in severe forms of personality disorders. The aim of this study is to explore the acceptability and the feasibility of an integrative mindful compassion group intervention for patients who were diagnosed with personality disorders and reported prominent perfectionistic traits. We pilot-tested an 8-week group intervention in a sample of patients (n = 5) who had completed individual psychotherapy. The intervention integrated mindful compassion practices with the perfectionism social disconnection model. Primary outcomes were the accessibility and feasibility of the intervention as evaluated through quantitative and qualitative measures. Secondary outcomes were differences between pre- and post-assessment in perfectionism traits and self-criticism. Neither adverse events nor drop-outs were reported. All the participants confirmed high acceptability and positive outcomes in terms of developing new healthy strategies. At post-assessment, the perfectionist dimensions that had shown the highest scores at the pre-assessment exhibited a reliable change. The study highlights the acceptability and feasibility of the proposed intervention. Further researches are needed in order to confirm the suitability of the intervention.

Keywords: *Compassion, Mindfulness, Perfectionism, Personality disorder, Self-criticism*

INTRODUCTION

Perfectionism is reputed to be a "multifaceted and multilevel personality style" (Hewitt, Flett & Mikail, 2017, p. 25) or personality disposition (Stoeber, 2017) that is characterized by requiring perfection of the self and/or others and by an overly critical stance in evaluating one's or others' behavior characterized by an inner dialogue of self-disparagement. Different models have been proposed, most of which hypothesize its causal role in the development of diverse forms of severe psychopathology (Bardone-Cone et al., 2007; Egan, Wade & Shafran, 2011; Smith et al., 2018). Theoretical formulations and collected evidence suggest that perfectionism may also play a role in personality disorders (PDs). Several studies have reported how perfectionism is associated with PDs traits in both clinical and non-clinical samples (Dimaggio, Semerari, Carcione, Nicolò & Procacci, 2015; Hewitt & Flett, 1991, 1993; Hewitt, Flett & Turnbull, 1992; Sherry, Hewitt, Flett, Lee-Baggley & Hall, 2007). Such evidences support the need for a multidimensional perspective on perfectionism as a comprehensive style driving core elements of psychopathology, regardless of an inhibited or dysregulated pattern of personality (Ayearst, Flett & Hewitt, 2012). Indeed, perfectionism may represent, for example, a core factor for either obsessive-compulsive personality disorder (Goodwin, Haycraft, Willis & Meyer, 2011) or borderline personality disorder (Chen, Hewitt, Flett & Roxborough, 2019).

Hewitt and colleagues (Hewitt et al., 2017) have proposed a Comprehensive Model of Perfectionistic Behavior (CMPB) that is rooted in the theoretical, clinical and experimental evidence in favor of a multidimensional perspective The CMPB has progressively integrated the accumulating data about perfectionism, and includes three elements: (i) the trait components or trait dimensions (self-oriented, other-oriented, socially prescribed perfectionism); (ii) the interpersonal components or self-presentational facets (perfectionistic self-promotion, nondisplay of imperfections, nondisclosure of imperfections); (iii) the intrapersonal or self-relational components or automatic perfectionistic cognitions. The complex CMPB may be effectively integrated with the foundational assumptions of PD as an enduring pattern of inner experience and behavior that is manifested in different areas such as cognition, affectivity, impulse control and interpersonal functioning (American Psychiatric Association, 2013). Moreover, its focus on a comprehensive view of self- and interpersonal- functioning is theoretically and experimentally consistent with the emergence of alternative models of personality disorder (Hopwood, 2018; Widiger et al., 2019). Perfectionism turns out to be a transdiagnostic dimension recurring among different categorical or trait-oriented diagnoses of PDs (Ayearst et al., 2012).

From Hewitt and colleagues' perspective, "perfectionism is an interpersonal personality style that develops within a relational context" (Hewitt et al., 2017, p. 99). In outlining a specifically designed clinical conceptualization and treatment, they proposed a model of the development of perfectionism (PSDM; Perfectionism Social Disconnection Model). PSDM extensively describes how persons may construe specific internal working models of others and self, leading to perfectionistic behaviors, traits and cognitions, and diverse forms of psychopathology. And by doing so, perfectionistic persons tend to express, since the early experiences as children, specific affects (shame, aloneness, depressive states, anger, etc.) that seem to relate to a recurring and profound experience of vulnerability. In particular, shame has been comprehensively studied in the literature on perfectionism, highlighting an overwhelming sense of humiliation and mortification (Stolorow, 2010), also expressed by self-criticism and a sort of attack against one's self (Gilbert & Andrews, 1998).

When treating PDs, perfectionism is frequently a maintaining factor and poses a potential risk for relapse (Cheli, MacBeth, Popolo & Dimaggio, 2020; Dimaggio et al., 2018). The overcontrolled and often pro-social characteristics associated with perfectionism may be reinforced by the environment and pursued by the patients themselves as desirable qualities (Lynch, Hempel & Dunkley, 2015). The recurrence of painful cyclical relational patterns (CRPs) that "emerges in response to aversive affective states arising from unmet attachment needs" (Hewitt et al., 2017, p.161) may result in a PD shaped by an overwhelming shame-based self-criticism linked to automatic ruminative self-statements regarding the attainment of perfection. The rationale of the present pilot-study relies upon the hypothesis that an integrated treatment approach (Livesley, 2012) to PDs characterized by a perfectionistic style may benefit from a module aimed at reducing the affective state of shame-based self-criticism. First, shame regulation has been proven to be a significant predictor of personality pathology (Schoenleber & Berenbaum, 2012), and, broadly speaking, of human

suffering (DeYoung, 2015; Gilbert & Andrews, 1998). Second, self-criticism turns out to be a significant mediator between shame and psychopathology, even if when compared with rumination (Cheung, Gilbert & Irons, 2004; Pinto-Gouveia, Castilho, Matos & Xavier, 2013). Third, an integrative treatment such as Compassion Focused Therapy (CFT) has tested interventions aimed at supporting patients in distinguish between a shame-based self-criticism and a self-compassionate correction in addition to other interventions (Gilbert, 2009; Gilbert & Procter, 2006). Fourth, shame-based criticism has proven to activate a threat system that, in turn, induces a physiological cascade inhibiting the reflective functioning and hyperactivating the defensive responses (Petrocchi & Cheli, 2019). In our clinical experience, we have tested how this vicious cycle may represent for many PDs not only a maintaining factor, but also a trigger for relapse after having concluded the individual intervention.

Hewitt et al. (2017) have conceptualized self-criticism as reflection of the self-relational component of the CMPB. That is, the internal dialogue that has with oneself, in this case a dialogue fraught with perfectionistic and highly self-disparaging themes, reflects the relationship one has with oneself (Hewitt, Mikail, Dang, Kealy & Flett, 2020). In the dynamic relational treatment developed by Hewitt et al., an important focus is on this relationship with self to help the person begin to develop self-compassion for the self and to develop the ability to self-soothe. This focus is addressed within the process and unfolding of the therapy and not structured as a specific intervention.

Thus, we outlined a pilot-study aimed at exploring the acceptability and the feasibility of an intervention aimed at consolidating the achieved changes and preventing relapses in perfectionistic patients diagnosed with and treated for PD. By considering the aforementioned hypotheses and assumptions, we developed an integrative group intervention based on both the conceptualization of perfectionism in CMPB and PSDM (Hewitt et al., 2017), and the CFT practices for promoting a self-compassionate enhancement (Gilbert & Choden, 2014). Diverse evidences may support this attempt. On the one hand, perfectionistic patients can benefit from mindfulness-based intervention, even if they might have problems in implementing these kind of practices (Flett, Nepon, Hewitt & Rose, 2020). On the other hand, the use of mindful compassion practices, as specifically focused on shame-based criticism, has reported significant evidences in favor of their application both on PDs and as integrative interventions (Kirby et al., 2017). CFT hypothesizes that the soothing system, a mammalian affect regulation system normally triggered by cues of social safeness, is poorly accessible in people whose threat system is hyperactivated by shame-based self-criticism. Therefore, the primary aim of CFT is to increase compassion for one's own distress, as a way to strengthen the ability to generate self-soothing responses to one's own suffering. We outlined a mindful compassion group intervention integrating CMPB and PSDM as core components of both a shared conceptualization of perfectionism with the participants and a few specifically designed practices.

METHODS

Sample

Five consecutive patients who were diagnosed with a PD (American Psychiatric Association, 2013) were recruited in the study after having provided informed consent. The ethical approval was given by the Ethical Committee of the Center for Psychology and Health Tages Charity (Ref. No. 01-2017/070120). Patients were eligible if: (i) they were diagnosed with a PD in last 7 months in accordance with the *Structured Clinical Interview for DSM-5 Personality Disorders* (First, Williams, Benjamin & Spitzer, 2016); (ii) they have concluded in the last month an individual *Metacognitive Interpersonal Therapy* (*TMI*; Dimaggio et al., 2007) reporting a remission from PD; (iii) they were reporting significant levels of perfectionism (equal to or higher than the mean of the normative clinical sample) in at least one scale of the *Multidimensional Perfectionism Scale* (Hewitt & Flett, 2004; Hewitt, Flett, Turnbull-Donovan & Mikail, 1991). The male to female ratio was 3:2, ages ranged between 23 to 37 (see Table 1). At the beginning of the individual psychotherapy one patient had been diagnosed with narcissist personality disorder (NPD), two with borderline personality disorder (BPD), and two with obsessive-compulsive personality disorder (OCPD).

Measures

– *Structured Clinical Interview for DSM-5 Personality Disorders* (S*CID-5-PD*; First et al., 2016): the SCID-5-PD is a semi-structured diagnostic interview for PDs as

Table 1 – Descriptives of the sample at pre-assessment

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Sample
Diagnosis	NPD	BPD	BPD	OCPD	OCPD	–
Sex	M	F	M	M	F	–
Education	College	College	Graduation	PhD	Graduation	–
Occupation	Student	Self-employed	Self-employed	Researcher	Self-employed	–
Relationship status	Single	Single	Stable relationship	Single	Married	–
Age	26	35	31	37	32	32.2 (4.2)
MPS-SO	65	60	60	65	61	62.2 (2.6)
MPS-OO	52	56	52	57	56	54.6 (2.4)
MPS-SP	54	50	56	51	50	52.2 (2.7)
FFMQ-O	33	30	26	30	32	30.2 (2.7)
FFMQ-D	32	32	32	32	33	32.2 (.4)
FFMQ-AA	24	22	26	24	30	25.2 (3.0)
FFMQ-NJ	28	30	32	35	29	30.8 (2.8)
FFMQ-NR	24	20	22	24	26	23.2 (2.3)
FSCRS-HS	2	2	1	1	1	1.4 (.5)
FSCRS-IS	13	10	11	12	11	11.4 (1.1)
FSCRS-RS	22	18	23	26	24	22.6 (3.0)

Legenda. MPS-SO = Multidimensional Perfectionism Scale – Self-oriented; MPS-OO = Multidimensional Perfectionism Scale – Other-oriented; MPS-SP = Multidimensional Perfectionism Scale – Socially prescribed; FFMQ-O = Five Facets Mindfulness Questionnaire – Observe; FFMQ-D = Five Facets Mindfulness Questionnaire – Describe; FFMQ-AA = Five Facets Mindfulness Questionnaire – Act with awareness; FFMQ-NJ = Five Facets Mindfulness Questionnaire – Non-judge; FFMQ-NR = Five Facets Mindfulness Questionnaire – Non-react; FSCRS-HS = Forms of Self-criticizing/Attacking and Self-reassuring Scale – Hated-self; FSCRS-IS = Forms of Self-criticizing/Attacking and Self-reassuring Scale – Inadequate-self; FSCRS-RS = Forms of Self-criticizing/Attacking and Self-reassuring Scale – Reassured-self.

Note. We report socio-demographic data and quantitative measures at pre-assessment for each patient. In the last column we score the Mean and Standard deviation (between parentheses) for the whole sample. Diagnosis of personality disorder refers to SCID-5-PD interview at the beginning of individual psychotherapy: borderline personality disorder (BPD); narcissistic personality disorder (NPD); obsessive-compulsive personality disorder (OCPD).

defined by the DSM-5. The procedure allows the clinician to capture the construct embodied in the diagnostic criteria of the 10 PDs. The SCID-5-PD reports good inter-rater reliability at both dimensional and categorical PD diagnoses.

- *Multidimensional Perfectionism Scale* (*MPS*; Hewitt et al., 1991): MPS is a 45-item measure on a 1-to-7 Likert scale designed to measure three dimensions of perfectionistic behavior: self-oriented perfectionism (MPS-SO), other-oriented perfectionism (MPS-OO), and socially prescribed perfectionism (MPS-SP). Higher scores indicate a greater level of perfectionism. Cronbach's alpha ranges from .79 to .89 for the three subscales, test-retest reliabilities range from .75 to .80 over 3 months, and subscale intercorrelations range from .25 to .40.
- *Five Facets Mindfulness Questionnaire* (*FFMQ*; Baer, Smith, Hopkins, Krietemeyer & Toney, 2006): the FFMQ is a 39-item questionnaire that measures five facets of mindfulness: observe (FFMQ-O), describe (FFMQ-D), act with awareness (FFMQ-AA), non-judge (FFMQ-NJ), and non-react (FFMQ-NR). Items were scored on a five-point Likert scale ranging from 1 to 5 and computed by summing the scores on the individual items, with higher scores indicating greater mindfulness. Cronbach's alpha ranges from .75 to .91 for the three subscales.
- *Forms of Self-criticizing/Attacking and Self-reassuring Scale* (*FSCRS*; Gilbert, Clarke, Hempel, Miles & Irons, 2004): the FSCRS is a 22 - item measure, which requires participants to rate a selection of positive and negative statements on a 5-point Likert scale ranging from 0 to 4. The scale measures self-reassurance (reassured-self; FSCRS-RS) and two types of self-criticism: inadequate-self (FSCRS-IS) and hated-self (FSCRS-HS), summing the scores of individual items. Cronbach's alpha ranges from .86 to .90 for the three subscales.
- *Written Open Questions* (*WOQ*): two written open questions were included in the assessment. During the initial assessment participants were asked to describe their expectation and desired goals before starting the intervention (WOQ-1), whereas the final assessment includes a question about how they evaluated the intervention and its effectiveness in respect to previously defined goals (WOQ-2).
- *Semi-structured Interview* (*SSI*): one month after the end of the intervention all the patients accessed a semi-structured interview. SSI included open questions about: SSI-1, how they generally evaluated the intervention; SSI-2, how they evaluated its general effectiveness in respect to their own goals; SSI-3, how they evaluated its specific effectiveness in respect to perfectionism; SSI-4, how they perceived the mindful compassion practices; SSI-5, how they perceived the shared conceptualization of perfectionism; SSI-6, how they evaluated the group format in respect to the individual one.

Procedures

Once the patients signed the informant consent form, they accessed the initial assessment (t_0), comprising psychometric measures (MPS; FFMQ; FSCRS), clinical interview (SCID-5-PD) and qualitative measures (WOQ-1). After having completed the 8-week intervention (t_1), patients completed the final assessment (MPS; FFMQ; FSCRS; WOQ-2). One month after the final assessment (t_2) all the patients were interviewed (SCID-5-PD; SSI). The primary outcome of the study was the acceptability of the intervention defined on the base of the following criteria: (i) no adverse events (e.g. self-harm behavior, suicidal ideation, etc.); (ii) maintenance of PD remission at t_2; (iii) rate of drop-out (≤10%); (iv) rate of attendance to sessions (no more than 1 session skipped for each participant); (v) rate of positive evaluation at the qualitative measures (WOQ; SSI) by participants (≥80%). The secondary outcomes were: (i) an individual reduction ($t_{0\text{-}1}$) of perfectionism and self-criticism (MPS and FSCRS scale with the higher score for each participant); (ii) a group reduction ($t_{0\text{-}1}$) of perfectionism and self-criticism (MPS and FSCRS scales); (iii) an individual increase ($t_{0\text{-}1}$) of mindfulness (FFMQ scales scores for each participant); (iv) a group increase ($t_{0\text{-}1}$) of mindfulness (FFMQ scales scores).

The group intervention involved two therapists, both with 5 years of experience in CFT. The protocol included specific schedules (e.g. psychoeducation; practices; homework; workbook; etc.) for all the sessions, and treatment adherence was evaluated at the end of each session. Moreover, specific slots for each therapist and session were defined, as a way to always have an observer of participants' engagement and therapist's adherence. All the information collected by the therapists were finally integrated with the qualitative measures reported by the patients (WOQ; SSI).

Analysis

We report the descriptives of the clinical measures. Pre-post changes in individual scores were evaluated trough the Reliable Change Index (RCI; Jacobson & Truax, 1991). Reliable changes (RCI≥1.96) were scored by using the normative data of non-clinical samples, since subjects were recruited after having concluded an individual psychotherapy and reported remission from PD's diagnosis. Pre-post changes in the sample (n = 5) were investigated through Student's t test, despite the low sample size (deWinter, 2013). Qualitative measures (i.e. the written answers at WOQ and the transcripts of SSI) were explored through hermeneutic phenomenological methods (Rennie, 2012).

Intervention

The intervention was an integrative mindful compassion group therapy (see Table 2). The structure was outlined on the base of standard mindfulness-based interventions (MBIs), comprising eight 2-hour group sessions and one day of silence lasting 4 hours (Didonna, 2009). The contents and the phases of the intervention were rooted in two different frameworks. On the one hand, the sequence of and the types of practices were defined in accordance with the mindful compassion protocol (Gilbert & Choden, 2014). On the other hand, the shared conceptualization of perfectionism and its role in triggering, maintaining, and inducing relapses in PD was proposed through the CMPB and PSDM (Hewitt et al., 2017). We also included specifically designed mindful compassion practices using Hewitt and colleagues (2017) dynamic relational approach. For example, we co-construed with the participants an individual Cyclical Relational Pattern that was used as the object of a compassionate enquiry practice (see Hewitt, Mikail, Flett & Dang, 2018).

The protocol was organized in four phases, similarly to group psychotherapy of perfectionistic behavior, in order to highlight the specific focuses of each of them and "the fluid yet predictable nature of group development " (Hewitt et al., 2017, p. 259). In Table 2 we report the focuses, the mindful compassion practices and the shared conceptualization of perfectionism we proposed in the four phases. Every single session was organized according to the classic MBIs' format: the therapists ask about previous week and practices and briefly discuss with participants; they introduce the focus of the session and share specific workbooks including shared conceptualization and the proposed practices; participants experiment practices and share feedbacks with therapists; the therapists conclude the session by anticipating next focus session and propose practices as homework.

Table 2 – Structure of the intervention

	Focus of the phase	Practices	Shared conceptualization
1st phase – week 1-2	Awareness	Mindfulness practices	Automatic ruminative cognitions
2nd phase – week 3-4	Acceptance	Working with acceptance	Perfectionism trait dimensions
3rd phase – week 5-6	Fear of compassion	Working with imagery	Perfectionistic self-presentation facets
Day of silence (7th week)	Circle of compassion	Impermanence and widening compassion	Group cohesiveness and termination
4th phase – week 7-8	Compassionate self	Compassion for self and others	Cyclical relational pattern

In proposing the practices and sharing the feedback there was a substantial difference compared to the MBI protocols: therapists never forced the participants to either use specific postures during meditation and or consider the homework a mandatory request. Such an approach was rooted in CFT attempt of helping patients activate the soothing system (Gilbert, 2009) and, at the same time, in defusing the recurring perfectionistic CRPs (Hewitt et al., 2017).

RESULTS

Course of the intervention

The course of the intervention was seemingly shaped by the outlined four phases and a few specific critical incidents. All the participants reported a good engagement and curiosity at the beginning of the first session. They took a collaborative approach to the rules and norms proposed by the therapists, despite the first critical incident corresponded to the first attempt to share an explicit conceptualization of perfectionism. Patient 4 let emerge a criticism, by remarking that he considered perfectionism to be a positive factor rather than negative. This incident was elaborated in terms of a hypothesis to be test, and by highlighting the complexity and heterogeneity of the construct of trait. The group cohesiveness and bond with therapists progressively increased allowing the resolution of the second critical incident. Patient 1 reported an extremely disturbing relational event that occurred between the second and the third session. In telling this event (happened external to the group), Patient 1 started crying and reporting how frequently he was self-critical, albeit covertly. The group proactively and emotionally responded, by supporting and reassuring Patient 1. They all stated how they were not considering how painful could have been his self-criticism, and how sharing this pain was judged an act of trust in the group. The third critical incident was a group one. When discussing about their recurrent CRPs (session 6) everyone agreed on how dysfunctional and painful their patterns were. This event and the proposed practices let emerge a deep shared conceptualization of their perfectionistic personality styles. The fourth event corresponded to the day of silence. On the one hand, they experienced the emotional impact of a long group session where they agreed on an emerging significant bond between each other. On the other hand, they started thinking about the approaching termination of the intervention. The last incident must be considered the last session, in terms of the final, albeit critical, step in elaborating the end of the group, and the individual and shared meaning of their relational experience.

Quantitative outcomes

No patient reported any adverse events or drop-outs during the intervention. Moreover, all the patients maintained the remission of PD at the one-month follow-up (t_2) and the rate of attendance was as expected (no one skipped a session). In respect to secondary outcomes (see Table 3), all the participants reported a reliable change (RCI ranging between 1.97 and 3.15) at MPS and FSCRS scale with the higher score at pre-assessment, except for Participant 4 at FSRCS. A significant difference (p<.05) was found in two (MPS-SO; MPS-OO) out of three of MPS scales, whereas the remaining scale reported a fringe value (p = .056). No one difference was found in FSRCS scales, except for FSCRS-RS (t = 3.04; p = .016; df = 8). All the patients reported a reliable change (RCI ranging between −2.05 to −3.08) at FFMQ-NR, whereas four out of five at FFMQ-NJ (RCI ranging between −1.97 to −2.75). No reliable change was found in the other FFMQ scales with only one exception (Patient 1 showed a reliable change at FFMQ-O). Similarly, we found no differences in the whole sample between pre-and post- assessment at all the FFMQ scales.

Qualitative outcomes

All participants reported positive feedback about the intervention at WOQ-2 and SSI. They also highlighted to have achieved their desired goals at WOQ-1 (mainly expressed in terms of better knowing one's self and learning new healthy strategies), and to have discovered new gains they did not considered at the beginning of the intervention. They especially highlighted the discovery of mindful practices and group experience as powerful tools in pursuing wellbeing. Another unexpected result was their deep understanding of the role of perfectionism in their daily life and how they were able to use compassion as a soothing and effective way to look at themselves. They remarked how the personal experience of practices and the relational experience of a group intervention were accelerators in expanding what they previously learned during the individual therapy.

Table 3 – Quantitative measures over time

	Reliable change index					Student's t	
	Patient 1	*Patient 2*	*Patient 3*	*Patient 4*	*Patient 5*	*t*	*p*
MPS-SO	.38	2.05*	.9	.13	2.05*	3.405	.009*
MPS-OO	2.1	1.23	2.34*	.99	.12	4.178	.003*
MPS-SP	.86	1.64	3.15*	0	1.26	2.227	.056
FFMQ-O	−2.05*	−1.17	−.58	−.29	0	1.519	.167
FFMQ-D	.29	−.86	0	−.29	−.86	1.633	.141
FFMQ-AA	.68	−.86	−.34	−1.36	−.68	.662	.526
FFMQ-NJ	−2.36*	−2.75*	−1.97*	−.39	−1.97*	1.633	.141
FFMQ-NR	−3.08*	−3.08*	−2.05*	−2.05*	−2.05*	.323	.754
FSCRS-HS	1.65	1.65	0	0	0	1.28	.236
FSCRS-IS	2.12*	3.71*	2.12*	.53	2.38*	3.04	.016*
FSCRS-RS	−.32	−.64	−.64	0	−.64	.675	.518

Legenda. MPS-SO = Multidimensional Perfectionism Scale – Self-oriented; MPS-OO = Multidimensional Perfectionism Scale – Other-oriented; MPS-SP = Multidimensional Perfectionism Scale – Socially prescribed; FFMQ-O = Five Facets Mindfulness Questionnaire – Observe; FFMQ-D = Five Facets Mindfulness Questionnaire – Describe; FFMQ-AA = Five Facets Mindfulness Questionnaire – Act with awareness; FFMQ-NJ = Five Facets Mindfulness Questionnaire – Non-judge; FFMQ-NR = Five Facets Mindfulness Questionnaire – Non-react; FSCRS-HS = Forms of Self-criticizing/Attacking and Self-reassuring Scale – Hated-self; FSCRS-IS = Forms of Self-criticizing/Attacking and Self-reassuring Scale – Inadequate-self; FSCRS-RS = Forms of Self-criticizing/Attacking and Self-reassuring Scale – Reassured-self.

Note. We report the reliable change index (RCI) for each patient and the Student's *t* for the whole sample (last two columns). The reliable changes (RCI≥1.96) and the significant differences ($p<.01$) between pre- and post- assessment are highlighted (*).

DISCUSSION

The aim of this study was to pilot-test the acceptability and the feasibility of a mindful compassion group intervention for PD's patient with predominant perfectionistic traits. The proposed protocol was rooted in CFT and integrated CMPB and PSDM as the core components of both a shared conceptualization of perfectionism with participants and a few specifically designed practices. The reported results seem to confirm both the acceptability and the feasibility, also highlighting promising evidences in favor of a potential effectiveness in reducing perfectionism and self-criticism and increasing mindfulness.

All the defined primary outcomes were achieved. Neither adverse event nor drop-out were reported. All the participants maintained at 1-month follow-up the remission of PD gained at the end of the individual psychotherapy. Moreover, no one skipped a session and the rate of positive evaluation was 100%. In respect to secondary outcomes, we report contrasting results. On the one hand, all the participants, with only one exception, reported a reliable change at MPS and FSCRS scale with the higher score at pre-assessment, and two out of three of the MPS scales showed a significant change in the whole group. Of note, the remaining MPS scale highlighted an almost significant value (p = .056), and the Participant 4 not reporting a reliable change at FSCRS highlighted at pre-assessment values at least one SD below the sample means. On the other hand, only one scale of FFMQ (FFMQ-NR) reported reliable changes among all the participants, whereas FFMQ-NJ in four out of five. Again, Patient 4 was the one not reporting a reliable change, and who highlighted an FFMQ-NJ value at pre-assessment one SD above the sample mean.

The results and their biases may be interpreted through four intertwined hypotheses. First, the intervention seems to acceptable and feasible, despite a few relevant limitations. Indeed, the low sample size and the observational methodology urge us not to generalize the results and to carefully consider the possible implications at both theoretical and clinical level. Second, the inclusion criteria may represent a nuanced bias due to a sample of remitted patients. We might have expected dubious results on clinical efficacy due to the previous therapeutic gains. At the same time, the good acceptability might be overestimated by a previous successful psychotherapy. Third, the limited recurrence of reliable changes, in both the single participants and in the sample as a whole, may be similarly underestimated by the recruitment of remitted patients. Once the intervention will be tested on newly diagnosed patients, its effectiveness may further increase. Finally, the complexity of personality trajectories urges us to consider the differential effect of a group intervention in respect to the phases of such trajectories. Our preliminary results can just support the feasibility of testing our protocol in different populations and phases of a wider treatment program.

All that said, the collected measures seem to support the need for further researches aimed at exploring the effectiveness of a mindful compassion group intervention that may integrate CMPB and PSDM. The intervention was effective in reducing the most problematic dimension of perfectionism in each participant, and of self-criticism in four out of five participants. Moreover, they reported a significant increase in the facets of mindfulness connected to the abilities of nonjudging and nonreacting to experience. We may hypothesize that the intervention was able to promote a different look at personal and relational life, in a non-judgmental and compassionate manner. The use of mindful compassion practices might be extremely useful in "discovering more adaptive and flexible ways of meeting the need for security, connection, and self-regard" (Hewitt et al., 2017, p. 150). Indeed, the perniciousness of perfectionistic traits seem to ask for a complex and integrate treatment approach (Livesley, 2012). An approach that can remind our patients how social disconnection may be the painful illusion that arises from not knowing how to balance compassion for us with compassion for others.

References

AMERICAN PSYCHIATRIC ASSOCIATION (2013). DSM-5 Diagnostic Classification. In *Diagnostic and Statistical Manual of Mental Disorders*. doi.org/10.1176/appi.books.9780890425596.x00diagnosticclassification

AYEARST, L.E., FLETT, G.L. & HEWITT, P.L. (2012). Where is multidimensional perfectionism in DSM-5? A question posed to the DSM-5 personality and personality disorders work group. *Personality Disorders: Theory, Research, and Treatment*, *3* (4), 458-469. doi.org/10.1037/a0026354

BAER, R.A., SMITH, G.T., HOPKINS, J., KRIETEMEYER, J. & TONEY, L. (2006). Using self-report assessment methods to explore facets of mindfulness. *Assessment*, *13* (1), 27-45. doi.org/10.1177/1073191105283504

BARDONE-CONE, A.M., WONDERLICH, S.A., FROST, R.O., BULIK, C.M., MITCHELL, J.E., UPPALA, S. & SIMONICH, H. (2007). Perfectionism and eating disorders: Current status and future directions. *Clinical Psychology Review*, *27* (3), 384-405. doi.org/10.1016/j.cpr.2006.12.005

CHELI, S., MacBETH, A., POPOLO, R. & DIMAGGIO, G. (2020). The intertwined path of perfectionism and self-criticism in a client with obsessive-compulsive personality disorder. *J Clin Psychol*, *76*, 2055-2066. doi.org/10.1002/jclp.23051

CHEN, C., HEWITT, P.L., FLETT, G.L. & ROXBOROUGH, H.M. (2019). Multidimensional perfectionism and borderline personality organization in emerging adults: A two-wave longitudinal study. *Personality and Individual Differences*, *146*, 143-148. doi.org/10.1016/j.paid.2019.04.011

CHEUNG, M.S.-P., GILBERT, P. & IRONS, C. (2004). An exploration of shame, social rank and rumination in relation to depression. *Personality and Individual Differences*, *36* (5), 1143-1153. doi.org/10.1016/S0191-8869(03)00206-X

De WINTER, J. (2013). Using the Student's t-test with extremely small sample sizes. *Practical Assessment, Research & Evaluation*, *18*. doi.org/10.7275/e4r6-dj05

DeYOUNG, P.A. (2015). *Understanding and treating chronic shame*. Routledge.

DIDONNA, F. (2009). *Clinical handbook of mindfulness*. Springer.

DIMAGGIO, G., LYSAKER, P.H., CALARCO, T., PEDONE, R., MARSIGLI, N., RICCARDI, I., SABATELLI, B., CARCIONE, A. & PAVIGLIANITI, A. (2015). Perfectionism and personality disorders as predictors of symptoms and interpersonal problems. *American Journal of Psychotherapy*, *69* (3), 317-330. doi.org/10.1176/appi.psychotherapy.2015.69.3.317

DIMAGGIO, G., MacBETH, A., POPOLO, R., SALVATORE, G., PERRINI, F., RAOUNA, A., OSAM, C.S., BUONOCORE, L., BANDIERA, A. & MONTANO, A. (2018). The problem of overcontrol: Perfectionism, emotional inhibition, and personality disorders. *Comprehensive Psychiatry*, *83*, 71-78. doi.org/10.1016/j.comppsych.2018.03.005

DIMAGGIO, G., SEMERARI, A., CARCIONE, A., NICOLÒ, G. & PROCACCI, M. (2007). *Psychotherapy of personality disorders: Metacognition, states of mind and interpersonal cycles*. Routledge.

EGAN, S.J., WADE, T.D. & SHAFRAN, R. (2011). Perfectionism as a transdiagnostic process: A clinical review. *Clinical Psychology Review*, *31* (2), 3-12. doi.org/10.1016/j.cpr.2010.04.009

FIRST, M.B., WILLIAMS, J.B., BENJAMIN, L.S. & SPITZER, R.L. (2016). *SCID-5-PD: Structured clinical interview for DSM-5 personality disorders*. American Psychiatric Association Publishing.

FLETT, G.L., NEPON, T., HEWITT, P.L. & ROSE, A.L. (2020). Why perfectionism is antithetical to mindfulness: A conceptual and empirical analysis and consideration of treatment implications. *International Journal of Mental Health and Addiction*. doi.org/10.1007/s11469-020-00252-w

GILBERT, P. (2009). *The compassioate mind*. New Harbinger Publications.

GILBERT, P. & ANDREWS, B. (1998). *Shame: Interpersonal behavior, psychopathology, and culture*. Oxford University Press.

GILBERT, P. & CHODEN, K. (2014). *Mindful compassion: How the science of compassion can help you understand your emotions, live in the present, and connect deeply with others*. New Harbinger Publications.

GILBERT, P., CLARKE, M., HEMPEL, S., MILES, J.N.V. & IRONS, C. (2004). Criticizing and reassuring oneself: An exploration of forms, styles and reasons in female students. *British Journal of Clinical Psychology*, *43* (1), 31-50. doi.org/10.1348/014466504772812959

GILBERT, P. & PROCTER, S. (2006). Compassionate mind training for people with high shame and self-criticism: Overview and pilot study of a group therapy approach. *Clinical Psychology and Psychotherapy*, *13* (6), 353-379. doi.org/10.1002/cpp.507

GOODWIN, H., HAYCRAFT, E., WILLIS, A.M. & MEYER, C. (2011). Compulsive exercise: The role of personality, psychological morbidity, and disordered eating. *International Journal of Eating Disorders*, *44* (7), 655-660. doi.org/10.1002/eat.20902

HEWITT, P.L. & FLETT, G.L. (1991). Dimensions of perfectionism in unipolar depression. *Journal of Abnormal Psychology*, *100*, 98-101. psycnet.apa.org/doi/10.1037/0021-843X.100.1.98

HEWITT, P.L. & FLETT, G.L. (1993). Dimensions of perfectionism, daily stress, and depression: A test of the specific vulnerability hypothesis. *Journal of Abnormal Psychology*, *102*, 58-65.

HEWITT, P.L. & FLETT, G.L. (2004). *Multidimensional Perfectionism Scale (MPS)*: Technical manual. Toronto, Canada: Multi-Health Systems.

HEWITT, P.L., FLETT, G.L. & MIKAIL, S.F. (2017). *Perfectionism: A relational approach to conceptualization, assessment, and treatment.* New York:The Guilford Press.

HEWITT, P.L., FLETT, G.L. & TURNBULL, W. (1992). Perfectionism and multiphasic personality inventory (MMPI) indices of personality disorder. *Journal of Psychopathology and Behavioral Assessment*, *14* (4), 323-335. doi.org/10.1007/BF00960777

HEWITT, P.L., FLETT, G.L., TURNBULL-DONOVAN, W. & MIKAIL, S.F. (1991). The Multidimensional Perfectionism Scale: Reliability, validity, and psychometric properties in psychiatric samples. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *3* (3), 464. doi.org/10.1037/1040-3590.3.3.464

HEWITT, P.L., MIKAIL, S.F., DANG, S.S., KEALY, D. & FLETT, G.L. (2020). Dynamic-relational treatment of perfectionism: An illustrative case study. *J Clin Psychol*, *76*, 2028-2040. doi.org/10.1002/jclp.23040

HEWITT, P.L., MIKAIL, S.F., FLETT, G.L. & DANG, S. (2018). Specific formulation feedback in dynamic-relational group psychotherapy of perfectionism. *Psychotherapy*, *55*, 179-185. doi.org/10.1556/2006.5.2016.004

HOPWOOD, C.J. (2018). A framework for treating DSM-5 alternative model for personality disorder features. *Personality and Mental Health*, *12* (2), 107-125. doi.org/10.1002/pmh.1414

JACOBSON, N.S. & TRUAX, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59* (1), 12-19. doi.org/10.1037/0022-006X.59.1.12

KIRBY, J.N., TELLEGEN, C.L. & STEINDL, S.R. (2017). A meta-analysis of compassion-based interventions: Current state of knowledge and future directions. *Behavior Therapy*, *48* (6), 778-792. doi.org/10.1016/j.beth.2017.06.003

LIVESLEY, W.J. (2012). Integrated treatment: A conceptual framework for an evidence-based approach to the treatment of personality disorder. *Journal of Personality Disorders*, *26* (1), 17-42. doi.org/10.1521/pedi.2012.26.1.17

LYNCH, T.R., HEMPEL, R.J. & DUNKLEY, C. (2015). Radically open-dialectical behavior therapy for disorders of over-control: Signaling matters. *American Journal of Psychotherapy, 69* (2), 141-162. doi.org/10.1176/appi.psychotherapy.2015.69.2.141

PETROCCHI, N. & CHELI, S. (2019). The social brain and heart rate variability: Implications for psychotherapy. *Psychology and Psychotherapy: Theory, Research and Practice*, *92* (2), 208-223. doi.org/10.1111/papt.12224

PINTO-GOUVEIA, J., CASTILHO, P., MATOS, M. & XAVIER, A. (2013). Centrality of shame memories and psychopathology: The mediator effect of self-criticism. *Clinical Psychology: Science and Practice*, *20* (3), 323-334. doi.org/10.1111/cpsp.12044

RENNIE, D.L. (2012). Qualitative research as methodical hermeneutics. *Psychological Methods*, *17* (3), 385-398. doi.org/10.1037/a0029250

SCHOENLEBER, M. & BERENBAUM, H. (2012). Shame regulation in personality pathology. *Journal of Abnormal Psychology*, *121* (2), 433-446. doi.org/10.1037/a0025281

SHERRY, S.B., HEWITT, P.L., FLETT, G.L., LEE-BAGGLEY, D.L. & HALL, P.A. (2007). Trait perfectionism and perfectionistic self-presentation in personality pathology. *Personality and Individual Differences*, *42* (3), 477-490. doi.org/10.1016/j.paid.2006.07.026

SMITH, M.M., SHERRY, S.B., CHEN, S., SAKLOFSKE, D.H., MUSHQUASH, C., FLETT, G.L. & HEWITT, P.L. (2018). The perniciousness of perfectionism: A meta-analytic review of the perfectionism-suicide relationship. *Journal of Personality*, *86* (3), 522-542. doi.org/10.1111/jopy.12333

STOEBER, J. (2017). *The psychology of perfectionism: Theory, research, applications*. London: Routledge. doi.org/10.4324/9781315536255

STOLOROW, R.D. (2010). The shame family: An outline of the phenomenology of patterns of emotional experience that have shame at their core. *International Journal of Psychoanalytic Self Psychology*, *5* (3), 367-368. doi.org/10.1080/15551024.2010.485347

WIDIGER, T.A., BACH, B., CHMIELEWSKI, M., CLARK, L.A., DEYOUNG, C., HOPWOOD, C.J., … & THOMAS, K.M. (2019). Criterion A of the AMPD in HiTOP. *Journal of Personality Assessment*, *101* (4), 345-355. doi.org/10.1080/00223891.2018.1465431