19.00 €

vol. LXVII. September-December 2019

N° 286

BPA

3/3

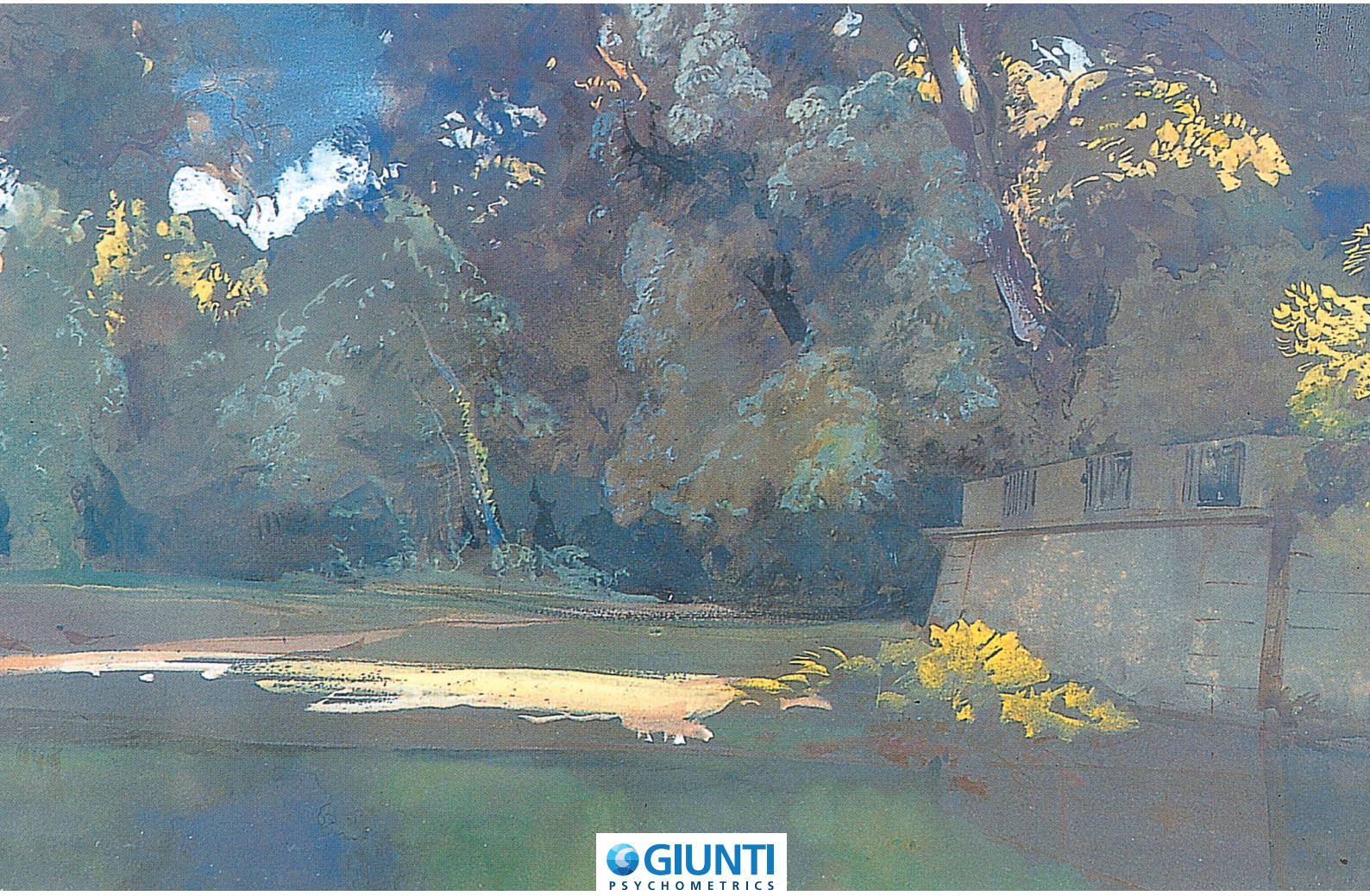
Four-monthly Journal
ISSN 0006-6761

Bollettino di Psicologia Applicata

APPLIED PSYCHOLOGY BULLETIN

Indexed in PsycINFO® – Scopus Bibliographic Database

Review | Research | Experiences & Tools



GIUNTI PSYCHOMETRICS

CONTENTS

⬢ Research

■ Review

▲ Experiences & Tools

DOI: 10.26387/bpa.286.1

Psychodynamic profiles of non-suicidal self-injury in adolescence: From the intra-psychic to the interpersonal dimension

Anna Gargiulo[1], Pasquale Dolce[2], Giorgia Margherita[1]

[1] Department of Humanistic Studies, University of Naples Federico II
[2] Department of Public Health, University of Naples Federico II

anna.gargiulo2@unina.it

ABSTRACT. L'autolesività non suicidaria (ANS), intesa come danno intenzionale che l'individuo infligge alla propria superficie corporea senza intento suicidario, è particolarmente diffusa tra gli adolescenti. La letteratura scientifica ha cercato di classificare l'ANS, focalizzandosi principalmente sulle caratteristiche sindromiche. Molti altri studi ne hanno approfondito le funzioni, tuttavia pochi hanno indagato le stesse in termini di criteri utili ad inquadrare clinicamente i comportamenti autolesionistici. Lo studio si propone di identificare possibili profili di autolesionismo, incrociando elementi relativi al quadro sindromico (ad esempio, la frequenza o la tipologia della condotta) con meccanismi latenti che ne sono alla base (funzioni psichiche), quali la regolazione emotiva e l'investimento corporeo. I soggetti coinvolti nella ricerca sono stati 108 adolescenti che avevano messo in atto condotte autolesive (*M* = 14.6, *DS* = .9; 34 maschi e 74 femmine). L'analisi del cluster degli aspetti sindromici ha mostrato due tipologie di ANS, Ripetitiva e Episodica, che sono state incrociate con le caratteristiche latenti. La distribuzione delle funzioni dell'autolesionismo ha consentito di individuare due profili principali: autolesionismo grave e autolesionismo moderato. Infine, viene proposto un uso delle funzioni come criteri utili alla base di una diagnosi di autolesionismo.

SUMMARY. Nonsuicidal self-injury (NSSI), considered deliberate and self-inflicted destruction of one's own body tissue without suicidal intent, is particularly widespread among adolescents. The literature has attempted to classify NSSI, focusing primarily on syndromal features. Many studies have addressed the functions of NSSI but very few have investigated using these functions as criteria to assess nonsuicidal self-injury behaviours. This study aimed to identify profiles for nonsuicidal self-injury behaviours, matching the manifest syndromal aspects (e.g., frequency, variety of the methods) with the latent functional ones like psychic functions, especially emotion regulation, and body investment. A sample of 108 adolescents with a history of nonsuicidal self-injury (M = 14.6, SD = .9; 34 males and 74 females) participated in the study. Cluster analysis of syndromal aspects have shown two types of NSSI, Repetitive and Episodic, which were matched to latent features. NSSI functions distribution led to the emergence of two main profiles: Serious nonsuicidal self-injury and Moderate nonsuicidal self-injury. Some reflections on the use of functions as diagnostic criteria for NSSI are discussed.

Keywords: *Nonsuicidal self-injury, Profiles, Adolescence, Psychic function*

INTRODUCTION

In recent years we have seen an increase in nonsuicidal self-injury behaviours (NSSI), defined as deliberate and self-inflicted destruction of one's own body tissue without suicidal intent and for purposes not socially or culturally sanctioned (e.g. cutting, burning and scratching the skin or hitting; Nock, 2010).

It is well-known that nonsuicidal self-injury is particularly widespread amongst the adolescent and young adult populations, increases in early adolescence and declines in late adolescence, with an average age of onset of 14 years (Brown & Plener, 2017; Cerutti, Manca, Presaghi & Gratz, 2011; Cipriano, Cella & Cotrufo, 2017; Plener, Schumacher, Munz & Groschwitz, 2015). Furthermore, females are more likely to implement the indicated behaviours than males (Bresin & Schoenleber, 2015; Gargiulo & Margherita, 2014; Valencia-Agundo, Corbet Burcher, Ezpeleta & Kramer, 2018), with gender differences linked to aetiology, trajectories and contexts (Gargiulo & Margherita, 2019; Gargiulo, Tessitore, Le Grottaglie & Margherita, 2020; Whitlock & Rodham, 2013). NSSI is particularly widespread among school populations, thus 18% of teenagers who attend school report having severely self-harmed at least once in their lifetime (Lewis & Heath, 2015; Swannell, Martin, Page, Hasking & St John, 2014). Although studies have focused on the school context, there is a need to intercept and better comprehend self-injury behaviours at school (Berger, Hasking & Reupert, 2014; Gargiulo, 2020).

Within a scientific discussion (Plener & Fegert, 2015) regarding the nomenclature and classification of the behaviour, which aimed to distinguish it on the one hand from suicide attempts and on the other from generalized self-harming behaviours, the *Diagnostic and Statistical Manual of Mental Disorder 5* proposed the definition of Nonsuicidal self-injury as a specific psychiatric diagnosis and not only as a symptom or syndrome present in different personality disorders (APA, 2013). This classification allowed for better differential diagnosis, particularly in adolescence, by preventing early stigmatization through a diagnosis of suicide or borderline personality (Gargiulo et al., 2014). In this sense, part of the literature focused on implementing studies in NSSI in adolescence following the diagnostic criteria of DSM-5, such as the number of episodes in the last year, to establish whether it is occasional or Repetitive NSSI (Madjar, Zalsmanb, Mordechaia & Shovalb, 2017; Manca, Cerutti & Presaghi, 2005; Manca, Presaghi & Cerutti, 2014; Sarno, Madeddu & Gratz, 2010; Shaffer & Jacobson, 2009).

One of the research areas that has seen the most development in recent years in the field of NSSI in adolescents is a focus on the functions (Lewis & Santor, 2010; Nock & Prinstein, 2004), which have been theorized as intrapersonal and interpersonal (Klonsky, 2007). The term intrapersonal refers to the functions aimed at changing an individual's internal state (emotions, thoughts and physical sensations), whereas the term interpersonal refers to functions that aim to alter the external setting, for example, withdrawal of demands or increased social support (Turner et al., 2012). In the vast majority of research, participants reported intrapersonal functions, mainly affect regulation (e.g. coping with negative emotions) and self-punishment (e.g. expression of anger at self) (Gratz, 2007; Klonsky 2007; Klonsky & Glenn, 2009). Interpersonal or social functions were also reported, though less frequently than intrapersonal ones. Among these, interpersonal influence (e.g. communicating internal distress to others) was the most frequently reported (Klonsky, 2007; Nock & Prinstein, 2004), followed by revenge function, getting revenge on others (Klonsky, 2007). Nonetheless, it should be noted that more than one function can be adopted and functions can vary over time.

Emotion regulation (to regulate, control or express intense and pervasive emotions, such as anger, boredom and sadness, or generally painful moods such as depression, guilt and shame) is the function of NSSI most commonly referred to by those who engage in the behaviour, in particular by adolescents (Klonsky, 2007; Laye-Gindhu & Schonert-Reichl, 2005; Madge et al., 2008; Nock & Prinstein, 2005). Self-injurious behaviour is usually preceded by negative emotions and high arousal, and followed by feelings of calm and relief (Di Pierro, Sarno, Gallucci & Madeddu, 2014; Klonsky, 2009). Therefore, emotional dis-regulation (the inability to recognize, accept, control and be aware of one's emotions) was then conceptualized as a possible risk factor of nonsuicidal self-injury in adolescence (Adrian, Zeman, Erdley, Lisa & Sim, 2011). NSSI has been related to infrequent use of effective coping strategies, lack of emotional expression, and lack of emotional clarity (Dicé, Maiello, Dolce & Freda, 2017; Gratz & Roemer, 2004; Martino et al., 2019). In this sense, a part of the literature has shown that adolescents who self-harm have difficulty putting their sufferings into words and consider it a means of communication that allows them to share emotions of anger and anguish, and which allows their families and those round them to realize how much they are suffering

(Crouch & Wright, 2004; Fortune, Sinclair & Hawton, 2008; Moyer & Nelson, 2007).

In addition, research has investigated the role of negative bodily attitudes in the relationship between emotion dysregulation and nonsuicidal self-injurious behaviour (Muehlenkamp, Bagge, Tull & Gratz, 2013). In particular, negative body regard (e.g. how one perceives, cares for, and experiences one's own body) increases the propensity for an individual to harm their body when he/she is emotionally dys-regulated (Muehlenkamp, 2012). Thus, the negative affective evaluations of the body (e.g. body dissatisfaction, lack of emotional investment in the body) are considered important risk factors for NSSI among college students (Mulay, West, Wallner Samstag & Diamond, 2017). The self-injurious behaviour is a form of attack on the body, related to hatred for one's own body; this hate, as well as body dissatisfaction and disregard, allows the subject to see their own body as an object separated from the self, making it easier to harm it; this can explain pain tolerance during the self-injury act. The self-objectification also contributes to negative body regard, increasing participation in self-harmful behaviours (Orbach, 1996).

Although many studies have investigated the functions of NSSI, few have addressed the functions of NSSI as criteria for clinical assessment (Klonsky, Glenn, Styer, Olino & Washburn, 2015; Nock & Prinstein, 2004). In this regard, the new *Psychodynamic Diagnostic Manual (PDM-2)*, which, from a psychodynamic perspective, offers a diagnosis based on the dimensions of symptom patterns as well as mental functioning and personality, suggested that clinicians should base their assessment on the history of nonsuicidal self-injury and its current manifestations, and, above all, on its functions. More specifically, nonsuicidal self-injurious behaviour has been read as a common and nonspecific psychiatric symptom found in a variety of disorders and also in adolescents without a specific psychiatric diagnosis (Lingiardi & McWilliams, 2017). Thus, there is growing evidence to suggest that NSSI functions have different implications for treatment, prognosis, and suicide risk (Klonsky & Glenn, 2009; Nock & Prinstein, 2005).

Therefore, trying to integrate these important assessment perspectives, categorical and dimensional, we believe that the diagnostic process is possible thanks to the intersection of different ways of looking at the symptom (Rossi Monti & D'Agostino, 2018). In this sense, the aim of this work was to carry forward a vision that identifies profiles of nonsuicidal self-injury among adolescent population, taking into account both its manifest clinical features, such as frequency and variety of methods, and its latent dimensions, like psychic functions and bodily investment, and improving the diagnostic process with the richness and complexity of psychoanalytic constructs. In a psychodynamic framework, which is the conceptual model that we apply, profiling is conceptualised as an interesting vision in which the diagnosis may include not only the psychopathology, but also the uniqueness of the individual, his/her subjectivity and his/her resources. Therefore, our study proposed to offer a new approach for the assessment of NSSI, that of profiling.

AIMS

Although the literature has largely investigated the different types of NSSI (Repetitive and Episodic; Brunner et al., 2007; Manca et al., 2014), the correlation between emotional regulation and body investment in NSSI (Cerutti, Manca & Presaghi, 2010; Muehlenkamp et al., 2013), as well as the functions of NSSI (Klonsky et al., 2015), our study aimed to collect together all these features in order to investigate different levels of the behaviour. As is well known, NSSI is a complex behaviour, comprising different dimensions. Thus, the study started from the hypothesis that some factors (manifest and latent clinical characteristics, gender and age of onset) may be present in the development of different profiles of NSSI.

Therefore, the aim of the study was to identify profiles of nonsuicidal self-injury by matching the manifest syndromal features (e.g., frequency of NSSI, the urgency i.e. the time lapse between thought and act, perception of the pain during injury, variety of methods used to injure oneself) to NSSI functions. These functions include not only emotional regulation and bodily investment but also latent functional dimensions that may be conceptualized as psychodynamic drivers motivating or reinforcing NSSI. We choose to investigate the intra-psychic dimensions of emotion regulation and body investment because respectively the first is the most common motivation referred by those who self-injure, and the second belongs to the crucial area of definition of Self and bodily boundaries, an important area in a psychodynamic perspective.

In addition, the study addressed NSSI in adolescence, as adolescents are considered the most at risk group for the behaviour.

METHOD

Participants and procedure

The participants were recruited in public high schools in an urban area in the South of Italy and involved in the study as part of a wider research project on nonsuicidal self-injury and risk behaviours in adolescence. The questionnaires were administered to a total sample of 589 adolescents, who completed the questionnaires in their classrooms during school hours. They were asked to respond anonymously. A sub-sample of 108 adolescents (mean age $M = 14.6$, $SD = .9$; 34 males and 74 females) reporting at least one episode of nonsuicidal self-injury behaviour, was eligible for the present study. Most of the students attended Scientific high schools (49%), while the rest attended Social Sciences (33%), Industrial Technology Institute (11%), and Classics (6%). All participants were Italian. The meetings for data collection were followed by a group discussion with adolescents in their classrooms; furthermore, meetings with parents and teachers were organized in order to inform them about the research findings.

The study was carried out after agreements had been made with the schools and their ethical commissions had approved the methods and aims of the research. Participation was voluntary, informed consent was given, and the privacy policy of the educational institution was respected. This research was approved by the Ethical Committee of the University of Naples Federico II.

Measures

All participants were provided with the *Inventory of Statements about Self-injury* (*ISAS*; Klonsky & Glenn, 2009), the *Body Investment Scale* (*BIS*; Orbach & Mikulincer, 1998) and the *Difficulties in Emotion Regulation Strategies* (*DERS*; Gratz & Roemer, 2004).

The ISAS (Klonsky & Glenn, 2009; Italian translation by Maura Manca) is a self-report questionnaire developed in two parts that considers both syndromal and functional aspects of NSSI. In the first part, the frequency of 12 nonsuicidal self-injurious behaviours which are performed deliberately and without suicidal intent are assessed. Examples of nonsuicidal self-injurious behaviours: banging/self-hitting, biting, burning, carving, cutting, wound picking, pinching, rubbing skin against rough surfaces and severe scratching. Participants were asked to estimate the number of times they have performed each behaviour. Five additional questions assess descriptive and contextual factors, including the age of onset, the experience of pain during NSSI, whether it was performed alone or with other people around, the time between the urge to self-injure and the act, and whether the individual wants to stop self-injuring or not. The last four have a multiple-choice format. In the second part 13 functions of the behaviours under two groups (intra-psychic and interpersonal) are assessed through 39 items. Five intra-psychic functions are emotion-regulation, anti-dissociation, anti-suicide, self-punishment, and marking distress. Eight interpersonal functions are: interpersonal boundaries, interpersonal influence, revenge, sensation seeking, peer-bonding, toughness, autonomy, and self-care. Each function was assessed by three items, rated as 0 = not relevant, 1 = somewhat relevant, or 2 = very relevant to the individual's experience of NSSI. The reliability and validity study of ISAS were done by Klonsky and Glenn (2009) and internal consistency for intra-psychic and interpersonal functions was found to be .88 and .80 respectively.

The BIS (Cerutti et al., 2010; Orbach & Mikulincer, 1998) is a brief 24-item self-report measure of emotional investment in the body, with particular reference to distorted body perceptions and the tendency to protect and damage one's body. Each item is scored on a 5-point scale, ranging from 1 (I do not agree at all) to 5 (I strongly agree). Nine of the items are reverse-scored, and items are summed up within each scale to obtain a total subscale score. The BIS includes four subscales: the body image, which includes items about body image, feelings and attitudes; the body touch includes items investigating personal comfort in situations of physical contact with others; the body care consists of items that investigate care behaviours towards one's body; the body protection investigates the desire to protect one's body. An exploratory factor analytic study with Israeli youths provided support for a four-factor solution. Estimates of internal consistency reported for the scale in Israeli youths ranged from .75 to.92. The reliability study of BIS Italian version using Cronbach's alpha was .65 (Cerutti et al., 2010). Estimates of concurrent validity were also reported by Orbach and Mikulincer (1998) for the instrument development and validation of Israeli samples. The scale already proved to be predictive of both self-harming behaviours and suicidal tendencies in the first study that the authors carried out with adolescents and young

adults aged between 13 and 19. In this research, those with higher scores reported greater self-esteem as well as having experienced adequate maternal care, they were more likely to indicate an ability to enjoy the pleasures of the body and its sensual aspects.

The DERS (Gratz & Roemer, 2004) is a 36-item self-report measure that assesses individuals' typical levels of emotion dysregulation across six domains: non-acceptance of emotional responses; difficulties in pursuing goal-directed behaviours when experiencing negative emotions; difficulties in controlling impulsive behaviours when experiencing negative emotions; lack of emotional awareness; limited access to emotion regulation strategies; and lack of emotional clarity. Higher values indicate greater difficulties in emotion regulation. The DERS has demonstrated good reliability (Cronbach's α = .93) and adequate construct and predictive validity and is significantly associated with objective (i.e., behavioural, physiological, and neurological) measures of emotion regulation (Gratz & Roemer, 2004; Gratz, 2007). The Italian version of the DERS was found to have adequate internal consistency (α =.90; Sighinolfi et al., 2010).

Statistical analysis

Internal consistency of the scales was assessed using Cronbach's α.

Two Step Cluster algorithm developed in SPSS (Version 23), was performed to find homogeneous clusters of nonsuicidal self-injury with respect to manifest clinical features. This method was selected because it is capable of handling both continuous and categorical variables and group data so that subjects within groups are similar in terms of the considered variables. If the desired number of clusters is unknown, the Two Step Cluster algorithm identifies automatically the optimal number of clusters and the best partitions in clusters, minimizing the Bayesian Information Criterion (BIC). Otherwise, the number of clusters can also be fixed by the users if number of clusters known a priori or the computed solution is not satisfactory.

Cluster models are typically used to find groups (or clusters) of similar records based on the variables examined, where the similarity between members of the same group is high and the similarity between members of different groups is low. The results can be used to identify associations that would otherwise not be apparent.

The log-likelihood method was used for similarity measures, since both continuous and categorical variables were considered in the analysis.

The interpretation of profiles takes into account the distribution of each characteristic among clusters, looking at the categories of each variable that mainly characterized each profile and following an interpretive criterion to assign labels to them. The silhouette measure of cohesion and separation was used as a measure for the overall goodness-of-fit of the found cluster structure.

We used a data-driven approach to define clusters because we aimed to identify the clusters (obtained by optimizing a statistical criterion) that emerged from the spontaneous aggregation of features, to interpret, then, the data with the help of the psychodynamic literature. Thus, a data-driven approach combined with a knowledge-based approach was essentially used, a blending that is increasingly implemented in Psychological Research (Dolce, Marocco, Maldonato & Sperandeo, 2020). A very similar approach was used also in Freda, Savarese, Dolce & Picione (2019).

For quantitative variables, data were reported as mean (± standard deviation) or median [25th;75th percentile], as appropriate, and Student's t-test or Mann-Whitney U test were performed accordingly to test for statistically significant differences between the two clusters obtained from the cluster analysis. For qualitative variables, data were reported as number of participants (%), and χ^2 test or Fisher's exact test were performed, as appropriate, to test for statistically significant differences between the two obtained clusters.

All statistical analyses were performed by using R (R Core Team, 2018). The level of significance was set at α = .05.

RESULTS

Internal consistency of the scales was considered satisfactory for DERS and for intra-psychic and interpersonal factors of ISAS (Cronbach's α was equal to .77, .73 and .83, respectively), while for BIS Cronbach's α coefficient was equal to .6. However, as mentioned above, Cronbach's α was .65 (less than .7) also in the study where the BIS Italian version was validated (Cerutti et al., 2010).

From our analysis, two clusters of nonsuicidal self-injury among a nonclinical sample of adolescents emerged. The silhouette measure indicated a weak overall goodness-of-fit of the cluster structure, but the partition was still satisfactory

from an interpretive point of view and interesting associations were identified. We also explored the three-clusters solution, but the two clusters optimal solution was the most satisfactory from an interpretive point of view. The approach we used to interpreting the results is that of examining fields across clusters to determine how and how much (in terms of percentages) values are distributed among clusters.

The first cluster was defined Repetitive Nonsuicidal Self-Injury (R-NSSI) and the second one Episodic Nonsuicidal Self-Injury (E-NSSI); these findings were in line with the literature, which showed two main types of self-injurious behaviour (Brunner et al., 2007; Madjar et al., 2017; Manca et al., 2014; Sarno et al., 2010; Shaffer & Jacobson, 2009).

The R-NSSI cluster assembled adolescents who reported having self-injured from 5 to 20 times in their lives (71.4%), more than 20 times in their life (79.5%), more than 10 times in the last year (100%), those whose last episode of nonsuicidal self-injury dated back to a few hours before the interview (75%), and those who had never tried to quit hurting themselves (61.5%). This represented the most clinically serious cluster.

Meanwhile, the E-NSSI represented the cluster where the self-injurious behaviour was used as an occasional symptom. Here we found self-harmers who usually tended to scratch (81%) and hit themselves (81.58%), who had self-injured less than 5 times in their life (100%), and never (100%) or only once (100%) in the last year, those whose last episode of NSSI dated back to between 2 months and 1 year before the interview (73.8) or more than 1 year (100%), those who usually spent many hours thinking about hurting themselves before doing it (about 80%), and those who tried at least once to stop hurting themselves (65.3%).

Study sample

The characteristics of the sample are presented in Table 1. Overall, 68% of the sample was female and almost the same gender distribution was found in the clusters that emerged. The mean age in years was 14.6(±.9), and it was very similar in the two clusters. Mean age of onset of nonsuicidal self-injury was equal to 12.5(±2), with no significant variations between the two clusters. Finally, all participants were from high schools, and in particular the percentages were as follows: 49.1% from Scientific, 33.3% from Social Sciences, 11.1% from Industrial Technology and 6.5% from Classics. The distribution of the type of high school did not differ significantly between the two clusters.

Table 1 – Characteristics of participants

	Total N = 108	**R-NSSI N = 41**	**E-NSSI N = 67**	**p**
Sex				1
Male	34 (31.5)	13 (31.7)	21 (31.3)	
Female	74 (68.5)	28 (68.3)	46 (68.7)	
Age (years)	14.6 (±.9)	14.8 (±.8)	14.5 (±1)	.177
Age of onset of Self-harm	12.5 (±2)	12.3 (±2.3)	12.6 (±1.8)	.441
High school address				.104
Scientific	53 (49.1)	24 (58.5)	29 (43.3)	
Human Sciences	36 (33.3)	13 (31.7)	23 (34.3)	
Industrial Tec. Institute	12 (11.1)	1 (2.4)	11 (16.4)	
Classics	7 (6.5)	3 (7.3)	4 (6)	

Note. Data are reported as number of patients (%) or mean (± standard deviation), as appropriate. *p*-values are based on Student's *t*-test, χ^2 test or Fisher's exact test, as appropriate.

The differences in syndromal features between the two clusters

The differences between the two clusters in terms of each considered manifest syndromal feature are presented in Table 2. Significant differences concerning the number of episodes of NSSI in a lifetime emerged between the two clusters (variable "times in life", $p<.001$), notably that 75% of the subjects in the R-NSSI cluster nonsuicidal self-injured more than 20 times in their lifetime compared to 11.9% of the subjects with E-NSSI.

Another important difference regarded the number of episodes of NSSI in the last year ("times in the last year", $p<.001$); the category "more than 10 times" was present in 75% of the subjects belonging to the first cluster while it was completely absent in the second.

The two clusters also differed significantly regarding the "last episode" of NSSI ($p<.001$), as 43.9% of the subjects belonging to the R-NSSI cluster engaged in NSSI a "few hours" before the survey, compared to 9% of the subjects with E-NSSI.

Finally, the time interval between the idea of NSSI to the act itself was another difference between the two clusters ("time", $p = .003$), with 78% of the subjects with R-NSSI usually letting "less than 1 hour" pass from thought of self-injury to the act, compared to the 49% of subjects belonging to the E-NSSI cluster.

In conclusion, the difference between the two clusters was not statistically significant in terms of the variables: "behaviour", "methods", "pain during the act" and "loneliness during the act". However, from a descriptive point of view, limited to our sample, we can observe that both clusters tended to choose a single behaviour: cutting, which was the most common.

The distribution of functional features between the two clusters

Regarding the psychic functions, there were differences between the two clusters (see Table 3). The Repetitive nonsuicidal self-injury was characterized by high scores in affect regulation (3.71±1.66 *vs* 2.79±1.58, $p = .005$), self-punishment (3.27±2.15 *vs* 2.24±1.77, $p = .0079$), and anti-suicide (2.00±1.86 *vs* 1.25±1.71, $p = .0354$). Following Klonsky and Glenn (2009), these three functions are categorized as intra-psychic types. More specifically, the correlation between R-NSSI and anti-suicide function confirmed the findings of other studies (Manca et al., 2014; Nock, Joiner, Gordon, Lloyd-Richardson & Prinstein, 2006).

Episodic nonsuicidal self-injury, on the other hand, was characterized by high scores in interpersonal influence (.85±.99 *vs* 1.55±1.64, $p = .0156$), defined as interpersonal function (Klonsky & Glenn, 2009). From a descriptive point of view, a distinction between interpersonal and intra-psychic functions emerged. Thus, the episodic cluster tended to report high scores of interpersonal functions, such as sensation seeking, peer bonding, interpersonal influence, compared to those of the repetitive cluster. In this sense, our findings confirmed Klonsky and colleagues' two factors model of NSSI functions (2015).

In particular, by analysing the emotion regulation function, it was found that R-NSSI is characterized by more significant and higher scores in the subscale awareness of emotions compared to those of E-NSSI (see Table 4; 9.8±3.25 *vs* 8.18±3.13, $p = .011$). This factor concerns the degree of attention focused on one's emotional state, which, from our data, appears as a consistently less developed characteristic in the Repetitive nonsuicidal self-injury cluster.

From a descriptive point of view, our study showed that the scores of all the other dimensions of emotion regulation, like the non-acceptance, goals, strategies, impulse and clarity of the emotions, which represent the incapacity to regulate emotions, are higher in the R-NSSI cluster compared to the scores of E-NSSI.

The differences of bodily investment between the two clusters

The analysis of body investment reveals that the Episodic nonsuicidal self-injury cluster shows more significant and higher scores in the subscale of body protection compared to those of Repetitive NSSI (see Table 5; 18.1±4.78 *vs* 20.34±5, $p = .023$). Our study also showed that the scores for body image are equally distributed between the two clusters, whereas scores for body touch and body care are higher in the E-NSSI cluster compared to those in the R-NSSI.

DISCUSSION

From our study, two profiles emerged for nonsuicidal self-injury behaviours among a nonclinical sample of adolescents, namely: Serious NSSI and Moderate NSSI.

Table 2 – Manifest features related to clusters

	Total N = 108	**R-NSSI N = 41**	**E-NSSI N = 67**	**p**
Behaviour				.906
multiple	31 (28.7)	11 (26.8)	20 (29.9)	
single	77 (71.3)	30 (73.2)	47 (70.1)	
Method				.045
cutting	58 (53.7)	25 (61)	33 (49.3)	
bite	18 (16.7)	10 (24.4)	8 (11.9)	
scratching	21 (19.4)	4 (9.8)	17 (25.4)	
hitting	11 (10.2)	2 (4.9)	9 (13.4)	
Times in life				**<.001**
1	14 (13)	0 (0)	14 (20.9)	
2-5	41 (38)	0 (0)	41 (61.2)	
5-20	14 (13)	10 (24.4)	4 (6.0)	
>20	39 (36.1)	31 (75.6)	8 (11.9)	
Times in the last year				**<.001**
never	20 (18.5)	0 (0)	20 (29.9)	
1	16 (14.8)	0 (0)	16 (23.9)	
2-10	41 (38)	10 (24.4)	31 (46.3)	
>10	31 (28.7)	31 (75.6)	0 (0)	
Last episode				**<.001**
few hours	24 (22.2)	18 (43.9)	6 (9.0)	
1 month	20 (18.5)	12 (29.3)	8 (11.9)	
2-12 months	42 (38.9)	11 (26.8)	31 (46.3)	
>1 year	22 (20.4)	0 (0)	22 (32.8)	
Pain				.089
yes	44 (40.7)	15 (36.6)	29 (43.3)	
sometimes	45 (41.7)	22 (53.7)	23 (34.3)	
no	19 (17.6)	4 (9.8)	15 (22.4)	
Loneliness				.186
yes	74 (68.5)	25 (61.0)	49 (73.1)	
sometimes	22 (20.4)	12 (29.3)	10 (14.9)	
no	12 (11.1)	4 (9.8)	8 (11.9)	
Time				**.003**
<1 hour	65 (60.2)	32 (78.0)	33 (49.3)	
1-6 hours	13 (12)	0 (0)	13 (19.4)	
7-24 hours	11 (10.2)	3 (7.3)	8 (11.9)	
>1 day	19 (17.6)	6 (14.6)	13 (19.4)	
Stop				.074
yes	95 (88)	33 (80.5)	62 (92.5)	
no	13 (12)	8 (19.5)	5 (7.5)	

Note. Data are reported as number of patients (%) or mean (± standard deviation), as appropriate. *p*-values are based on Student's *t*-test, χ^2 test or Fisher's exact test, as appropriate.

Table 3 – Results of psychic functions related to clusters

	Total N = 108	R-NSSI N = 41	E-NSSI N = 67	p
Affect regulation	3.14 ± 1.67	3.71 ± 1.66	2.79 ± 1.58	**.0050**
Interpersonal boundaries	1.37 ± 1.53	1.37 ± 1.58	1.37 ± 1.51	.9809
Self-punishment	2.63 ± 1.97	3.27 ± 2.15	2.24 ± 1.77	**.0079**
Self-care	1.58 ± 1.25	1.78 ± 1.35	1.46 ± 1.18	.2026
Anti-dissociation	1.49 ± 1.25	1.71 ± 1.25	1.36 ± 1.24	.1597
Anti-suicide	1.54 ± 1.80	2.00 ± 1.86	1.25 ± 1.71	**.0354**
Sensation-seeking	.63 ± 1.12	.54 ± 1.03	.69 ± 1.17	.5003
Peer-bonding	.94 ± 1.47	.66 ± 1.39	1.10 ± 1.51	.1277
Interpersonal influence	1.29 ± 1.47	.85 ± .99	1.55 ± 1.64	**.0156**
Toughness	1.70 ± 1.64	1.59 ± 1.63	1.78 ± 1.66	.5604
Marking distress	1.81 ± 1.55	2.15 ± 1.74	1.60 ± 1.39	.0736
Revenge	.75 ± 1.14	.61 ± 1.00	.84 ± 1.23	.3216
Autonomy	1.40 ± 1.51	1.63 ± 1.56	1.25 ± 1.47	.2053

Note. Data are reported as mean (± standard deviation) or median [25th; 75th percentile] as appropriate. *p*-values are based on Student's *t*-test or Mann-Whitney test, as appropriate.

Table 4 – Results of emotion regulation relating to clusters

	Total N = 108	**R-NSSI N = 41**	**E-NSSI N = 67**	**p**
DERS-Non-Acceptance	15.64(±7.33)	16.76(±8.04)	14.96(±6.83)	.217
DERS-Goals	16.32(±5.32)	16.93(±5.31)	15.96(±5.34)	.360
DERS-Strategies	25.04(±6.45)	25.98(±7.59)	24.46(±5.62)	.238
DERS-Impulse	17.08(±6.49)	17.83(±6.14)	16.63(±6.70)	.352
DERS-Clarity	17.02(±4.93)	17.54(±5.32)	16.70(±4.68)	.395
DERS-Aware	8.80(±3.26)	9.80(±3.25)	8.18(±3.13)	**.011**
DERS-Tot	99.9(±23.69)	104.83(±25.9)	96.88(±21.9)	.091

Note. Data are reported as mean (± standard deviation). *p*-values are based on Student's *t*-test.

Table 5 – Results of body investment relating to clusters

	Total N = 108	**R-NSSI N = 41**	**E-NSSI N = 67**	**p**
BIS-Image	15.83(±6.86)	15.95(±7.18)	15.76(±6.72)	.890
BIS-Touch	19.55(±4.68)	18.73(±4.70)	20.04(±4.64)	.158
BIS-Care	20.19(±4.25)	19.32(±4.99)	20.72(±3.67)	.097
BIS-Protection	19.49(±5.02)	18.10(±4.78)	20.34(±5.00)	**.023**
BIS-TOT	75.06(±14.2)	72.10(±14.82)	76.87(±13.6)	.090

Note. Data are reported as mean (± standard deviation). *p*-values are based on Student's *t*-test.

The Serious nonsuicidal self-injury profile was characterized by repetitive and chronic behaviours, intra-psychic functions and low body investment, representing the most clinically compromised profile. In particular, the dimensions of compulsiveness and impulsiveness of the behaviour, which characterized the Repetitive NSSI cluster, were clinical indicators of a very deep and pervasive anguish in those who self-injure. In addition, these features were associated with intra-psychic functions, such as the regulation of emotions and tensions, the punishment of the self through the body, and the use of nonsuicidal self-injury as a defence against suicidal attempts. The correlation we found between intra-psychic functions and clinical severity is in line with the literature (Klonsky & Glenn, 2009; Nock & Prinstein, 2005), suggesting that endorsement of intrapersonal functions may be indicative of NSSI that is more persistent and more likely to progress to medically severe forms of self-injury, including suicide attempts.

Intra-psychic functions are related to the need to defend and sustain the integrity of the Self. Thus, in a psychodynamic framework, nonsuicidal self-injury is considered an extreme way of preserving the self-cohesion and of protecting the boundaries of self against the loss of identity (Simpson & Porter, 1981): the scars may create a continuity of existence, connecting episodes of dissociation or preserving past events or emotions that could not be integrated into the sense of identity (Miller & Bashkin, 1974). The self-injurious behaviour as a form of skin containment (Turp, 2007) makes it possible to maintain a sense of identity in the face of overwhelming internal emotion. In particular, in adolescence, when it is difficult to symbolize and mentalize the affects (Haza & Keller, 2005; Rossouw & Fonagy, 2012), and when the instinct can physiologically exceed the capacity of representation, the impulsive behaviours are used to express, through the body, emotions that cannot be put into words yet (Blos, 1967).

In these scenarios, the rejected and injured body becomes a theatre of conflict. Adolescents who endorsed Repetitive NSSI were more likely to have a negative body image, to exhibit a lower tendency to protect it, in line with the main studies (Brunner et al., 2007; Cerutti, Manca, Presaghi & Gratz, 2012; Muehlenkamp, Swanson & Brausch, 2005), and to experience difficulty being with others. For all these reasons, the Serious NSSI profile is characterized by intra-psychic management of the malaise, using the nonsuicidal self-injurious behaviour as a psychopathological compromise to manage conflicts on an individual level.

The second profile, the Moderate nonsuicidal self-injury, was characterized by episodic behaviours, interpersonal functions and high body investment. It represented the profile where the nonsuicidal self-injury was used as an occasional symptom. The characteristics of this group showed a pathologically less structured clinical profile. Regarding the dimension of body investment, people with this profile tended to protect and take care of themselves, and to be more open to other people. Those who belong to this profile reported functions and motivations that have to do with the area of relationships, for example the interpersonal influence: in this sense, the symptom was probably used to communicate with others, to share the behaviour with peers and make it a group ritual, aimed at giving it new meanings. This explains why this profile is oriented towards relationships, whereby the symptom is already a tool to search for the other, for help, even if presented in a still pre-symbolic form. NSSI can also pinpoint to the secondary gains of attention and control over others. Thus, as shown in our previous studies (Margherita & Gargiulo, 2018), the area of relationships in those who self-harm is denoted by ambivalence and conflict; the other is thus painfully present through its very absence, and is continually called on, and, at the same time, rejected. Therefore, from a psychodynamic perspective of object relations (Briggs, Lemma & Crouch, 2008; Lemma, 2010), self-injurious behaviour has been read as an unconscious attempt at separation (e.g. violently cutting away the other, who seems to be living within their own body), a way to distinguish between the self and others, to create boundaries and protect against feelings of being engulfed or fear of loss of identity (Suyemoto, 1998).

In conclusion, our findings are in line with the literature concerning the emerging of two main and different types of NSSI (Brunner et al., 2007; Madjar et al., 2017; Manca et al., 2014; Sarno et al., 2010), as well as with results confirmed by Klonsky and colleagues' two factors model of NSSI functions, intrapersonal and social (2015).

No significant gender differences were found between the two NSSI profiles, confirming current studies about the gender variable between Repetitive and Episodic nonsuicidal self-injury (Brunner et al., 2007; Manca et al., 2014). However, since previous studies evidenced gender differences in terms of methods and functions (Gargiulo & Margherita, 2014), we assumed that our small sample size prevented gender differences from emerging; in our opinion this could assume a clinical relevance in the assessment of NSSI among

adolescents and we hope it will be examined further in future studies.

Finally, no difference was observed for the age of onset of the first episode of NSSI. This result is in line with other studies (Cerutti et al., 2011) and adds evidence that nonsuicidal self-injury tends to emerge in the adolescent community population at the same age, independently of whether it is Repetitive or Episodic.

Limitations

This study is not free from limitations. We are aware that the sample size, limited to a geographic area of Italy, constitutes a limitation for the study; therefore, in our future research, we will increase the sample. In particular, we will balance it for the gender variable in order to test this aspect when profiling self-harm. Another limitation is the lack of certified clinical data at the baseline, due to the impossibility of submitting a clinical test at school; however, in our future research, we will consider also a clinical sample, and we will compare the clinical to the nonclinical one.

CONCLUSION

The aim of the study was to identify profiles for nonsuicidal self-injury behaviours among a nonclinical sample of adolescents. Two main psychodynamic profiles emerged as follows: Serious nonsuicidal self-injury, which represents the more clinically compromised profile, seemed to be more oriented to an intra-psychic level of management of malaise, while the Moderate profile is projected on to the interpersonal one, and more open to the possibility of sharing the pain with others.

Findings from the present study have some clinical implications. First of all, the results suggest that for a large percentage of adolescents, NSSI may represent an episodic behaviour that is not always linked to a clinically compromised profile. In this sense, it is important to have a diagnostic process that does not stigmatize young people (Gargiulo et al., 2014). These data highlight the importance of longitudinal studies in adolescence in order to monitor the occurrence of self-injurious behaviours over time, as well as understanding whether occasional occurrences of NSSI are precursors of Repetitive NSSI, and under which conditions a remission or transition into Repetitive NSSI might occur.

Furthermore, if we identify the specific psychological aspects of each profile, then clinical practices can be developed to address the specific needs and psychological functioning of young self-harmers. Psychic functions could be used by clinicians during the diagnostic process to compile more detailed and sensitive clinical assessments of nonsuicidal self-injury. In this way, it might be possible to understand whether Moderate NSSI may constitute an early indicator of severe personality disturbance. Finally, understanding the functions of NSSI can be critical for treating individuals engaging in NSSI.

Understanding the NSSI experience in the young population is a first step towards developing patient-tailored programs and to supporting preventive interventions at school. In the future, therefore, we will aim to profile self-harmers, maybe integrating quantitative with qualitative research, and thereby integrating nomothetic understanding and idiographic knowledge of clinical presentations.

References

ADRIAN, M., ZEMAN, J., ERDLEY, C., LISA, L. & SIM, L. (2011). Emotional dysregulation and interpersonal difficulties as risk factors for nonsuicidal self-injury in adolescent girls. *Journal of Abnormal Child Psychology*, *39*, 389-400.

AMERICAN PSYCHIATRIC ASSOCIATION (2013). *DSM-5. Diagnostic and statistical manual of mental disorders (5th ed.)*. Arlington, VA: American Psychiatric Publishing.

BERGER, E., HASKING, P. & REUPERT, A. (2014). Response and training needs of school staff towards student self-injury. *Teaching and Teacher Education*, 44, 25-34. doi.org/10.1016/j.tate.2014.07.013

BLOS, P. (1967). The second individuation process of adolescence. *The Psychoanalytic Study of the Child*, *22*, 162.

BRESIN, K. & SCHOENLEBER, M. (2015). Gender differences in the prevalence of nonsuicidal self-injury: A meta-analysis. *Clinical Psychology Review*, *38*, 55-64.

BRIGGS, S., LEMMA, A. & CROUCH, W. (2008). *Relating to self-harm and suicide: Psychoanalytic perspectives on practice, theory and prevention*. London: Routledge.

BROWN, R.C. & PLENER, P.L. (2017). Non-suicidal self-injury in adolescence. *Current Psychiatry Reports*, *19*, 20. doi:https://doi.org/10.1007/s11920-017-0767-9

BRUNNER, R., PARZER, P., HAFFNER, J., STEEN, R., ROOS, J. & KLETT, M.F.R., (2007). Prevalence and psychological correlates of occasional and repetitive deliberate self-harm in adolescents. *Arch Pediatr Adolescent Med*, *161*, 641-649.

CERUTTI, R., MANCA, M. & PRESAGHI, F. (2010). Body investment and deliberate self-harm in Italian university student sample. *Minerva Psichiatrica*, *51*, 3.

CERUTTI, R., MANCA, M., PRESAGHI, F. & GRATZ, K.L. (2011). Prevalence and clinical correlates of deliberate self-harm among a community sample of Italian adolescents. *Journal of Adolescence*, *34*, 337-347.

CERUTTI, R., MANCA, M., PRESAGHI, F. & GRATZ, K.L. (2012). Deliberate self-harm behaviour among Italian young adults: Correlations with clinical and non clinical dimensions of personality. *Am Journal of Orthopsychiatry*, *82*, 298-308.

CIPRIANO, A., CELLA, S. & COTRUFO, P. (2017). Nonsuicidal self-injury: A systematic review. *Frontiers in Psychology*, *8* (8), 46-59. doi: 10.3389/fpsyg.2017.01946

CROUCH, W. & WRIGHT, J. (2004). Deliberate self-harm in an adolescent unit: A qualitative investigation. *Clinical Child Psychology and Psychiatry*, *9* (2), 185-204.

DICÉ, F., MAIELLO, A., DOLCE, P. & FREDA, M.F. (2017). Le trasformazioni dialogiche delle preoccupazioni in pediatria di famiglia: Un'indagine empirica. *Psicologia della Salute*, *3*, 74-96. doi: 10.3280/PDS2017-003004

DI PIERRO, R., SARNO, I., GALLUCCI, M. & MADEDDU, F. (2014). Nonsuicidal self-injury as an affect-regulation strategy and the moderating role of impulsivity. *Child and Adolescent Mental Health*, *9* (4), 259-264. doi.org/10.1111%2Fcamh.12063

DOLCE, P., MAROCCO, D., MALDONATO, M.N. & SPERANDEO, R. (2020). Towards a machine learning predictive-oriented approach to complement explanatory modeling. An application for evaluating psychopathological traits based on affective neurosciences and phenomenology. *Frontiers in Psychology*, *11*, 446.

FORTUNE, S., SINCLAIR, J. & HAWTON, K. (2008). Adolescents' views on preventing self-harm: A large community study. *Social Psychiatry and Psychiatric Epidemiology*, *43* (2), 96-104.

FREDA, M.F., SAVARESE, L., DOLCE, P. & PICIONE, R.L. (2019). Caregivers sensemaking of children's hereditary angioedema: A semiotic narrative analysis of the sense of grip on the disease. *Frontiers in Psychology*, *10*, 2609.

GARGIULO, A. (2020). Narratives of self-harm at school: Identifying trajectories to develop psychological interventions in educational contexts. *Europe's Journal of Psychology*, *16* (1), 95-111. doi.org/10.5964/ejop.v16i1.1883

GARGIULO, A. & MARGHERITA, G. (2014). Autolesività non suicidaria e genere: Rassegna teorica e riflessioni psicodinamiche. *Infanzia e Adolescenza*, *13* (2), 119-128. doi:10.1710/1624.17656

GARGIULO, A. & MARGHERITA, G. (2019). Narratives of self-harm: The experience of young women through the qualitative analysis of blogs. *Mediterranean Journal of Clinical Psychology*, *7* (1), 1-19. doi: 10.6092/2282-1619/2019.7.2055

GARGIULO, A., PLENER, P.L., BAUS, N., MARGHERITA, G., BRUNNER, R., KAES, M. & KAPUSTA, N. (2014). Autolesività non suicidaria (NSSI) e Disturbo da comportamento suicidario (SBD) nella recente pubblicazione del DSM-5. [Nonsuicidal self-injury (NSSI) and Suicidal behavior disorder (SBD) in DSM-5]. *Minerva Psichiatrica*, *55* (2), 83-90.

GARGIULO, A., TESSITORE, F., LE GROTTAGLIE, F. & MARGHERITA, G. (2020). Self-harming behaviours of asylum seekers and refugees in Europe: A systematic review. *International Journal of Psychology.* doi: 10.1002/ijop.12697

GRATZ, K.L. (2007). Targeting emotion dysregulation in the treatment of self-injury. *Journal of Clinical Psychology. Session*, *63* (11), 1091-1103.

GRATZ, K.L. & ROEMER, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. *Journal of Psychopathology and Behavioral Assessment*, *26*, 41. doi.org/10.1023/B:JOBA.0000007455.08539.94

HAZA, M. & KELLER, P.H. (2005). Scarifications chez l'adolescent suicidaire. Une tentative pour penser? *Adolescence*, *3*, 733-742.

KLONSKY, E.D. (2007). The function of deliberate self-injury. A review of evidence. *Clinical Psychology Review*, *27*, 226-239.

KLONSKY, E.D. (2009). The functions of self-injury in young adults who cut themselves: Clarifying the evidence for affect-regulation. *Psychiatry Research*, *166*, 260-268.

KLONSKY, E.D. & GLENN, C.R. (2009). Assessing the functions of non-suicidal self-injury: Psychometric properties of the Inventory of Statements About Self-injury (ISAS). *J Psychopathol Behav Assessment*, *31*, 215-219.

KLONSKY, E.D., GLENN, C.R., STYER, D.M., OLINO, T.M. & WASHBURN, J.J. (2015). The functions of nonsuicidal self-injury: Converging evidence for a two-factor structure. *Child Adolesc Psychiatry Ment Health*, *9*, 44. doi.org/10.1186/s13034-015-0073-4

LAYE-GINDHU, A. & SCHONERT-REICHL, K.A. (2005). Nonsuicidal self-harm among community adolescents: Understanding the "Whats" and "Whys" of self-harm. *Journal of Youth and Adolescence*, *34*, 447-457.

LEMMA, A. (2010). *Under the skin. A Psychoanalytic study of the body modification*. London: Routledge.

LEWIS, S.P. & HEATH, N.L. (2015). Nonsuicidal self-injury among youth. *J Pediatr*, *166* (3), 526-530. doi: 10.1016/j.jpeds.2014.11.062.

LEWIS, S.P. & SANTOR, D.A. (2010). Self-harm reasons, goal achievement, and prediction of future self-harm intent. *J Nerv Ment Dis*, *198*, 362-369.

LINGIARDI, V. & McWILLIAMS, N. (2017). *Manuale diagnostico psicodinamico. Seconda edizione. PDM-2*. Raffaello Cortina Editore.

MADGE, N., HEWITT, A., HAWTON, K., WILDE, E.J.D., CORCORAN, P., FEKETE, S., HEERINGEN, K., DE LEO, D. & YSTGAARD, M. (2008). Deliberate self-harm within an international community sample of young people: Comparative findings from the Child & Adolescent Self-harm in Europe (CASE) Study. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *49* (6), 667-677. doi: 10.1111/j.1469-7610.2008.01879.x

MADJAR, N., ZALSMANB, G., MORDECHAIA, T.R.B. & SHOVALB, G. (2017). Repetitive vs occasional non-suicidal self-injury and school-related factors among Israeli high school students. *Psychiatry Research*, *257*, 358-360.

MANCA, M., CERUTTI, R. & PRESAGHI, F. (2005). Sviluppo e validazione di un questionario per la rilevazione della sindrome da autolesionismo ripetitivo. *Atti del VI Congresso Nazionale della Associazione Italiana di Psicologia (AIP) - Sezione di Psicologia Clinica*.

MANCA, M., PRESAGHI, F. & CERUTTI, R. (2014). Clinical specificity of acute versus chronic self-injury: Measurement and evaluation of repetitive non-suicidal self-injury. *Psychiatry Research*, 215, 111-119.

MARGHERITA, G. & GARGIULO, A. (2018). A comparison between pro-anorexia and non-suicidal self-injury blogs: From symptom-based identity to sharing of emotions. *Psychodynamic Practice*, *24* (4), 346-363. doi: 10.1080/14753634.2018.1535326

MARTINO, M.L., GARGIULO, A., LEMMO, D., DOLCE, P., BARBERIO, D., ABATE, V., AVINO, F. & TORTORIELLO, R. (2019). Longitudinal effect of emotional processing on psychological symptoms in women under 50 with breast cancer. *Health Psychology Open*, 1-9. doi: 10.1177/2055102919844501

MILLER, F. & BASHKIN, E.A. (1974). Depersonalization and self-mutilation. *Psychoanalytic Quarterly*, 43, 638-649.

MOYER, M. & NELSON, K. (2007). Investigating and understanding self mutilation: The student voice. *Professional School Counseling*, *11* (1), 42-48.

MUEHLENKAMP, J. (2012). Body regard in nonsuicidal self-injury. *Journal of Cognitive Psychotherapy*, *26* (4).

MUEHLENKAMP, J.J., BAGGE, C.L., TULL, M.T. & GRATZ, K.L. (2013). Body regard as a moderator of the relation between emotion dysregulation and nonsuicidal self-injury. *Suicide Life Threat Behav*, *43* (5), 479-493.

MUEHLENKAMP, J.J., SWANSON, J.D. & BRAUSCH, A.M. (2005). Self-objectification, risk taking, and self-harm in college women. *Psychol Women Q*, *29*, 24-32.

MULAY, A.L., WEST, M.L., WALLNER SAMSTAG, L. & DIAMOND, E. (2017). Negative body attitudes as a risk factor for non-suicidal self-injury among college students. *Suicidology Online*, *8* (2), 11-19.

NOCK, M.K. (2010). Self-Injury. *Annual Review of Clinical Psychology*, *6*, 339-363.

NOCK, M.K., JOINER, T.E., GORDON, K.H., LLOYD-RICHARDSON, E. & PRINSTEIN, M.J. (2006). Non-suicidal self-injury among adolescents: Diagnostic correlates and relation to suicide attempts. *Psychiatry Res*, *144*, 65-72.

NOCK, M.K. & PRINSTEIN, M.J. (2004). A functional approach to

the assessment of self-mutilative behavior. *J Consult Clin Psychol*, *72*, 885-890

NOCK, M.K. & PRINSTEIN, M.J. (2005). Contextual features and behavioral functions of self mutilation among adoltescens. *Journal of Abnormal Psychology*, *114*, 140-146.

ORBACH, I. (1996). The role of body experience in self-destruction. *Clin Psychol Psychiatry*, *1* (4), 607-619.

ORBACH, I. & MIKULINCER, M. (1998). The body investment scale: Construction and validation of a body experience scale. *Psychological Assessment*, *10*, 415-425.

PLENER, P.L. & FEGERT, J.M. (2015). Nonsuicidal self-injury: A condition for further study. *Child and Adolescent Psychiatry and Mental Health*, *9*, 30. doi.org/10.1186/s13034-015-0067-2

PLENER, P.L., SCHUMACHER, T.S., MUNZ, L.M. & GROSCHWITZ, R.C. (2015). The longitudinal course of non-suicidal self-injury and deliberate self-harm: A systematic review of the literature. *Borderline Personality Disorder and Emotion Dysregulation*, *2* (1), 2. doi.org/10.1186/s40479-014-0024-3

R CORE TEAM (2018). *R: A language and environment for statistical computing. R Foundation for statistical computing*. Wien. Retrieved from https://www.R-project.org/

ROSSI MONTI, M. & D'AGOSTINO, A. (2018). Diagnosis or diagnosing? From symptoms to a phenomenological-psychoanalytic stance on the patient's world. *Psychoanalytic Review*, *105* (2), 209-222. doi: 10.1521/prev.2018.105.2.209

ROSSOUW, T.I. & FONAGY, P. (2012). Mentalization-based treatment for self-harm in adolescents: A randomized controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, *51* (12). doi.org/10.1016/j.jaac.2012.09.018

SARNO, I., MADEDDU, F. & GRATZ, K.L. (2010). Self-injury, psychiatric symptoms, and defense mechanisms: Findings in an Italian nonclinical sample. *European Psychiatry*, *25*, 136-145.

SHAFFER, D. & JACOBSON, C. (2009). *Proposal to the DSM-V childhood disorder and mood disorder work groups to include Non Suicidal Self-Injury (NSSI) as a DSM-V disorder. Proposed revisions: Disorders usually first diagnosed in infancy, childhood or adolescence. DSM-5 development*. Arlington, VA: American Psychiatric Association.

SIGHINOLFI, C., NORCINI PALA, A., CHIRI, L.R., MARCHETTI, I. & SICA, C. (2010). Difficulties in Emotion Regulation Scale (DERS): Traduzione e adattamento italiano. *Psicoterapia Cognitiva e Comportamentale*, *16*, 141-170.

SIMPSON, C.A. & PORTER, G.L. (1981). Self-mutilation in children and adolescents. *Bulletin of the Menninger Clinic*, *45* (5), 428-438.

SWANNELL, S.V., MARTIN, G.E., PAGE, A., HASKING, P. & ST JOHN, N.J. (2014). Prevalence of nonsuicidal self-injury in nonclinical samples: Systematic review, meta-analysis and meta-regression. *Suicide Life Threat Behav*, *44* (3):273-303. doi: 10.1111/sltb.12070

SUYEMOTO, K. (1998). The functions of self mutilation. *Clinical Psychology Review*, *18*, 531-554.

TURP, M. (2007). Self-harm by omission: A question of skin containment. *Psychodynamic Practice*, *13* (3), 229-244. doi:https://doi.org/10.1080/14753630701455812

VALENCIA-AGUNDO, F., CORBET BURCHER, G., EZPELETA, L. & KRAMER, T. (2018). Nonsuicidal self-injury in community adolescents: A systematic review of prospective predictors, mediators and moderators. *Journal of Adolescence*, *65*, 25-38.

WHITLOCK, J.L. & RODHAM, K. (2013). Understanding nonsuicidal self-injury in youth. *School Psychology Forum: Research in Practice*, *7* (4), 1-18.

DOI: 10.26387/bpa.286.2

Urdu translation of Flourishing Scale: Evidence for the validity and measurement invariance across gender

Sadia Niazi[1], Adnan Adil[2]

[1] Lecturer, Department of Psychology, University of Sargodha
[2] Assistant Professor, Department of Psychology, University of Sargodha

livespirit786@yahoo.com

ABSTRACT. Una valutazione completa del benessere comporta la misurazione delle prospettive sia edoniche che eudaimoniche di benessere e la *Flourishing Scale*, composta da 8 item, valuta entrambi questi aspetti del benessere. Pertanto, la ricerca corrente si propone di tradurre la *Flourishing Scale* in urdu e validarla per la popolazione pakistana. La Flr.S è stata tradotta nella lingua urdu attraverso la procedura di back translation standard. Si è esaminata l'affidabilità test-retest (15 giorni) per le versioni urdu-inglese e inglese-urdu in un campione mirato di studenti universitari (N = 60). In un campione casuale separato di 574 studenti universitari pakistani (maschi = 235 e femmine = 339), sono state somministrate le versioni in urdu della Flr.S e della *Core Self-evaluation Scale* per stabilire la validità di costrutto e fattoriale della versione in urdu della Flr.S. L'affidabilità test-retest di due settimane per le versioni urdu-inglese e inglese-urdu era >.90. L'analisi fattoriale confermativa (CFA) della Flr.S ha rivelato una soluzione a fattore singolo, dimostrando l'invarianza configurale, metrica e scalare tra i generi. Le medie latenti dei ragazzi e delle ragazze sulla Flr S erano comparabili: la Flr.S ha dimostrato una validità di costrutto. I nostri risultati suggeriscono che la versione urdu della Flr.S è una misura affidabile e valida del costrutto del fiorire per entrambi i generi nelle popolazioni di lingua urdu.

SUMMARY. A comprehensive assessment of well-being involves the measurement of both hedonic and eudaimonic perspectives of well-being and the 8-item Flourishing Scale assesses both of these aspects of well-being. Therefore, the current research translated Flr.S into Urdu and validated it for the Pakistani population. Flr.S was translated into the Urdu language through the standard back-translation procedure. Test-retest reliability (15 days) for the Urdu-English and English-Urdu versions was established in a purposive sample of university undergraduates (N = 60). In a separate random sample of 574 Pakistani university undergraduates (males = 235 and females = 339), the Urdu version of Flr.S and Urdu version of Core Self-evaluation Scale were administered to establish the factorial and the construct validity of Urdu Flr.S. The test-retest reliability of two-week for the Urdu-English and English-Urdu versions of Flr.S was >.90. The confirmatory factor analysis (CFA) of Flr.S revealed a single factor solution, which demonstrated evidence for the configural, metric, and scalar invariance across genders. The latent means of males and females on the Flr.S were comparable: the Flr.S demonstrated construct validity. Our findings suggest that the Urdu version of the Flr.S is a reliable and valid measure of flourishing for both genders in the Urdu speaking population.

Keywords: *Flourishing, Translation, Cross-language validation, Measurement invariance*

INTRODUCTION

Flourishing means to live within an optimal range of human ability. It is a mixture of both functioning and feelings. Flourishing is considered a sign of a high level of mental well-being (Huppert, 2009). However, the concept of flourishing is based on recent theories of social and psychological well-being. One of the most prominent conceptions of psychological well-being is based on the eudaimonic and hedonic paradigms (Woyciekoski, Stenert & Hutz, 2012). Hedonia is subjective happiness that involves seeking pleasure and avoiding pain (Diener, Oishi & Lucas, 2003) and it can be considered as subjective well-being characterized by satisfaction with life and positive affect. On the other hand, eudaimonia refers to the personal experience of self-realization, personal growth, and meaning in life (Ryan & Deci 2001). More precisely, it can be considered as psychological well-being (Waterman et al., 2010). The instruments developed in consonance of this conception of psychological well-being assess people's perception and life evaluation of happiness (Diener et al., 2003).

Previous literature showed that most of the self-report measures for the assessment of well-being were either based on eudaimonic perspective (e.g., Ryff & Keyes, 1995; Waterman et al., 2010) or hedonic perspective (scales assessing negative affect, life satisfaction, and positive affect). Based on the eudaimonic perspective, Ryff and Keyes' (1995) *Psychological Well-being Scale* comprised six theoretically derived dimensions including personal growth, autonomy, the meaning of life, constructive relations with others, mastery, and acceptance of self. The CFA of the *Psychological Well-being Scale* through maximum likelihood estimation in a sample of 1108 adults supported the proposed six-factor model of psychological well-being (χ^2 = 339, *df* = 120, AGFI = .89, BIC = −167). Moreover, the *Psychological Well-being Scale* was found to be positively related to *Life Satisfaction Index* (Neugarten, Havighurst & Tobin, 1961) and *Affect Balance Scale* (Bradburn, 1969); and negatively related to *Zung Depression Scale* (Zung, 1965), which established evidence for its construct validity (Ryff & Keyes, 1995). Currently, the Ryff's *Psychological Well-being Scale* has three versions. The longest version comprises 84 items (14 for each scale), the mid-length version includes 54 items (9 per scale), and the shortest version comprises 18 items (3 per scale).

Questionnaire of Eudaimonic Well-being (Waterman et al., 2010) is a 21-item scale, which assesses subjective well-being. This questionnaire was based on eudaimonic identity theory (Waterman, 2007) and its unidimensional factorial structure was confirmed in a CFA using the maximum likelihood estimation (χ^2 = 22.59, *df* = 5, CFI = .99, NNFI = .98, RMSEA = .065, SRMR = .018) in a large sample of college students (*N* = 1728) of the USA. The factor loading of the items ranged from .60 to .80 and the Cronbach alpha value of scale was .85. The construct validity of *Questionnaire of Eudaimonic Well-being* was established as it was found to be positively correlated with self-esteem (measured through the *Rosenberg Self-Esteem Scale*; Rosenberg, 1965) and internal locus of control (measured through *Locus of Control Scale*; Côté, 1997) and negatively related to general anxiety (measured through *Beck Anxiety Inventory*; Beck, Steer & Carbin, 1988) and social anxiety (measured through *Social Interaction Anxiety Scale*; Habke, Hewitt, Norton & Asmundson, 1997).

On the other hand, instruments based on the hedonic perspective have also been developed and these measures mainly cover three dimensions of hedonic well-being including life satisfaction, negative affect, and positive affect. For instance, *Satisfaction with Life Scale* was developed by Diener, Emmons, Larsen and Griffin (1985). This 5-item scale covered cognitive features of well-being. In an exploratory factor analysis (EFA) using principal axis factoring, Diener et al. (1985) found a unidimensional structure of the scale, which accounted for 66% variance. The item loadings ranged from .60 to .85. The *Satisfaction with Life Scale* showed favorable psychometric properties such as the high value of Cronbach alpha (.87) and test-retest reliability of two months (*r* = .82). The authors also established the construct validity of *Satisfaction with Life Scale* as it was found to be positively related to Bradburn's *Affect Balance Scale* (1969) and the *Rosenberg Self-Esteem Scale* (1965).

Positive and Negative Affect Schedule (PANAS) was developed by Watson, Clark and Tellegen (1988) for the assessment of the affective feature of well-being. The EFA with principal axis factoring extraction revealed two distinct factors, which accounted for 62.8% variance. The item loadings of all the items remained >.50, which showed that all items of PANAS were good indicators of their corresponding factor. The Cronbach's alpha of positive affect was .86 whereas the same for the negative affect was .87. The PANAS also demonstrated good temporal stability over two months (*r* = .87).

The 15-item *Subjective Well-being Scal*e (Lawrence & Liang, 1988) integrated the *Affect Balance Scale* (Bradburn, 1969) and the *Life Satisfaction Index A* (Neugarten et al.,

1961) into one measure following the subjective well-being model of Liang (1985), which hypothesized subjective well-being to be comprised of four dimensions of including happiness, congruence, positive affect, and negative affect. The CFA of the scale using the maximum likelihood method in a large American sample yielded subjective happiness as a second-order factor with happiness, positive affect, negative affect, and congruence as the first-order factors. The first order loading ranged from -.64 to .89 and the measurement model demonstrated a good fit to the data. Furthermore, the authors established its measurement invariance across age and gender.

Lyubomirsky and Lepper (1999) developed a 4-item measure named *Subjective Happiness Scale* that was capable of measuring the extent to which a person was happy or not. The principal component analysis of the initial 13 items retained four items and suggested a single factor solution in a sample of college students. *Subjective Happiness Scale* was validated on a sample of 2732 adults. Lyubomirsky and Lepper (1999) indicated that this scale had a high Cronbach alpha value (.87) and test-retest reliability of three weeks was .79. The evidence for the convergent validity of the *Subjective Happiness Scale* was established as Lyubomirsky and Lepper (1999) found it to be positively related with self-esteem (measured through *Rosenberg Self-Esteem Scale*; Rosenberg, 1965), and optimism (measured through *Life Orientation Test*; Scheier & Carver, 1985).

The preceding review of the literature suggests that a complete perspective on psychological well-being needs the amalgamation of both hedonic and eudaimonic dimensions into one construct that could be reliably assessed. The construct of flourishing involves both hedonic and eudaimonic dimensions and can be conceptualized as the modern conception of well-being as it refers to a more global view of well-being, which not only covers life satisfaction but also includes personal growth, sense of purpose, self-acceptance, self-esteem, and competence (Keyes, 2006).

Built on the notion for assimilating the viewpoints of earlier assessments of well-being, Diener et al. (2009, 2010) developed the *Flourishing Scale* (*Flr.S*). The Flr.S is a mixture of psychological, emotional, and social well-being that includes the meaning, happiness, purpose in life, engagement, mastery, personal growth, being involved in one's work, being optimistic, and positive social relations with others. The Flr.S assesses core aspects of psychosocial functioning such as the fulfillment of competence and affiliation needs and self-acceptance as well as the ownership of psychological wealth such as engagement and flow (Diener et al., 2010; Huppert & So, 2013).

The Flr.S comprises of eight items. Diener et al. (2009, 2010) explored the factor structure of the Flr.S through EFA where the factors were extracted through the principal axis factoring in a sample of 689 university students of the USA. The EFA revealed a single factor solution where item loading ranged from .61 to .77. This single factor structure accounted for 53% variance. Cronbach's alpha of the scale was .80. The Flr.S was found to be positively related to *Satisfaction with Life Scale* (Diener et al., 1985), Ryan and Deci's *Basic Need Satisfaction in General Scale* (2000), and Ryff's *Scale of Psychological Well-being* (2008). This established the evidence for the construct validity of the Flr.S.

Flr.S is one of the most widely used measures for assessing psychological well-being around the globe. It has also been translated into different languages i.e. Spanish (Checa, Perales & Espejo, 2018), Italian (Giuntoli, Ceccarini, Sica & Caudek, 2017), French (Villieux, Sovet, Jung & Guilbert, 2016), Chinese (Sumi, 2014), Brazilian (da Fonseca et al., 2015), and Portuguese (Silva & Caetano, 2013). Across all the translations of the Flr.S, the same factor structure has been confirmed as was reported for the original English version. In the present study, the Flr.S was not only translated into Urdu language but it was also validated on a large sample of students. This validation will allow researchers to use Flr.S within Pakistan or in other areas of the world where people speak Urdu.

Numerous studies have examined the gender differences in flourishing, for instance, Keyes and Simoes (2012) found a high level of flourishing in females as compared to males. Hone, Jarden and Schofield (2014) also reported a higher level of flourishing in women as compared to men. Contrary to the aforementioned results, some studies also reported non-significant gender differences in flourishing (Howell & Buro, 2015; Tang, Duan, Wang & Liu, 2016). Howell and Buro (2015) asserted that owing to the inconsistent findings concerning gender differences on the Flr.S, therefore, it needed further examination.

When measuring gender differences, it is important to determine whether Flr.S measures the same latent structure across gender. The present study is an empirical attempt to bridging this gap as it has explored whether the measurement structure underlying Flr.S is comparable across gender. To the best of our knowledge, it is the first study that has explored the latent structure of Flr.S across gender to establish evidence for its invariance across the two genders. For this

purpose, the CFA was performed to assure whether the original single-factor structure of Flr.S can be replicated in the Pakistani sample (N = 574). Secondly, multi-group CFAs were performed to measure invariance across gender and various models have been tested for establishing configural, metric, and scalar invariance of the scale. Finally, latent mean scores on the Flr.S were compared across both the genders.

To examine the construct validity of the Urdu translated Flr.S, the present study has also examined the pattern of relationships of flourishing with positive and negative core self-evaluation. Core-self-evaluation (CSE) is considered an overall perception of an individual's capability and worth as a human being (Judge et al., 1998) and it comprises of four components including self-esteem (the overall evaluation of one's worth), emotional stability (the ability to feel composed and protected), generalized self-efficacy (assessment of one's ability to effectively accomplish a variety of tasks), and locus of control (the faith that events in the lives of individuals come as a result of their actions or because of powerful others or fate). According to CSE theory, these four faces of personality unite to illuminate an individual's overall judgment of the worth that s/he has as a person (Judge, Locke, Durham & Kluger, 1998).

Positive core self-evaluation (PCSE) may act as a protective factor and contribute positively to psychological well-being. It is a broad latent trait and individual high on the PCSE is likely to be optimistic, composed, efficient, and self-assured. People having positive self-evaluations tend to be more emotionally stable, are more self-efficacious, have a higher degree of self-esteem, and demonstrate an internal locus of control (Judge & Bono, 2001). People's evaluations about their selves determine what they can do and what they can become; this self-evaluation contributes to better psychological and physical health. It promotes healthy functioning such as increasing coping ability, success, and satisfaction with life (Mann, Hosman, Schaalma & De Vries, 2004).

In contrast, people with negative core self-evaluation (NCSE) have poor self-esteem and unstable self-concept. They are emotionally unstable and have an external locus of control, which can play a significant role in the development of an array of mental and social problems (Mann et al., 2004). Judge, Erez, Bono and Thoresen (2003) suggested that a high level of psychological well-being is an outcome of PCSE because people with positive self-evaluations experience more positive emotions.

Zimmerman (2000) asserted that self-esteem and self-efficacy could explain significant variance in psychological well-being and they are key elements of psychological health. Brown (1998) argued that feeling good and mentally healthy is a basic human need that can be fulfilled when one has positive self-evaluation. Roddenberry and Renk (2010) found that people with an external locus of control had poor psychological health than that of their counterparts with an internal locus of control. Additionally, people with a high degree of self-esteem were less likely to be mentally ill as compared to their counterparts having a low degree of low self-esteem. These pieces of research evidence are suggesting that flourishing should have a positive relationship with the PCSE and a negative relationship with the NCSE.

METHOD

The present research comprised of two studies. Study 1 involved the translation of the Flr.S into the Urdu language, its cross-language validation, and assessment of its psychometric properties. Study 2 involved the assessment of factorial validity and measurement invariance of the Flr.S across gender.

STUDY 1

Study 1 comprised of two phases. Phase I involved the Urdu translation of Flr.S whereas phase II aimed at establishing the cross-language validation of the Flr.S.

Phase I: Translation and validation of the Flourishing Scale

This scale is free to use and formal permission for its translation into the Urdu language was sought from the author. In the first phase, the backward translation procedure (Brislin, 1986) was adopted. This procedure was further divided into four steps. The first step was a forward translation, which involved translation from English to the Urdu by following a parallel back-translation procedure. This step was performed to create conceptually equivalent translations for the culture of the target language.

Three bilingual experts (two assistant professors and one lecturer) who were native speakers of the target language and had fluency in English performed forward translations,

emphasizing the conceptual equivalence. While translation, they monitored that the translation and contents associated with semantic features of the original version must be maintained. The second step was the settlement of items for attaining the finest translated items, three independent Urdu translated versions were evaluated through a committee approach for assessing the conjectural consistency among items. They were requested to analyze each item and select one translated version which delivered the best sense of meaning for each item. The committee members made their assessments item-by-item and selected the best translation for each item. This gave us the finalized forward Urdu translations that comprised of best-translated items finalized through the agreement of all bilingual experts. In the next step, the back translation was performed. This step was planned to determine the theoretical equivalence of the finalized forward translation and the original version of the Flr.S Therefore, three bilingual experts who were unfamiliar with the original version of the scale were approached to translate the Urdu translated version of the Flr.S back into English. Thus, three independent English translations of Urdu translated Flr.S were achieved. Finally, the three bilingual expert's committee (two assistant professors and one lecturer) reviewed and compared the back-translated version of Flr.S with the original version. After the agreement on the translated version, to ensure the accuracy of Urdu translation with the original scale, the final version was sent to the original designer (Diener & Biswas-Diener, 2008), who finally approved the Urdu version.

Phase II: Cross-language validation

Cross-language validation of Urdu Flr.S was undertaken by comparing the Urdu translated version with the original English version of the scale. This step aided in establishing the excellence of the Urdu translated version to assess its empirical correspondence with the original English version. For this purpose, four groups of participants were given Flr.S in English to English, English to Urdu, Urdu to Urdu and Urdu to English orders twice with an interval of one week. Firstly, two groups were given Urdu and two groups were given English version of the Flr.S. Subsequently, after one week, all four groups were given Flr.S again but this time two of the groups received the same language version and two received different language version of the scale. The participants' assignment to these four groups was random. The purpose behind this step was to measure the incongruity or similarity between Urdu and English versions of scales. The sample was distributed in such a way as to control the learning effect that might occur due to the testing of original and translated versions. This empirical equality was calculated by finding the correlations of test and retest phase of a week.

Sample

For cross-language validation, the sample of (N = 60) students was randomly selected from the University of Sargodha. The minimum age of students was 20 (range 20-35, M = 25.3, SD = 5.3) and they all were bilingual in English and Urdu. Further, the sample was grouped into four conditions. Each group comprised of 15 participants in the test and retest phase.

Instruments

Flourishing Scale was constructed by Diener and Biswas-Diener (2008) and was used to assess the psychological well-being. It is a brief 8-item scale that helps in measuring respondents' self-perceived success in imperative areas of life such as purpose, self-esteem, relationships, and optimism. The items of scales were scored by using a 7-point Likert scale ranging from 1 = strongly agree to 7 = strongly disagree. Diener and Biswas-Diener (2008) reported high reliabilities (α = .87, test-retest of a week = .76) and convergence validity (.78) of Flr.S with *Satisfaction with Life Scale* (Diener et al., 1985).

Procedure

This study was designed to assess the validation of translated versions of scales which were administrated twice to the four groups of bilingual Pakistani married people in Urdu-Urdu, English-English, English-Urdu and Urdu-English sequence.

The tests were applied to participants in two settings. In the first trial, two groups were given Urdu versions and two were given English versions of both scales. In the second trial after 7 days, the same 60 participants were contacted to make their responses again, but in this trial, two groups were

given the same language version but the other two groups were given the opposite language version. This procedure was used to identify the point of equivalence or discrepancy between Urdu and English versions of scales. This empirical equivalence was evaluated by finding the correlations of test and retest phase of a week, which were depicted in Table 1.

STUDY 2: FACTORIAL VALIDITY

Since our data were normally distributed, we conducted CFA with maximum likelihood estimation for the confirmation of factor structure and measurement model of the Urdu Flr.S. For that purpose, various fit indices were examined and the fit criteria were established with the help of Root Mean Square Error of Approximation (RMSEA≤.50; the smaller is better); Normed Fit Index (NFI>.90); Comparative Fit Index (CFI>.90) (1990) and Goodness of Fit Index (GFI>.90). Table 3 depicts the model fit indices results and factor loading of CFA for Urdu translated Flr.S.

Sample

The sample (N = 574) of the present study was selected through cluster sampling. Since there was the probability of unidentified internal heterogeneity in the population of university students, therefore, the present study adopted a cluster sampling technique. Out of 36 teaching departments of the University of Sargodha (Main Campus), 18 departments were randomly selected as clusters through draw method. In that way, 50% of the entire population was considered as part of the study to make the sample more representative. In the next step, all students of BS (Honors, Semester VII) programs in various academic disciplines were selected. The sample included 235 males and 339 females with an age range of 18 to 25 years (M = 21, SD = 1.8 years).

Instrument

The 8-item Urdu version of the Flr.S was used to assess flourishing on on a 7-point Likert scale. The test-retest reliability of this scale over one week was .93.

Urdu translated version of the *Core Self-evaluation Scale* (Zia, 2016) was used to measure the core self-evaluations of the respondents. It comprises of 12 items which are rated on a 5-point Likert scale with the response rate from 1 = strongly disagree to 5 = strongly agree. Item no. 2, 4, 6, 8, 10, and 12 measure NCSE and item no. 1, 3, 5, 7, 9, and 11 measure PCSE. The internal consistency of the scale was satisfactory (α = .84; Zia, 2016).

Procedure

Through random sampling, participants were personally approached in their classrooms. They were directed about the purpose of the study and were briefed about the relevant response format. Participant's queries were resolved and then they were requested to respond to each statement honestly. They were guaranteed the privacy and confidentiality of the information they provided. For scale completion, there was no time limit. Six hundred participants returned the scales while 574 scales were complete in all aspects and suitable for additional data analysis. Participants were appreciated at the end for their help and support.

RESULTS

Cross-Language Validation

Table 1 shows the correlations between all test-retest phases of Flr.S. Results depicted significant correlations among translated versions, which ranged from .93 to .97. Generally, results in Table 1 depicted empirical equivalence of Urdu translated version of Flr.S to its English version. Table 1 also portrays that the highest correlations existed between English-Urdu and Urdu-English versions.

Descriptive statistics of the Urdu Flr.S

Table 2 displays the descriptive statistics of all items of the Flr.S in terms of the entire sample and by the gender. The ratio of skewness coefficients to their standard error and the ratio of kurtosis coefficients to their standard error remained less than 3, which indicated that in the large sample, the distribution of items did not significantly deviate from normality (Brookshier & Boyd, 2016; Tabachnick & Fidell, 2003, 2007).

Table 1 – Correlations between Urdu and English versions of Flr.S (N = 60)

Scales	r
Test-retest Urdu-Urdu	.93***
Test-retest Urdu-English	.95***
Test-retest English-English	.94***
Test-retest English-Urdu	.97***

***p<.001.

Table 2 – Descriptive statistics for items of Flr.S (N = 574)

	Total		Males		Females	
Item	M±SD	Sk[a]	M±SD	Sk[b]	M±SD	Sk[c]
1	5.16±1.69	−1.20	5.16±1.7	−1.07	5.16±1.6	−1.31
2	5.52±1.33	−1.59	5.64±1.3	−1.63	5.44±1.3	−1.58
3	5.63±1.32	−1.54	5.61±1.3	−1.43	5.65±1.3	−1.63
4	6.02±1.13	−1.85	5.91±1.2	−1.48	6.10±1.0	−2.19
5	5.84±1.18	−1.60	5.85±1.1	−1.42	5.84±1.0	−1.72
6	5.84±1.27	−1.75	5.67±1.4	−1.41	5.96±1.1	−2.06
7	6.10±1.12	−2.02	6.05±1.0	−1.55	6.13±1.1	−2.33
8	5.92±1.17	−1.75	5.85±1.2	−1.46	5.97±1.1	−2.00

Note. [a] Standard error of skewness = .10; [b] Standard error of skewness = .15; [c] Standard error of skewness = .15.

Internal consistency

The Urdu Flr.S demonstrated a satisfactory level of internal consistency. The Cronbach's alpha of the scale for the whole sample was .80. Across the two genders, the Urdu Flr.S was also found to be quite reliable. For the males' sample, its Cronbach's alpha was .80 whereas the same for the females' sample was .79. Both subscales of the *Core Self-evaluation Scale* were also found to be internally consistent.

Factorial validity

To determine whether the original factor structure of the Flr.S can be replicated in the sample of the current study, the CFA was performed using the maximum likelihood estimation because our data were normally distributed. The results showed that the proposed model showed a very good model fit. The results of CFA indicated that the chi square to *df* ratio was 2.3 (χ^2 = 33, *df* = 14, *p*<.05), and other indices of model fit also showed a good model fit between the data and the model. The values of CFI, GFI, and NFI were all above .90 and hence met the criteria of fit indices. The value of RMSEA was .04 with a non-significant p-value (*p* = .48) with 95% CI (LL = .02, UL = .07) and the standardized RMR was well below the cut-off point of .05. The factor loading ranged from .35 to .74.

The values in Table 3 shows the standardized factor loading of all items of Flr.S in the whole sample, the sample of males only, and the sample of females only. The factor loadings of all items on a single latent factor were ≥.30, which supported the unidimensional structure of the Flr.S in the whole sample as well as in the separate groups of males and females. Furthermore, the fit indices indicated that the data fitted well to the model in the whole sample as well as its subgroups.

Table 4 displays the correlations and covariance of Flr.S computed in the entire sample, which suggests that all items of the Flr.S are positively correlated with one another.

Measurement invariance

Table 5 depicts the invariance analyses across gender for the CFA of the Flr.S through maximum likelihood estimation since our data were normally distributed. For the assessment of configural invariance, the factor structure of the scale was kept the same across both the gender i.e., the same number of factors with the same indicators were specified for both men and women. Results revealed that the data fitted well to both the genders, which provided the evidence that the measure was configuraly invariant across the two genders. For the assessment of metric invariance, the factor loadings of the measure were constrained to be

Table 3 – Standardized factor loadings and fit indices (N = 574)

	Standardized factor loading of indicators								Fit indices of models						
	1	2	3	4	5	6	7	8	χ^2	df	GFI	CFI	NFI	RMSEA	St.RMR
Total	.35	.45	.50	.53	.60	.72	.74	.53	33**	14	.98	.98	.97	.04	.02
Males	.30	.41	.50	.56	.62	.80	.79	.44	37***	14	.97	.97	.95	.07	.03
Females	.46	.52	.48	.47	.58	.65	.70	.63	20	12	.97	.98	.95	.04	.03

Legenda. df = degree of freedom; GFI = Goodness of Fit Index; CFI = Comparative Fit Index; NFI = Normed Fit Index; RMSEA = Root Mean Square Error of Approximation; RMR = Root Mean Residual.
**p<.01, *** p<.001.

Table 4 – Covariance and correlation matrices of Urdu version of Flr.S (N = 574)

	Flr.S1	Flr.S2	Flr.S3	Flr.S4	Flr.S5	Flr.S6	Flr.S7	Flr.S8
Flr.S1	–	.38***	.30***	.20***	.22***	.24***	.26***	.18***
Flr.S2	.87	–	.44***	.26***	.43***	.26***	.37***	.34***
Flr.S3	.69	.78	–	.36***	.43***	.30***	.38***	.22***
Flr.S4	.39	.39	.55	–	.52***	.39***	.36***	.26***
Flr.S5	.45	.52	.68	.71	–	.43***	.42***	.31***
Flr.S6	.53	.45	.51	.57	.66	–	.55***	.41***
Flr.S7	.50	.55	.58	.46	.57	.78	–	.39***
Flr.S8	.37	.54	.35	.35	.43	.61	.52	–

Note. Below the diagonal is the covariance matrix.
****p*<.001.

Table 5 – Invariance tests for the Flr.S across gender (N = 574)

Model	χ^2	df	CFI	RMSEA	Model Comparison	$\Delta\chi^2$	Δdf	ΔCFI	ΔRMSEA
1. M_1	57.99	28	.976	.043					
2. M_2	69.61	35	.967	.046	2 vs 1	11.62	7	.009	.003
3. M_3	80.80	43	.957	.047	3 vs 1	22.81	15	.019	.004

Legenda. df = degree of freedom; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; M_1 = invariant form model (configural invariance), M_2 = invariant loading model (metric invariance), M_3 = invariant intercept model (scalar invariance).

equal across both genders. Results suggested that the data again fitted well to the model, which provided the evidence for metric invariance. Furthermore, the comparison of configuraly invariant and metrically invariant models revealed nonsignificant chi square difference tests and negligible differences in other measures of fit indices, which were well below the critical value suggested by Chen (2007). This indicated that data fitted to both models 1 and 2 equally well. Finally, the scalar invariance was estimated by constraining the factor loadings as well as intercepts of the measure to be equal across the two genders. This scalarly invariant model also demonstrated excellent fit to the data. Moreover, the comparison of scalarly invariant models with the configuraly invariant models suggested

non-significant chi square difference tests with negligible differences in CFI and RMSEA values, which were quite small as compared to the critical values suggested by Chen. This demonstrated full scalar invariance of the Flr.S across both the gender and revealed that both scalarly invariant models and the configuraly invariant models fitted to the data equally well.

Latent mean differences

To estimate the differences in the latent means of males and females on the Flr.S, females were selected as the reference group and its latent mean was fixed to zero. However, the latent mean of the males was estimated. The results revealed that the latent means of males and females did not differ significantly from each other (CR = 1.45, p = .15).

Construct validity

Table 6 shows the means, standard deviations, and reliability coefficients of Flr.S and CSE scales. There is no evidence of the restricted range in the scores. The ratio between the values of skewness and its standard error is less than 2, which suggests that the variables were symmetrical in their distribution. Table 6 provides the evidence for the convergent validity of the Flr.S as it indicates a significant positive correlation between Flr.S and PCSE and a significant negative correlation of Flr.S with NCSE. Furthermore, the relationship between flourishing and positive core self-evaluation was stronger than the relationship between flourishing and NCSE. These results provide evidence for the construct validity of Urdu translated Flr.S.

DISCUSSION

The primary purpose of the present study was to translate and validate Flr.S. for the Pakistani population. It also intended to establish the empirical evidence for the measurement invariance of the Urdu Flr.S across gender. Overall, results showed a single-factor structure of Flr.S with a reasonable level of internal consistency and temporal stability.

The findings of the present study indicate that the Urdu translated version of Flr.S has an adequate level of reliability and construct validity for the Pakistani population. The CFA of the Urdu Flr.S indicated that the single factor solution demonstrated a very good fit to the data, which is consistent with the pertinent literature. The original English version of the Flr.S and its various translations have confirmed the single factor structure of the scale and the present research was no exception. Thus, we have established the factorial validity of the Urdu Flr.S and empirically demonstrated its factorial equivalence with the original English version as well as the

Table 6 – Descriptive statistics, alpha reliability, and correlations of the scales of the present study (N = 574)

Variables	M	SD	α	2	3	Range		Sk[a]
						Actual	Potential	
1. Flr.S	46	6.6	.80	.39***	–.18***	14-56	8-56	–.05
2. PCSE	22	3.5	.83	–	–.28***	9-30	6-30	–.06
3. NCSE	18	3.5	.79	–	–	8-30	6-30	–.03

Legenda. Flr.S = Flourishing Scale; PCSE = Positive Core self-evaluation subscale of Core Self-evaluation Scale; NCSE = Negative Core self-evaluation subscale of Core Self-evaluation Scale.

Note. [a] Standard error of skewness = .10; ***p<.001.

other translated versions of the scale. The brevity and simple factor structure make the Flr.S one of the most frequently used instruments for the measurement of well-being, specifically when investigators have a limited period or they want to administer a battery of instruments in a single study.

Albeit the Flr.S has been used in varied cultural and linguistic contexts and with numerous scale versions, none of the studies deliberated on the assessment of measurement invariance of the Flr.S across the gender. Factor uniformity requires to be confirmed statistically because it is crucial for the comparison and clarification of psychological constructs in dissimilar groups such as across the two genders.

We followed Vandenberg and Lance's (2000) proposal to investigate the measurement invariance with successively restrictive phases. In step 1, the configural invariance test was performed to induct a baseline model across groups. The configural invariance test permits factor loadings, intercepts, and residuals were assessed freely. The establishment of a configural invariance test infers that the conceptual framework is the same across groups. If the data is not in the favor of configural invariance test, the measurement invariance test should be dismissed. Our data delivered strong empirical support for the configural invariance as the equivalent factorial structure of the Flr.S was fitted well to both males' and females' sample, therefore we advanced to step 2 in which we tested the metric invariance model. In this model, all factor loadings were embarrassed the same. The metric invariance is a weak invariance test and the establishment of this test means that different groups reacted to the indicators in the same way. Our results supported the metric invariance of the Flr.S. since the model with constrained loadings and the model with freely estimated loadings did not significantly vary in terms of their fit to the data.

In step 3, we performed the scalar invariance model in which the factor loading and indicator intercepts were constrained to be identical across groups. The scalar invariance test is a strong invariance test, which advocates that the measurement model has alike scale across the different groups. Our results provided support for the scalar invariance of the Flr.S as the model with constrained factor loadings and intercepts and the model where these parameters were freely estimated did not significantly differ in terms of their fit to the data. The metric and scalar invariances tests were inspected by measuring the change in the χ^2, CFI, and RMSEA values. The establishment of this test is a prerequisite before the latent means can be compared across groups.

Finally, we did the comparison of latent mean differences between gender. Explicitly, a full scalar invariance model was used as the baseline. To compare latent mean between genders, we constrained the females' group latent mean to 0 and the latent means of the males' group was free to estimate. We used the value of the critical ratio (CR) to calculate latent mean differences. CR is calculated by parameter estimate divided by its standard error, which tests whether the coefficient is significantly different from 0. A CR value larger than 1.96 indicates statistically significant differences in the latent means (Byrne, 2013). A positive CR suggests that the comparison group has a greater latent mean than the reference group. Equally, a negative CR submits that the comparison group's latent mean is lesser than the reference group. In our case, the CR was non-significant, which advocated that the latent means of both males and females on the Flr.S were equivalent.

We also established the evidence for the construct validity of the Urdu Flr.S. The FLr.S was found to be correlated with the PCSE and the NCSE in the expected direction, which provided the evidence for the convergent validity. More specifically, our results suggested that flourishing was positively related to positive core self-evaluation and it was negatively related to the negative core self-evaluation. Furthermore, the Flr.S was more strongly related to the positive core self-evaluation than the negative core self-evaluation, which yielded evidence for the discriminant validity of the Urdu Flr.S.

Pertinent literature supports this pattern of relationships between flourishing and core self-evaluations. For instance, Valkenburg, Peter and Schouten (2006) claimed that evaluations regarding self, such as self-esteem and self-efficacy might influence one's degree of well-being. Bornstein, Davidson, Keyes and Moore (2003) defined psychological well-being as the successful state of performance throughout life, which can lead to the integration of cognitive, physical, and socio-emotional functions that results in productive activities. There are many sources of psychological well-being among which PCSE is an important factor involving self-esteem, self-efficacy, internal locus of control, and emotional stability. Diener and Suh (2000) claimed that these sources can be individual's perceptions about his/her self, appraisals of one's happiness, life happiness, emotional state, personal worthiness, personal value, positive evaluation of one's success, and happiness. All these psychological factors have a positive influence on the psychological well-being of individuals (Bornstein et al., 2003). According to

Kelly (2017), well-being is considered as a positive sense of personal efficacy, strong self-esteem, and emotional stability, which are positively associated with better health outcomes; contrarily, low self-esteem is strongly associated with poor health outcomes.

CONCLUSION

The current research provided details of the Urdu translations of the Flr.S and established the evidence for its factorial structure, internal consistency, temporal stability, and construct validity. Our results on the measurement invariance of the Urdu Flr.S demonstrated the construct of flourishing remains the same across both genders in the Pakistani population. Our findings suggest that the Urdu Flr.S is a psychometrically sound measure that can be reliably used to assess well-being. It has demonstrated a good degree of temporal stability and internal consistency. Furthermore, its single-factor structure was comparable to that of the original English version, which makes it equivalent to the original scale in terms of the factor structure. It also demonstrated construct validity as its positive relationship with the positive core self-evaluations was stronger than its negative correlation with the negative core self-evaluations albeit the statistical significance of both correlations. Our findings on the measurement invariance indicate that the Urdu Flr.S has equivalent structure, similar responses, and comparable latent scores for both males and females, therefore, the construct of flourishing seems to be the same for both the genders.

Limitations and suggestions

The present study was not free of weaknesses. This study was based on a sample of university undergraduate students that could not be a true representative of the general population. Forthcoming research with longitudinal design and the larger, more diverse sample will be more valuable to continue the research of the measurement invariance of the 8-item version of the Flr.S. In the present study the invariance across gender was determined, the same procedure should be employed to determine variance across other demographics. Notwithstanding these restrictions, the findings provided support for the use of the Flr.S as a reliable and valid measurement tool for the assessment of well-being across both the genders. In conclusion, even if the current research displayed the factor invariance of the 8-item version of the Flr.S across gender, further cross-cultural studies are essential to well understand the structure and the nature of this construct. This will lead to the growth of knowledge and warrant the generalizability of the Flr.S as well as its applicability in different educational and cultural contexts.

References

BECK, A.T., STEER, R.A. & CARBIN, M.G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review*, *8* (1), 77-100. doi.org/10.1016/0272-7358(88)90050-5

BORNSTEIN, M.H., DAVIDSON, L., KEYES, C.L. & MOORE, K.A. (Eds.) (2003). *Well-being: Positive development across the life course*. Psychology Press.

BRADBURN, N.M. (1969). T*he structure of psychological wellbeing*. Chicago: Aldine.

BRISLIN, R.W. (1986). Research instruments. *Field Methods in Cross-cultural Research*, *8*, 137-164.

BROOKSHIER, A. & BOYD, R. (2016). *SPSS data screening*. Retrieved May 30, 2020 from https://nau.edu/wp-content/uploads/sites/30/COERC-SPSS-Data-Screening-Tutorial.pdf

BYRNE, B.M. (2013). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Psychology Press.

CHECA, I., PERALES, J. & ESPEJO, B. (2018). Spanish validation of the Flourishing Scale in the general population. *Current Psychology*, *37* (1), 949-956. doi.org/10.1007/s12144-017-9581-0

CHEN, F.F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14* (3), 464-504. doi.org/10.1080/10705510701301834

CÔTÉ, J.E. (1997). An empirical test of the identity capital model. *Journal of Adolescence*, *20* (5), 577-597. doi.org/10.1006/jado.1997.0111

da FONSECA, P.N., da SILVA NASCIMENTO, B., BARBOSA, M., GOMES, L.H., VIONE, K.C. & VELOSO GOUVEIA, V. (2015). Flourishing scale: Evidence of its suitability to the Brazilian context. *Social Inquiry into Well-Being*, 1 (2), 33-40.

DIENER, E. & BISWAS-DIENER, R. (2008). *Happiness: Unlocking the mysteries of psychological wealth*. Malden, MA: Blackwell Publishing.

DIENER, E., EMMONS, R.A., LARSEN, R.J. & GRIFFIN, S. (1985). The satisfaction with Life Scale. *Journal of Personality Assessment*, *49* (1), 71-75. doi.org/10.1207/s15327752jpa4901_13

DIENER, E., OISHI, S. & LUCAS, R.E. (2003). Personality, culture, and subjective well-being: Emotional and cognitive evaluations of life. *Annual Review of Psychology*, *54* (1), 403-425. doi.org/10.1146/annurev.psych.54.101601.145056

DIENER, E. & SUH, E.M. (2000). Measuring subjective well-being to compare the quality of life of cultures. In E. Diener & E.M. Suh (Eds.), *Culture and subjective well-being*. Cambridge: MIT Press.

DIENER, E., WIRTZ, D., BISWAS-DIENER, R., TOV, W., KIM-PRIETO, C., CHOI, D. et al. (2009). New measures of well-being. In E. Diener (Ed.), *Assessing well-being: The collected works of Ed Diener*. New York: Russell Sage Foundation. doi: 10.1007/978-90-481-2354-4 12

DIENER, E., WIRTZ, D., TOV, W., KIM-PRIETO, C., CHOI, D., OISHI, S. et al. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research*, *97* (2), 143-156. doi:10.1007/s11205-009-9493-y

GIUNTOLI, L., CECCARINI, F., SICA, C. & CAUDEK, C. (2017). Validation of the Italian versions of the Flourishing Scale and the Scale of Positive and Negative Experience. *Sage Open*, *7* (1). doi.org/10.1177/2158244016682293

HABKE, A.M., HEWITT, P.L., NORTON, G.R. & ASMUNDSON, G. (1997). The Social Phobia and Social Interaction Anxiety Scales: An exploration of the dimensions of social anxiety and sex differences in structure and relations with pathology. *Journal of Psychopathology and Behavioral Assessment*, *19* (1), 21-39. doi.org/10.1007/BF02263227

HERRICK, C.A. & BROWN, H.N. (1998). Underutilization of mental health services by Asian Americans residing in the United States. *Issues in Mental Health Nursing*, *19* (3), 225-240. doi.org/10.1080/016128498249042

HONE, L., JARDEN, A. & SCHOFIELD, G. (2014). Psychometric properties of the Flourishing Scale in a New Zealand sample. *Social Indicators Research*, *119* (2), 1031-1045.

HOWELL, A.J. & BURO, K. (2015). Measuring and predicting student well-being: Further evidence in support of the flourishing scale and the scale of positive and negative experiences. *Social Indicators Research*, *121* (3), 903-915.

HUPPERT, F.A. (2009). Psychological well-being: Evidence regarding its causes and consequences. *Applied Psychology: Health and Well-Being*, *1* (2), 137-164. doi.org/10.1111/j.1758-0854.2009.01008.x

HUPPERT, F.A. & SO, T.T. (2013). Flourishing across Europe: Application of a new conceptual framework for defining well-being. *Social Indicators Research*, *110* (3), 837-861. doi.org/10.1007/s11205-012-0030-z

JUDGE, T.A. & BONO, J.E. (2001). Relationship of core self-evaluations traits - self-esteem, generalized self-efficacy, locus of control, and emotional stability - with job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology*, *86* (1), 80-92.

JUDGE, T.A., EREZ, A., BONO, J.E. & THORESEN, C.J. (2003). The Core Self-evaluation Scale: Development of a measure. *Personnel Psychology*, *56* (2), 303-331. doi.org/10.1111/j.1744-6570.2003.tb00152.x

JUDGE, T.A., LOCKE, E.A., DURHAM, C.C. & KLUGER, A.N. (1998). Dispositional effects on job and life satisfaction: The role of core evaluations. *Journal of Applied Psychology*, *83* (1), 17-34.

KELLY, S. (2017). *The effects of gender role orientation, self-esteem, and self-efficacy on help-seeking behaviors and wellbeing.* [Higher Diploma Final Year Project, Dublin Business School]. hdl.handle.net/10788/3307

KEYES, C.L. (2006). Subjective well-being in mental health and human development research worldwide: An introduction. *Social Indicators Research*, 77 (1), 1-10. doi.org/10.1007/s11205-005-5550-3

KEYES, C.L. & SIMOES, E.J. (2012). To flourish or not: Positive mental health and all-cause mortality. *American Journal of Public Health*, 102 (11), 2164-2172. doi: 10.2105/AJPH.2012.300918

LAWRENCE, R.H. & LIANG, J. (1988). Structural integration of the Affect Balance Scale and the Life Satisfaction Index A: Race, sex, and age differences. *Psychology and Aging, 3* (4), 375-384. doi.org/10.1037/0882-7974.3.4.375

LIANG, J. (1985). A structural integration of the Affect Balance Scale and the Life Satisfaction Index A. *Journal of Gerontology*, *40*, 552-561.

LYUBOMIRSKY, S. & LEPPER, H.S. (1999). A measure of

subjective happiness: Preliminary reliability and construct validation. *Social Indicators Research*, *46* (2), 137-155. doi: 10.1023/A:1006824100041

MANN, M.M., HOSMAN, C.M., SCHAALMA, H.P. & De VRIES, N.K. (2004). Self-esteem in a broad-spectrum approach for mental health promotion. *Health Education Research*, *19* (4), 357-372. doi.org/10.1093/her/cyg041.

NEUGARTEN, B.L., HAVIGHURST, R.J. & TOBIN, S.S. (1961). The measurement of life satisfaction. *Journal of Gerontology*, *16*, 134-143. doi.org/10.1093/geronj/16.2.134

RODDENBERRY, A. & RENK, K. (2010). Locus of control and self-efficacy: Potential mediators of stress, illness, and utilization of health services in college students. *Child Psychiatry and Human Development*, *41*(4), 353-370. doi.org/10.1007/s10578-010-0173-6

ROSENBERG, M. (1965). Rosenberg's Self-esteem Scale (RSE). *Acceptance and Commitment Therapy. Measures Package*, *61* (52), 18-60.

RYAN, R.M. & DECI, E.L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American Psychologist*, *55*, 68-78. doi:10.1037/0003-066X. 55.1.68

RYAN, R.M. & DECI, E.L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology*, *52* (1), 141-166. doi.org/10.1146/annurev.psych.52.1.141

RYFF, C.D. (2008). *Scales of psychological well-being*. The University of Wisconsin, Institute on Aging: The Author.

RYFF, C.D. & KEYES, C.L. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, *69* (4), 719-727. doi.org/10.1037/0022-3514.69.4.719

SCHEIER, M.F. & CARVER, C.S. (1985). Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Health Psychology*, *4* (3), 219-247. doi.org/10.1037/0278-6133.4.3.219

SILVA, A.J. & CAETANO, A. (2013). Validation of the Flourishing Scale and scale of positive and negative experience in Portugal. *Social Indicators Research*, *110* (2), 469-478. doi.org/10.1007/s11205-011-9938-y

SUMI, K. (2014). Reliability and validity of Japanese versions of the Flourishing Scale and the scale of positive and negative experience. *Social Indicators Research*, *118* (2), 601-615. doi.org/10.1007/s11205-013-0432-6

TABACHNICK, B.G. & FIDELL, L.S. (2003, May). *Preparatory data analyses*. Paper presented at the annual meeting of the Western Psychological Association, Vancouver, BC, Canada.

TABACHNICK, B.G. & FIDELL, L.S. (2007). *Using multivariate statistics (5th ed.)*. Boston, MA: Allyn & Bacon.

TANG, X., DUAN, W., WANG, Z. & LIU, T. (2016). Psychometric evaluation of the simplified Chinese version of the Flourishing Scale. *Research on Social Work Practice*, *26* (5), 591-599. doi.org/10.1177/1049731514557832

VALKENBURG, P.M., PETER, J. & SCHOUTEN, A.P. (2006). Friend networking sites and their relationship to adolescents' well-being and social self-esteem. *CyberPsychology & Behavior*, *9* (5), 584-590. doi.org/10.1089/cpb.2006.9.584

VANDENBERG, R.J. & LANCE, C.E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3* (1), 4-70. doi.org/10.1177/109442810031002

VILLIEUX, A., SOVET, L., JUNG, S.C. & GUILBERT, L. (2016). Psychological flourishing: Validation of the French version of the Flourishing Scale and exploration of its relationships with personality traits. *Personality and Individual Differences*, *88*, 1-5. doi.org/10.1016/j.paid.2015.08.027

WATERMAN, A.S. (2007). Doing well: The relationship of identity status to three conceptions of well-being. *Identity: An International Journal of Theory and Research*, *7* (4), 289-307. doi.org/10.1080/15283480701600769

WATERMAN, A.S., SCHWARTZ, S.J., ZAMBOANGA, B.L., RAVERT, R.D., WILLIAMS, M.K., BEDE AGOCHA, V. & BRENT DONNELLAN, M. (2010). The Questionnaire for Eudaimonic Well-Being: Psychometric properties, demographic comparisons, and evidence of validity. *The Journal of Positive Psychology*, *5* (1), 41-61. doi.org/10.1080/17439760903435208

WATSON, D., CLARK, L.A. & TELLEGEN, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54* (6), 1063-1070. doi:10.1037/0022-3514.54.6.1063

WOYCIEKOSKI, C., STENERT, F. & HUTZ, C.S. (2012). Determinantes do bem-estar subjetivo. [Determinants of subjective well-being]. *Psico*, *43* (3), 280-288.

ZIA, S. (2016). *Impact of parental acceptance-rejection and core self-evaluations on the pride and mental health of university students*. Unpublished BS Thesis. Sargodha: University of Sargodha, Pakistan.

ZIMMERMAN, B.J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, *25* (1), 82-91. doi:10.1006/ceps.1999.1016

ZUNG, W.W. (1965). A self-rating depression scale. *Archives of General Psychiatry*, *12* (1), 63-70. doi:10.1001/archpsyc.1965.01720310065008

DOI: 10.26387/bpa.286.3

MMPI-2 and post-traumatic stress disorder predictors: A meta-analysis

Giuseppe Agrusti[1], Paola Tellaroli[1,2], Claudia Scalise[1], Luca Mandolesi[1]

[1] R&D's Department Giunti Psychometrics, Florence
[2] Department of Developmental Psychology and Socialisation, University of Padova

bpa@giuntipsy.com

ABSTRACT. In letteratura sono disponibili numerosi studi che indagano la capacità dell'MMPI-2 di predire la presenza di Disturbo Post-Traumatico da Stress (PTSD) ma i risultati di tali studi non sono tuttavia sempre concordanti. La presente meta-analisi si propone di perseguire due obiettivi: valutare se esistano scale dell'MMPI-2 predittive del PTSD e del *malingering* della sintomatologia di tale disturbo, nonché riuscire a delineare un profilo tipo di soggetti con PTSD e soggetti *Faker* che, di contro, simulano la presenza o esagerano l'intensità del proprio quadro sintomatologico. I metodi usati sono la revisione sistematica e meta-analisi paired e network degli articoli seguendo le linee guida *PRISMA* e i più importanti database elettronici. Il presente lavoro è il primo che analizza le scale cliniche e le scale di validità che sono in grado di profilare lo stile di risposta tipico dei soggetti con PTSD e *Faker*, utile per predire la vulnerabilità dei soggetti al PTSD. Le analisi effettuate confermano che le scale cliniche 1 (Hs), 2 (D), 6 (Pa), 7 (Pt), 8 (Sc), le scale specifiche del PTSD (PK e PS) e le scale di validità (L, K, F, FB, FP) sono capaci di discriminare i soggetti con PTSD dalla popolazione generale.

SUMMARY. There are numerous studies available in literature that examine the capacity of MMPI-2 to predict the presence of Post-Traumatic Stress Disorder (PTSD) but the results of these studies are not always concordant. This meta-analysis has two objectives: to assess whether MMPI-2 scales exist in predicting PTSD and malingering of the disorder, as well as to define a typical profile for PTSD subjects and Faker subjects, who feign or exaggerate the intensity of their symptoms. The methods used are systematic review, pair-wise and network meta-analysis of the articles, following the PRISMA guidelines and the most important electronic databases. This work is the first of its kind to analyse clinical scales and validity scales able to profile response styles typical of subjects with PTSD and Fakers, useful in predicting subjects' vulnerability to PTSD. The analyses performed confirm that clinical scales 1 (Hs), 2 (D), 6 (Pa), 7 (Pt), 8 (Sc), specific PTSD scales (PK and PS) and validity scales (L, K, F, FB, FP) are able to discriminate subjects with PTSD from the general population.

Keywords: *MMPI, MMPI-2, PTSD, Trauma, Fakers, Malingering, Meta-analysis, Scale, Profile*

INTRODUCTION

An individual who is the victim (real or potential) or the spectator of events that threaten his/her own life or the life of others can develop *Post-Traumatic Stress Disorder* (PTSD). Classified as an anxiety disorder in DSM-IV-TR and as a disorder correlated to traumatic and stressful events in DSM-5, the incidence of PTSD is estimated to vary between 1% and 14% in the United States (American Psychiatric Association, 2007). PTSD rates very much depend on the nature of the event that caused the trauma. As observed by the National Institute of Mental Health (NIMH; EpiCentro, Italian National Institute of Health), PTSD can develop, for instance, in 2% of survivors after a natural disaster, in 28% of people involved in a mass terrorist attack, and in 29% of survivors and family members of victims of airplane crashes. There is only one study in Italy (Di Giorgio et al., 2003) that assesses the incidence of the disorder, considering the population affected by the earthquake in San Giuliano di Puglia. This study observes how 14% of the adults interviewed had possible PTSD, while the children and adolescents proved to be the category most at risk (49%).

The DSM-IV-TR criteria to diagnose PTSD will be presented further on, as they are referenced by the studies under examination. A PTSD diagnosis requires the person to have directly experienced or witnessed a traumatic event that threatens his/her physical integrity or the physical integrity of others (such as serious injury, sexual violence, natural catastrophes, war experiences, or serious accidents), which is associated to intense fear and feelings of impotence or horror (Criteria A). PTSD is characterised by three clusters of symptoms: (a) the traumatic event is relived persistently (for example: recurrent, involuntary, and intrusive flashbacks or nightmares); (b) persistent avoidance of trauma-related reminders and negative alterations in trauma-related cognition and mood (for example: trying to avoid unpleasant memories, feeling detached or estranged from others); (c) alterations in arousal (for example: irritable behaviour or problems in concentration, hypervigilance). In addition, this disorder stands out for its high rate of comorbidity, such as depression, panic attacks, substance abuse, dissociative symptoms, and personality disorders. This is why a multidimensional approach is recommended in clinical practice. The *Minnesota Multiphasic Personality Inventory – 2* (*MMPI-2*; Butcher et al., 1995, 2011) is one of the most widely used of the various tools (Greene, 2000) that have been developed to assess psychopathology and personality. This self-reporting tool consists of 567 "True/False" items and numerous specific scales to assess PTSD symptoms and various associated comorbidities (10 clinical scales, 16 supplementary, 15 content, 5 PSY-5, *Personality Psychopathology Five Scale*, and 27 subscales related to components of the content scales, 28 Harris-Lingoes subscales, and 3 subscales). The eight validity scales also allow malingering (Scheibe, Bagby, Miller & Dorian, 2001), or feigning illness, to be identified.

Most research conducted on using the MMPI-2 to assess PTSD focus on using two specifically developed scales to evaluate how the symptoms of this disorder are configured: *PTSD-Keane* (PK; Keane, Malloy & Fairbank, 1984) and the *PTSD-Schlenger* scale (PS; Schlenger & Kulka, 1989). The PK scale was developed by comparing psychiatric patients from the Veterans Administration Department with various Axis I diagnoses who manifested signs of PTSD to those who did not present this clinical condition. This scale was used with another group of patients from the Veterans Hospital and found confirmation in the cross-validation, also known as cross-validity, which consists in verifying the results obtained on a second independent sample of subjects. The PS scale, instead, was built by comparing Vietnam veterans with good emotional adaptation to those who manifested symptoms typical of PTSD. These two scales are independent from each other and can be used simultaneously for better diagnostic classification (Butcher et al., 1995, 2011). There is no unambiguous consensus on the predictive capacity of these two scales. Indeed, various others have suggested that the two scales are able to discriminate general maladjustment and emotional distress from PTSD in the strict sense of the term (Greene, 2000; Moody & Kish, 1989; Wise, 1996).

Another line of studies concentrated on the 10 clinical scales in the MMPI-2 in PTSD patients in order to profile the response styles and peak elevations typical of the disorder. In analysing the clinical scales of veterans with PTSD, there is often significant elevation in scale 2 (D; *Depression*) and scale 8 (Sc; *Schizophrenia*). The first reflects feelings of discouragement, pessimism, desperation, and personality aspects regarding an excessive sense of duty, aspirations to high standards, and the tendency to be intropunitive, while the second measures a wide variety of oddities, unusual experiences, and particular perceptions that are characteristic of how schizophrenia manifests. Profiles with elevations in these two scales are often coded as 28/82 (Fairbank, Keane

& Malloy, 1983; Frueh, Hamner, Cahill, Gold & Hamlin, 2000; Lyons & Wheeler-Cox, 1999; Orr et al., 1990; Talbert et al., 1994; Wilson & Walker, 1990; Wise, 1996). Studies that used the MMPI-2 with Vietnam veterans report peak elevations also in Scale 7 (Pt; *Psychasthenia*), correlated to a general state of anxiety and worry, adherence to high moral standards, self-criticism, and impulse control. This scale often proves to be higher than scale 2 (D; Albrecht et al., 1994; Baldrachi, Hilsenroth, Arsenault, Sloan & Walter, 1999; Forbes, Creamer & McHugh, 1999; Litz et al., 1991; Wetter, Baer, Berry, Robinson & Sumpter, 1993; Weyermann, Norris & Hyer, 1996). This suggests that a typical PTSD profile could be characterised as an 87/78 three-point code with scale 2 (D) following closely. Furthermore, in studies that focused on Gulf War veterans, there were elevations in scale 1 (Hs; *Hypochondria*), which reflects general concern with one's body or self, and scale 8 (Sc), leading back to an 18/81 code (Glenn et al., 2002). Studies on Croatian war veterans (Begic & Jokic-Begic, 2007) show very high elevations in scales 1 (Hs), 2 (D), and 3 (Hy, *Hysteria*); the latter reflects specific physical disorders or agitation, as well as denial of problems in their lives or lack of social anxiety. Validity scales complete these PTSD profiles. Many studies show that patients with PTSD often score higher in the F scale (*Frequency*), which measures the exaggeration of symptoms and detects atypical responses, and score lower in the L (*Lie*) and K (*Correction*) scales, which reflect the tendency to present the most favourable self-image and downplaying a psychological condition, respectively. The wide diversification of MMPI-2 profiles among the studies examined is most likely caused by the wide variability of the sample, the symptoms of the disorder itself, and the traumatic events (Elhai, Gold, Sellers & Dorfman, 2001; Rademaker, Kleber, Meijer & Vermetten, 2009).

There are other lines of research on the use of the MMPI-2 in assessing PTSD in literature that concentrate on patients that feign or exaggerate their symptoms: this phenomenon is known as *malingering*. This technical term indicates deliberately exaggerating or inventing physical or psychological symptoms in order to obtain some external benefit (American Psychiatric Association, 2007). There are many reasons that lead subjects to feign and/or exaggerate PTSD symptoms: for example, the possibility of receiving financial gain; the possibility of receiving treatment; in legal settings, it can be used to obtain the insanity defence and/or a reduced sentence (Elhai, Frueh, Gold, Gold & Hamner, 2000; Frueh et al., 1996; Resnick, 1997). These subjects that deliberately exaggerate for an external gain can be defined as *suspected malingerers*. However, as stated by van Impelen, Merckelbach, Jelicic and Merten (2014) not all subjects that exaggerate/invent symptoms can be defined as malingerers, as not all of them do it for an external benefit. What the tests can show is whether or not the symptoms are being exaggerated but it is not possible to detect the motive behind subjects presenting an exaggerated view of their condition (Boone, 2007). Literature shows that the incidence of *malingering* varies considerably. In reviewing literature by Rogers (2008), it was found that the incidence of *malingering* in forensics varies from 15.7% to 17.4%, with a large standard deviation of 14.4% (Rogers, Salekin, Sewell, Goldstein & Leonard, 1998). On the whole, subjects who feign/invent symptoms of mental illness were observed as often having the tendency to exaggerate in the generalisation. They might feign both specific symptoms of a psychiatric condition and cognitive deficits (in the sense that they might exaggerate the presence of symptoms and score poorly on cognitive tests). In addition, malingerers present their cognitive deficits in an extremely generalised manner more frequently than those who actually suffer from such conditions (Alwes, Clark, Berry & Granacher, 2008; Green, Rosenfeld, Belfi, Rohlehr & Pierson, 2012; Heinze & Purisch, 2001). However, malingerers might also be highly selective in presenting their symptoms or deficits. Indeed, if these subjects are clever or have gleaned information on the symptoms of the disorders, they will also be able to establish their account of the symptoms experienced using signs and symptoms specific to a particular disorder or disability. Given the importance of the phenomenon and the great variety of how the "feigned" symptoms are presented, most researchers and clinics currently agree that the clinical determination of malingering should not rely solely on a single measure and, as such, on a single tool; rather, it should use a series of tools and scales that are able to detect the various feigning strategies (Boone, 2009; Bush, Heilbronner & Ruff, 2014; Bush et al. 2005; Chafetz et al. 2015; Rogers, 2008; Rogers & Bender 2018). In recent years, literature has shown a growing focus on studies that use various tools to detect different feigning strategies used by subjects and how their test scores differ from subjects that actually experience a psychological and medical condition. Most of these studies concentrate on nonclinical samples coached to feign symptoms (*experimental simulators*), that is, subjects who were never diagnosed with any psychopathologies and were taught to fake experiencing the symptoms typical of a specific disorder, following the

criteria validated and standardised for simulation studies (Giromini et al., 2019). To date, there are several multiscale personality inventory tests that include one or more validity indicators designed to detect atypical response styles and exaggeration of symptoms. Of these, the MMPI-2 is likely the most used. Indeed, the MMPI-2 contains numerous scales that specifically recognise symptom exaggeration among people coached to feign them and the presence of actual disorders in subjects who actually have a disorder. Many studies focus on people who exaggerated/feigned having experienced symptoms characteristic of PTSD. Franklin & Thompson (2005), analysing all the studies focusing on using the MMPI-2 in assessing malingering, observed that the most widely used scales and indices are the F scale (*Frequency*); the FB scale (*Frequency-Back*), which examines the tendency to give unusual responses in the second half of the test; the FP scale (*Frequency-Psychopathology*), which measures responses pertinent to psychopathological aspects and infrequently provided by the general population; the F-K index (Gough Dissimulation Index), which indicates the tendency to control responses (*underreporting*) or an extremely high presence of symptoms (*overreporting*). Of these, the F and FB scales are more useful in assessing whether symptoms are being exaggerated. Most studies in literature tend to confirm the efficacy of the F family validity scales in the MMPI-2 (F, F_B e F_P) in differentiating the response styles of subjects who have been coached to exaggerate PTSD symptoms in exchange for monetary compensation (Bagby, Buis & Nicholson, 1995; Bagby & Marshall, 2005; Rogers, Sewell, Martin & Vitacco, 2003; Rogers, Sewell & Salekin, 1994). Instead, other studies suggest that these scales have no true predictive capacity in differentiating fakers from subjects actually suffering from the disorder (Babgy, Marshall & Bacchiochi, 2006; Elhai et al., 2000; Elhai et al., 2001; Elhai et al., 2002; Rogers et al., 2003). There are various explanations for this diversity of opinions in literature. One is that the reduced predictive capacity of validity scales is given by the fact that they assess the general response strategies typical of fakers and not specific to a certain disorder. Another possible reason is the fact that subjects with PTSD often have very serious and varied symptoms. This could lead to assessing peak elevations as untruthful in the F, FB and FP scales and, as a result, also believing that subjects who actually have PTSD are faking (Marshall & Bagby, 2006).

This work integrates into the range of literature on the use of the MMPI-2 to assess PTSD, using a meta-analytical approach in order to verify which MMPI-2 scales are truly useful in evaluating subjects with PTSD. The study has the following objectives: (a) identify the scales that prove to be important in predicting the symptoms typical of subjects suffering from PTSD; (b) identify the scales that allow subjects suffering from PTSD to be distinguished from those who feign/exaggerate the symptoms of the conditions (for the sake of clarity, these subjects shall be called Fakers; for further specification on the subjects from the various studies, see the Appendix 2, tab. A2-2); (c) create prototype profiles of subjects with PTSD and fakers. To achieve these objectives, in accordance with literature, this work will focus on validity scales and indices (L, K, F, F_B, F_P, F-K), as well as 9 of the 10 clinical scales (1, 2, 3, 4, 6, 7, 8, 9, 0). The decision was made not to consider scale 5 (Mf; *Masculinity/Femininity*) given that it requires a different standardisation process for men and women and because it was deemed unnecessary for the purpose of this study.

METHODS

The meta-analysis presented in this study was carried out following the PRISMA guidelines (Liberati et al., 2015).

Eligibility criteria

The eligibility criteria of the studies were presented according to the PICOS model (*participant*, *intervention*, control, *outcome*, *study design*). For the first objective, studies were included if: (a) they measure the presence of PTSD or Fake-PTSD with MMPI-2; (b) they require a control sample for the outcomes of interest; (c) they present the results of the study regarding the outcome (PTSD and Fake-PTSD) and report the characteristics of the sample; (d) they present the results for the outcomes of interest using case-control studies, cohort studies, randomised control trials (RCT); (e) they were published in English or Italian.

For the second objective, studies were included according to less restrictive criteria in terms of the types of studies permitted, extending them even to cross-sectional studies and analyses of case studies.

For both objectives, studies were excluded if: (a) they use qualitative research studies or single cases; (b) they are editorials, conference abstracts, abstracts, reviews; (c) it was not possible to obtain the full text.

Research strategy

Research was carried out using the most important electronic databases (PsycINFO, PsycTESTS, Web of Science, PubMed, EMBASE, CINAHL, Cochrane), also using *Medical Subject Headings* (*MeSH*). The terms used to search for the studies in all the databases were: *Minnesota Multiphasic Personality Inventory*, *MMPI-2, PTSD*, *trauma, Post-traumatic stress disorder, scale, subscale, diagnosis*. The following is the search string used:

(Minnesota Multiphasic Personality Inventory 2 OR MMPI-2) AND *(PTSD OR trauma OR Post-traumatic stress disorder)* AND *(scale OR subscale OR diagnosis)*

The bibliography search was carried out by examining all studies published until 09.17.2019.

Data extraction and quality assessment

Initially, three psychologists (Giuseppe Agrusti, Luca Mandolesi and Claudia Scalise) screened the titles and abstracts. When there was a doubt as to the eligibility of the study, the article was then read in full. While articles were being read in their entirety, the studies were evaluated based on information regarding: (a) population; (b) PTSD assessment tools; (c) control population; (d) results; (e) research design. Two researchers (Giuseppe Agrusti and Paola Tellaroli, see. Appendix 1) independently assessed the quality of the included studies, using the *Newcastle-Ottawa Scale* (*NOS*; Wells, et al., 2012) and its version adapted for cross-sectional studies (Modesti, et al., 2016).

The NOS scale assesses the quality of non-randomised trials, according to three parameters (*selection*, *comparability*, *exposure*) measured by eight items, which differ slightly for case-control and longitudinal studies. At most one point can be assigned to each item on the scale, with the exception of the *comparability* parameter, for which the maximum score is two points. Scores, therefore, can range from 0 to 9. The higher the score, the better the quality of the study. The specific version for cross-sectional studies consists of seven items, which can be assigned a maximum score of 10. In this work, the studies with a score lower than 4 were identified as having a high risk of bias and, therefore, eliminated from the analysis.

The total scores of each study are divided based on the total possible score. Studies with scores >75% were considered as being of *high quality*, those with scores ≥50% as being of *moderate quality*, and studies with scores <50% as being of *low quality*.

Statistical analysis

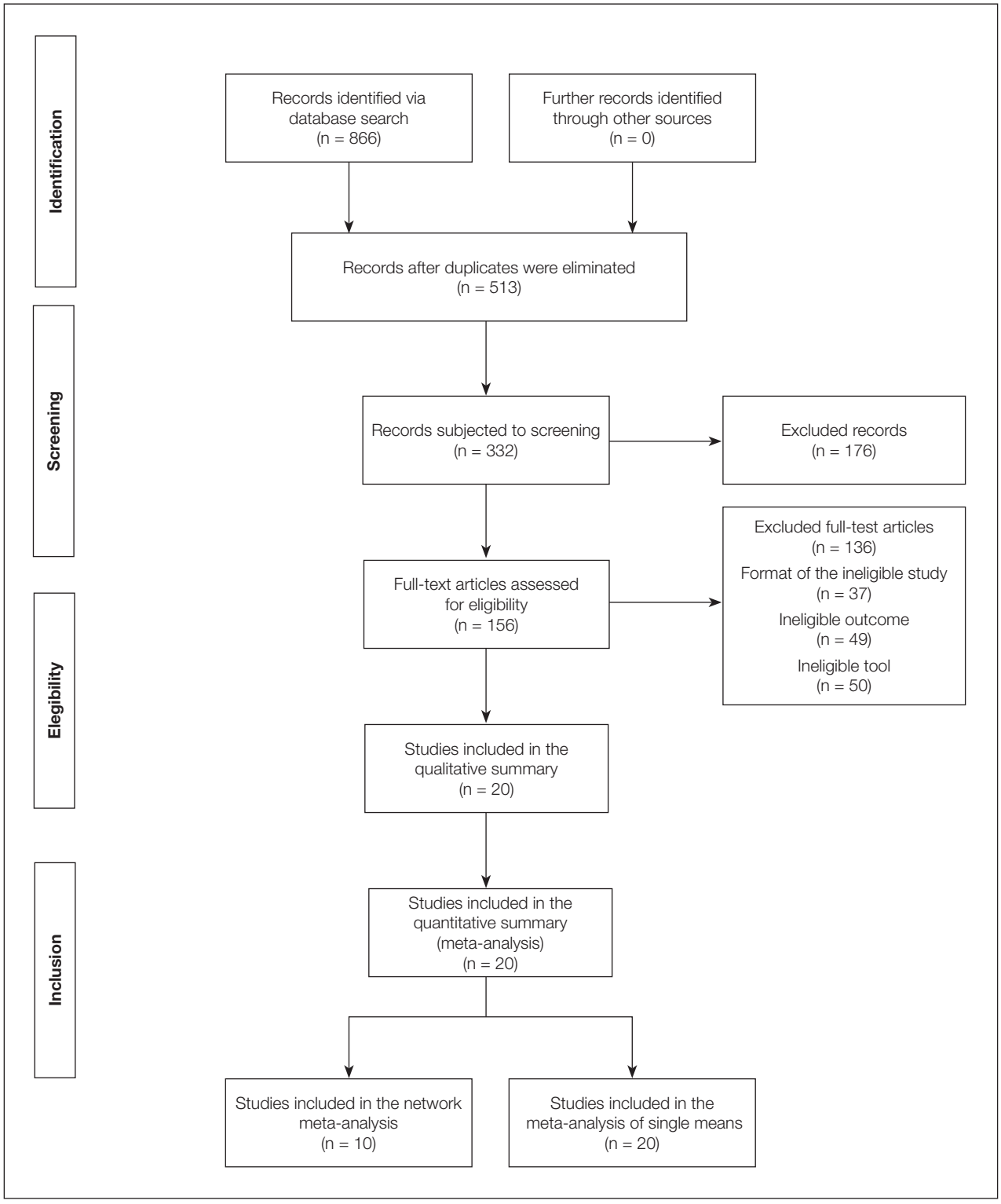
Pairwise and network meta-analyses with a frequentist approach were carried out using R packages (version 3.6.1 for Windows; R Core Team, 2019) *meta* (Schwarzer, 2007) and *netmeta* (Rücker, Krahn, König, Efthimiou & Schwarzer, 2019). Shapiro-Wilk tests for normality were run before aggregating the data if the number of combined studies was greater than 3 (Royston, 1995). Network meta-analyses of the mean difference (*MD*) were carried out for the first objective. A pairwise meta-analysis of the individual means for each scale of interest was carried out for the second objective, using the inverse-variance grouping method. Where there was no standard deviation, two different solutions were adopted: if there were statistics that allowed an estimate to be made, they were used (Higgins & Green, 2011); otherwise, thevalue was taken from another similar study included in the analysis (Furukawa, Barbui, Cipriani, Brambilla & Watanabe, 2006). Both fixed-effect and random-effects models were applied. The first used the inverse-variance weighted estimate, while the second used the DerSimonian-Laird estimate (DerSimonian & Laird, 1986) to take into account heterogeneity, quantified using the I^2 statistic. An I^2 ≥50% value indicated substantial heterogeneity (Higgins & Thompson, 2002), which, in that case, was explored via influence analysis, using the exclusion method (Cooper & Hedges, 1994). Funnel plots and testing for their asymmetry, based on a weighted linear regression of the treatment effect on its standard error (Egger, Smith, Schneider & Minder, 1997), were used to assess possible publication bias if the number of studies was greater than or equal to 10 (Sterne et al., 2011). All p values were two-tailed, with statistical significance set at less than .05.

RESULTS

Selection of the studies

The study selection process is described in Figure 1. Overall, 866 bibliographic citations were identified, which were reduced to 20 studies that meet the eligibility criteria

Figure 1 – Flow chart (PRISMA, 2009)



after several stages of verification. Thereafter, network meta-analyses were carried out for the first objective with 10 of the 13 selected studies, as the remaining (Arbisi, Ben-Porath & McNulty, 2006; Lange, Sullivan & Scott, 2010; Lees-Haley, 1992) had received a score of 3 in the quality assessment and, therefore, were considered as having a high risk of bias. In fact, the Lange study (Lange et al., 2010) presents a sample of solely university students as experimental and control subjects and there is no clear distinction between the two groups, in the Arbisi study (Arbisi et al., 2006), there is a percentage of subjects with PTSD even in the control sample, the experimental subjects in the Lees-Haley study (1992) do not meet criteria A of PTSD diagnosis (considered, therefore, pseudo-PTSD). The 10 selected studies were included as they provided for a clinical sample (subjects with PTSD and/or Fake-PTSD) and a control sample. For our second study, pairwise frequentist meta-analyses were carried out on the individual standardised means for each scale of interest using all 20 selected studies, as we were interested in using the highest possible number of means of the MMPI-2 scores from subjects with PTSD or Fake-PTSD and control samples, regardless of the fact that those samples were compared in the various studies.

Participant characteristics

Participant characteristics are described in the Appendix 2.

Assessment tools

The studies examined used various assessment tools in addition MMPI-2. Only two studies (Albrecht et al., 1994; Litz et al., 1991) use both the MMPI-2 and the MMPI to make a comparison between the two versions in assessing PTSD. The *Structured Clinical Interview for DSM* (*SCID*; Spitzer, Williams & Gibbon, 1987) is often used to diagnose PTSD. The interview is structured to diagnose most Axis I disorders and Axis II personality disorders, according to the DSM (Albrecht et al., 1994; Elhai et al., 2000; Glenn et al., 2002; Litz et al., 2010; Marshall e Bagby, 2006; Scheibe et al., 2001; Tolin et al., 2004; Weyermann et al., 1996). Other tools are associated to the interview, such as the *Mississipi Scale for Combat-Related PTSD* (*M-PTSD*; Keane, Caddel & Taylor, 1988), which is a diagnostic measure for combat-related PTSD (Albrecht et al., 1994; Glenn et al., 2002; Litz et al., 2010; Munley, Bains, Bloem & Busby, 1995; Rademaker et al., 2009; Tolin et al., 2004); the *Clinician Administered PTSD Scale* (*CAPS*; Blake et al., 1995), a clinical assessment scale for PTSD consisting in 30 items administered by a clinic qualified to assess PTSD symptoms, including their frequency and severity (Eakin et al., 2006; Elhai et al., 2000; Forbes et al., 1999; Glenn et al., 2002; Rademaker et al., 2009; Tolin et al., 2004); the *Posttraumatic Stress Disorder Checklist* (*PCL*; Weathers, Litz, Herman, Huska & Keane, 1993), a 17-item self-reporting tool that corresponds to DSM criteria for PTSD, used to measure symptom severity (Eakin et al., 2006; Forbes et al., 1999); the *Davidson Trauma Scale* (*DTS*; Davidson et al., 1997), a 17-item self-reporting tool that corresponds to DSM-IV symptoms and yields a total score and one corresponding to PTSD criteria B, C, and D (Glenn et al., 2002); the *Computerized Diagnostic Interview Schedule* (*C-DIS*; Blouin, Perez & Blouin, 1988), a structured interview to diagnose DSM-III-R disorders, in the Munley et al study (1995) only the part for PTSD diagnosis is used; finally, the *Self-Rating Inventory for PTSD* (*SIP*; Hovens, Bramsen & van der Ploeg, 2000), consisting of 22 items that correspond to clusters B, C, and D of the DSM-IV for PTSD (Rademaker et al., 2009).

These assessment tools are associated with tools to gather information on the traumatic events experienced by the subjects being examined: the *Combat Exposure Scale* (*CES*; Keane, Wolfe & Taylor, 1987), a 7-item tool with Likert scale, which measures the level of wartime stressors experienced (Albrecht et al., 1994; Forbes et al., 1999; Litz et al., 2010; Munley et al., 1995); the *Life Events Checklist* (*LEC*; Blake et al., 1995), a measure of exposure to potentially traumatic events, developed in conjunction with the CAPS to facilitate PTSD diagnosis (Eakin et al., 2006); finally, the *Impact of Event Scale* (*IES*; Horowitz et al., 1979), a 15-item set to assess the amount of distress associated to a specific event (Elhai et al., 2004).

To assess PTSD comorbidity, Glenn et al. (2002) use the *Cook-Medley Hostility Scale* (*Cook-Mendeley*; Barefoot, Dodge, Peterson, Dahlstromwi & Williams, 1989), an abbreviated form of the original scale consisting of 27 items to measure cynicism, hostility, and aggression; the *Beck Depression Inventory* (*BDI*; Beck & Steer, 1987; Beck, Steer & Garbin, 1988), a 21-item self-reporting tool that measures the general severity of depressive symptoms; the *State-Trait*

Anxiety Inventory (*STAI*; Spielberger, 1983), a 40-item self-reporting tool that measures state and trait anxiety.

Some studies assess overall functioning of the subjects examined (Munley et al., 1995; Scheibe et al., 2001) using the *Shipley Institute of Living Scale* (*SILS*) to measure intelligence, or the *Global Assessment of Functioning Index* (*GAF*; American Psychiatric Association, 2007) to assess the severity of mental illness and to what degree the symptoms influence the person's daily life on a scale of 0 to 100.

Configuration of the MMPI-2 scales in subjects with PTSD and in fakers

Network meta-analyses were carried out for the studies considered to be of at least moderate quality and with a control group for the 6 recurring clinical scales in literature (Hs; D; Hy; Pd; Pt; Sc; see Fig. 2), validity scales (F, F_B, F_P; see Fig. 3), and a specific PTSD scale (PK; as PS scores were only available for one study; see Fig. 4) to assess which of these scales was more significant in differentiating a subject with PTSD or fakers from the control group.

Clinical scales

Carrying out a network meta-analysis on the clinical scales, clinical scale 1 (Hs) shows a significant difference between the control group and the group with PTSD (*MD* = 20.41, CI 95% = [7.91; 32.90], *k* = 2) and the group of fakers (*MD* = 32.46, CI 95% = [18.18; 46.74], *k* = 3). Instead, when comparing the experimental group (PTSD) and the group of fakers, there is a trend in which the fakers score higher on average than the subjects with PTSD but these scores did not prove significant owing to overlapping confidence intervals. This result could be explained by high heterogeneity (I^2 = 92.5%), likely due to the scarce number of studies taken into consideration (*k* = 4), to the different nationalities of the subjects considered, and to the fact that one of the studies (Rademaker et al., 2009) involves armed peacekeepers and not actual war veterans.

As regards the network meta-analysis of scale 2 (D), there is a similar trend, that is, an evident significant difference between the control group and the group with PTSD (*MD* = 19.82, CI 95% = [11.60; 28.04], *k* = 4) and the group of fakers (*MD* = 28.56, CI 95% = [18.81; 38.31], *k* = 4). Analysing the differences between the experimental group and the group of fakers, there is a similar general trend with fakers scoring higher than subjects with PTSD but it is not possible to confirm that those with the disorder can actually be discriminated from those who are faking. This result could be explained by high heterogeneity (I^2 = 89.2%), which disappears by eliminating the studies by Elhai et al. (2000) and Marshall et al. (2006). It was not possible to formulate an explanation for this heterogeneity from the data in our possession.

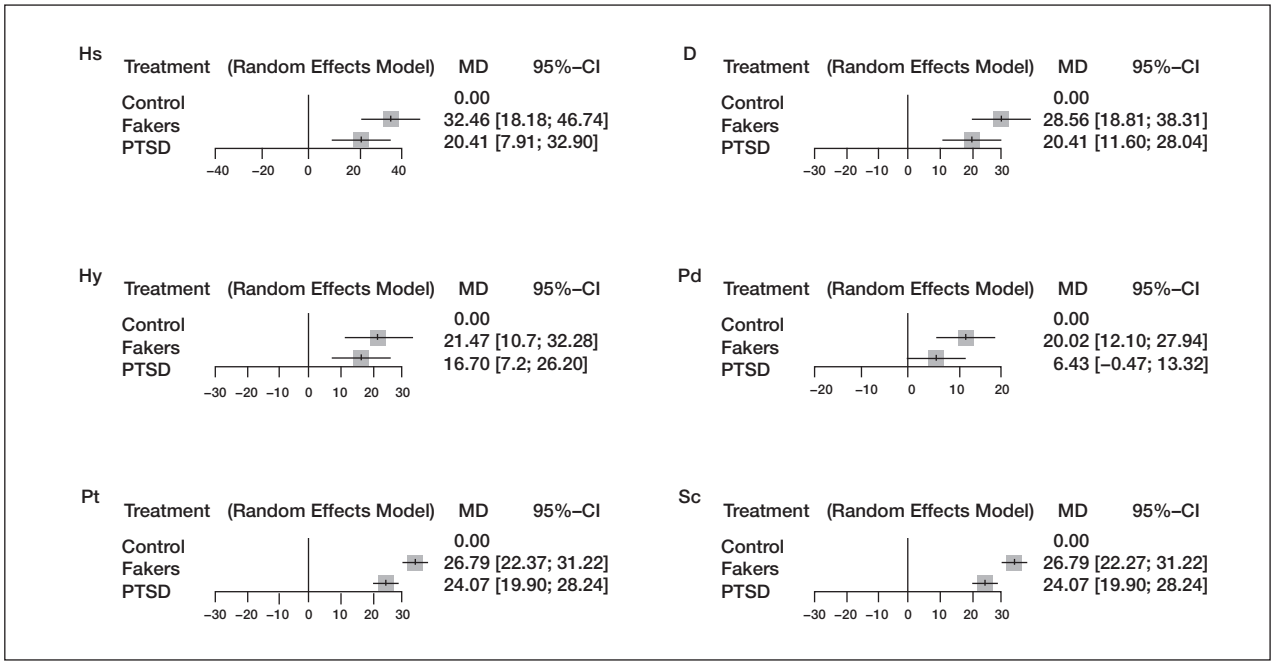
Observing the results for scale 3 (Hy), it is immediately evident that there is a clear significant difference between the control group of subjects with PTSD (*MD* = 16.70, CI 95% = [7.2; 26.20], *k* = 2) and the fakers (*MD* = 21.49, CI 95% = [10.7; 32.28], *k* = 3), but if the results of the two groups are compared against each other, once again, there is no true discriminating capacity (I^2 = 88.6%), likely due to the scarce number of studies (*k* = 4) examined.

Analysing the differences for scale 4 (Pd), there is a significant difference between the control group and the group of fakers (*MD* = 20.2, CI 95% = [12.10; 27.94], *k* = 3), but it is not possible to confirm this as regards the difference between the control group and the group with PTSD (*MD* = 6.43, CI 95% = [–.47; 13.32], *k* = 2). This could, once again, be due to the scarce number of studies available and to the subsequent high heterogeneity (I^2 = 80.5%) in the studies.

For scale 8 (Pt), significant differences stand out between the control group and the fakers (*MD* = 26.79, CI 95% = [22.37; 31.22], *k* = 3) and the group with PTSD (*MD* = 24.07, CI 95% = [19.90; 28.24], *k* = 1), but this difference does not prove significant between the experimental group and the group of fakers. In fact, the trend is similar to the other clinical scales; that is, fakers scored higher than the experimental group but it is not possible to differentiate them from subjects who actually experience the constellation of symptoms typical of the disorder.

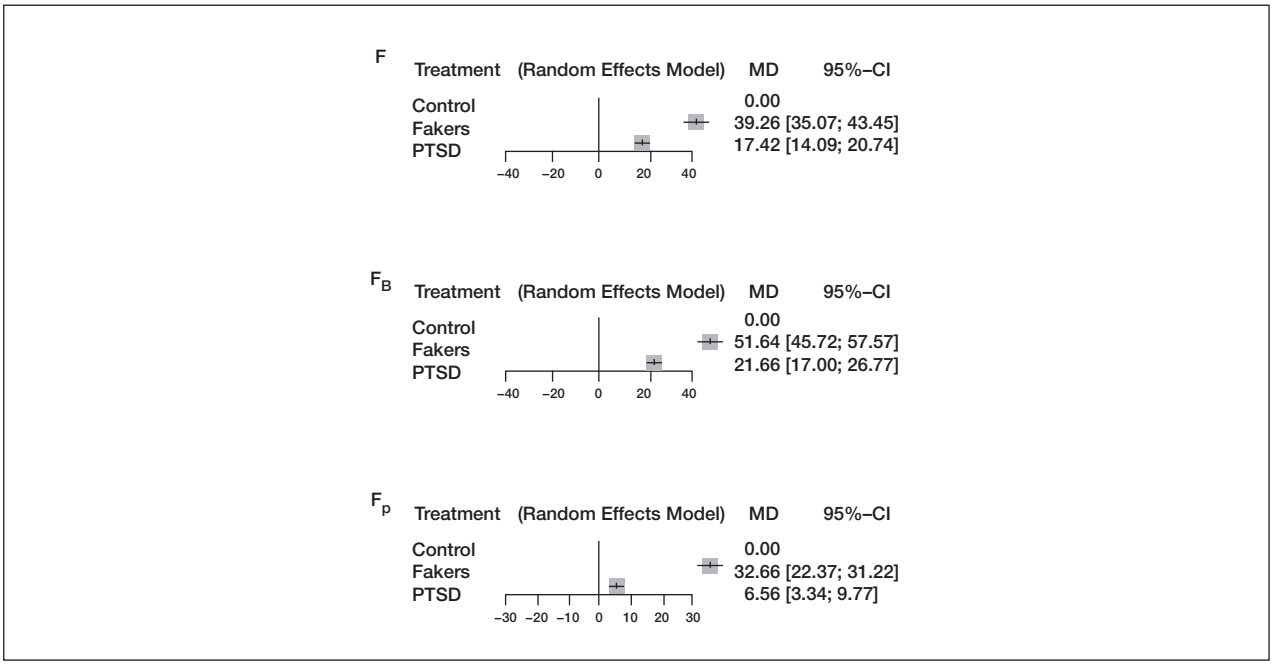
Finally, analysing the results of the network meta-analysis carried out for scale 9 (Sc), there is a significant difference between the control group and the group with PTSD (*MD* = 24.07, CI 95% = [19.90; 28.24], *k* = 1) and the group of fakers (*MD* = 26.79, CI 95% = [22.37; 31.22], *k* = 3), but it is not possible to discriminate the fakers from those who are actually suffering from the disorder. This result could be explained by the scarce number of studies taken into consideration (*k* = 3).

Figure 2 – Forest plot of the clinical scale network meta-analysis



Legenda. Hs = Hypochondria; Hy = Hysteria; Pt = Psychasthenia; D = Depression; Pd = Psychopathic Deviance; Sc= Schizophrenia.

Figure 3 – Forest plot of the validity scale network meta-analysis



Legenda. F = Frequency; F_B = Frequency-Back; F_P = Frequency-Psychopathology.

Validity scales

A network meta-analysis was carried out for the F validity scale, comparing the experimental group (subjects with PTSD) and the group of fakers to the control group. The results highlight the significant differences between the control group and the fakers (*MD* = 39.26, CI 95% = [35.07; 43.45], *k* = 6) and the group with PTSD (*MD* = 17.42, CI 95% = [14.09; 20.74]; *k* = 4; I^2 = 26%).

As regards the F_B validity scale, carrying out a network meta-analysis between the control sample and the other two samples (PTSD and fakers), there is a clear significant difference with the experimental group (*MD* = 21.88, CI 95% = [17.00; 26.77], *k* = 3; I^2 = 41.5%) and the group of fakers (*MD* = 51.64, CI 95% = [45.72; 57.57], *k* = 3), despite moderate heterogeneity (I^2 = 41.5%), due to the scarce number of studies examined.

Finally, the network meta-analysis for the FP scale shows a significant difference between the control sample and the experimental group (*MD* = 6.56, CI 95% = [3.34; 9.77], *k* = 3) and the group of fakers (*MD* = 32.86, CI 95% = [28.47; 37.25], *k* = 5), with low heterogeneity (I^2 = 24.2%), attributable to the study by Elhai et al. (2001), which has mainly female subjects with a history of sexual abuse as its experimental subjects. In fact, by omitting this study, heterogeneity almost completely disappears, confirming a significant difference.

What can be observed from the network meta-analyses is that the validity scales generally have a good discriminating capacity, as the fakers score clearly higher than the control subjects and subjects with PTSD in these scales.

Specific PTSD scale

Given the scarce number of studies presenting data on the *PTSD-Schlenger* (PS) scale, we were only able to analyse the *PTSD-Keane* (PK) scale. Analysing the network meta-analysis of this scale, the trend proved the same, that is, that the group of fakers scored higher compared to the subjects with PTSD. However, in this case there is a significant difference between the two aforementioned groups and the control group but this difference is not significant between the PTSD and faker groups since, as shown in the graph, the confidence intervals tend to overlap (*Fakers MD* = 32.08, CI 95% = [25.15; 39.00]; PTSD *MD* = 20.54, CI 95% = [14.35; 26.73]). This trend could be explained by high heterogeneity (I^2 = 98.1%), due to the scarce number of studies (*k* = 4) and the differences of the samples examined, which differ by gender, age, nationality, and type of war fought (see Fig. 4).

Typical profile of patients with PTSD

Carrying out a pairwise meta-analysis of single means for all 20 studies with MMPI-2 scores of subjects with PTSD and using the Welsh (1948, 1951) coding system, it was possible to sum up all the scores obtained in order to obtain a simple numeric expression that defines a prototype of the typical profile of subjects that present the constellation of PTSD symptoms (see Fig. 5):

82”7163 40’9 / FF_B”F_P-L/K# PKPS”

As can be seen from the Welsh coding system, subjects with PTSD present peak elevations in clinical scales 8 (*M* = 83.33; CI 95% = [79.31; 87.36]) and 2 (*M* = 82.15; CI 95% = [80.20; 84.11]), followed by high elevations in scales 7 (*M* = 79.26; CI 95% = [75.35; 83.18]), 1 (*M* = 76.30; CI 95% = [72.93; 79.67]), 6 (*M* = 76.22; CI 95% = [73.18; 79.25]), 3 (*M* = 74.98; CI 95% = [73.45; 76.51]), 4 (*M* = 71.82; CI 95% = [68.70; 74.94]), 0 (*M* = 70.52; CI 95% = [67.47; 73.58]), and moderate elevation in clinical scale 9 (*M* = 58.48; CI 95% = [55.57; 61.39]). Instead,

Figure 4 – Forest plot of the specific PTSD scale (PK-Keane) network meta-analysis

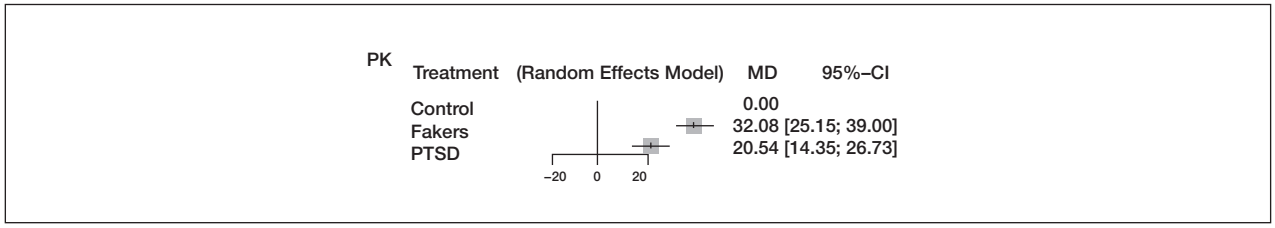
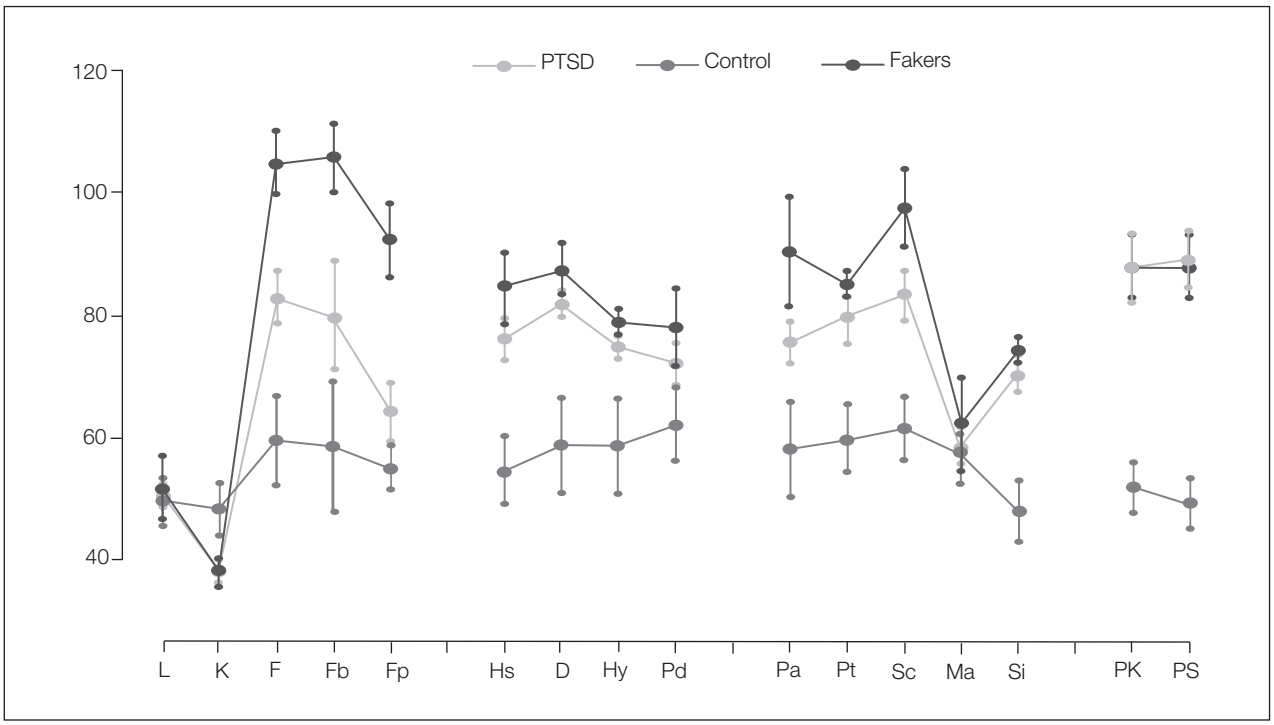


Figure 5 – Graph of the profiles of subjects with PTSD and Faker subjects compared to the control group



Legenda. L = Lie; K = Correction; F = Frequency; FB = Frequency-Back; FP = Frequency-Psychopathology; Hs = Hypochondriasis; D = Depression; Hy = Hysteria; Pd = Psychopathic Deviance; Pa = Paranoia; Pt = Psychasthenia; Sc = Schizophrenia; Ma = Hypomania; Si = Social Introversion; PK = PTSD-Keane; PS = PTSD-Schlenger.

in terms of the validity scales, peak elevations can be noted in the F (M = 83.13; CI 95% = [78.88; 87.39]) and F_B scales (M = 80.06; CI 95% = [70.97; 89.16]), moderate elevation in the F_P scale (M = 64.33; CI 95% = [59.40; 69.26]), mild elevation in the L scale (M = 50.16; CI 95% = [48.44; 51.87]), and low elevation in the K scale (M = 37.53; CI 95% = [36.39; 38.67]). Finally, as regards specific PTSD scales, very high elevation can be observed in the PK (M = 87.95; CI 95% = [82.77; 93.12]) and PS scales (M = 89.27; CI 95% = [84.64; 93.91]).

Typical profile of Fakers

Proceeding with a pairwise meta-analysis of single means for all the studies presenting MMPI-2 scores for fakers, it was possible to sum up the trend of the scores obtained and formulate a simple numerical expression that defines a prototype of the typical profile of fakers (see Fig. 5):

86271”<u>34</u>0’9- F_BF**F_P*L/K# <u>PKPS</u>”

As can be seen from the Welsh coding system, subjects with PTSD present peak elevations in clinical scales 8 (M = 97.53; CI 95% = [90.80; 104.26]) and 6 (M = 90.65; CI 95% = [81.70; 99.60]), followed by very high elevations in scales 2 (M = 87.65; CI 95% = [83.44; 91.86]), 7 (M = 85.31; CI 95% = [83.23; 87.38]), and 1 (M = 84.46; CI 95% = [78.34; 90.59]), high elevations in clinical scales 3 (M = 78.96; CI 95% = [76.96; 80.97]), 4 (M = 78.14; CI 95% = [71.76; 84.53]), and 0 (M = 74.37; CI 95% = [72.48; 76.26]), and moderate elevation in clinical scale 9 (M = 62.16; CI 95% = [54.47; 69.86]). Instead, in terms of the validity scales, peak elevations can

be noted in the F_B (*M* = 105.96; CI 95% = [100.49; 111.43]) and F scales (*M* = 105.18; CI 95% = [99.99; 110.36]), very high elevation in the F_P scale (*M* = 92.63; CI 95% = [86.72; 98.54]), mild elevation in the L scale (*M* = 51.62; CI 95% = [46.92; 56.33]), and low elevation in the K scale (*M* = 37.83; CI 95% = [35.53; 40.13]). Finally, as regards specific PTSD scales, very high elevation can be observed in the PK (*M* = 88.13; CI 95% = [83.45; 92.81]) and PS scales (*M* = 88.30; CI 95% = [83.49; 93.11]).

Risk of bias

The test results for funnel plot asymmetry for the meta-analysis of single means show that for nearly all the scales examined, there is good symmetry at the psychometric level. Only scales 3 (Hy; *p* = .07845) and 9 (Ma; *p* = .01354) could be at risk for publication bias.

DISCUSSION

The results of the analyses carried out suggest that the MMPI-2 is very useful is assessing the severity of PTSD symptoms. Exploring the association between MMPI-2 scores and PTSD symptoms, the peak elevation means of PTSD subject profiles in clinical scales 1 (Hs; *Hypochondriasis*), 2 (D; *Depression*), 6 (Pa; *Paranoia*), 7 (Pt; *Psychasthenia*), and 8 (Sc; *Schizophrenia*) can be observed as being consistent with previous research analysing this association. In fact, should one wish to attempt to describe the typical profile of a subject with PTSD by interpreting the peak elevations of such scales, there could be a concordance with the PTSD symptom clusters. Specifically, elevations in scale 1 (Hs) could reflect psychological reactivity, as well as the presence of general malaise and numerous vague somatic symptoms associated to an increase in symptoms of anxiety typical of PTSD. Elevations in clinical scale 2 (D) would reflect symptoms of depression, often reported by patients affected by PTSD. Indeed, people with elevated scores in this scale usually report weakness, fatigue, low energy; they often have trouble sleeping, a lack of interest in activities, tension; they are seen as being unhappy, pessimistic, and self-critical. Peak elevations in scale 6 (Pa) could be associated with aggression, acting out, hostility. People with elevated scores in this scale, in fact, are often hostile, resentful, argumentative; they are hypersensitive and hyperactive to the actions of others; they are often suspicious and defensive. Elevations, instead, in scale 7 (Pt) can be associated to symptoms of anxiety. People with elevations in these scales tend to be anxious, tense, agitated, and present a lack of concentration. Finally, peak elevations in scale 8 (Sc) can be associated to blunted affect, social alienation, and intrusive and/or dissociative symptoms, two clusters of PTSD symptomatology.

The meta-analyses, furthermore, confirm that specific PTSD scales, particularly the *PTSD-Keane* (PK) scale, are capable of optimally discriminating control subjects from subjects with PTSD. Elevated scores in these scales, in fact, indicate the presence of PTSD symptoms including anxiety, depression, emotional distress, disturbing thoughts, and trouble sleeping.

By analysing the validity scale scores, it can be seen that they are also in line with previous scientific literature. Indeed, they confirm the usefulness of the F family scales (F, F_B e F_P) in discriminating between subjects that actually have the disorder from those feigning/exaggerating the symptoms. Analysing all the validity scales as a whole, the trend is confirmed; that is, subjects with PTSD and fakers present peak elevations in the F scale and low scores in the L and K scales. Subjects with PTSD, in fact, have elevated scores in the F scale, typical of someone experiencing general distress who has had to face an excessive number of psychological problems. Fakers, on the other hand, paint a noticeably exaggerated picture in which they report an extreme number of symptoms that are more than likely not correlated to each other.

In general, from the individual network meta-analyses, the faker group scores for the validity scales are clearly higher than the group with PTSD.

Despite the fact that our analyses show a summary that generally confirms the existing literature, these results must be taken with caution. What was observed from the single means analyses is that even though the profiles of subjects with PTSD and faker subjects are significantly higher compared to the control subjects (indicating that the validity scales, clinical scales, and the two specific PTSD scales have good discriminating capacity), these profiles do not demonstrate good discriminating capacity among themselves. The graph of the profiles (see Fig. 5), indeed, shows that the only scales in which the confidence indexes do not overlap are the F family validity scales. This trend confirms the data in literature that attest to the difficulty in recognising fakers from those actually affected by PTSD due to the vast heterogeneity of

the symptomatology of the disorder itself and the traumatic events.

The results presented above are to be interpreted under certain limitations. First and foremost, all the measures used, with the exception of the PTSD diagnostic tools, are self-reporting tools, which could lead to bias in assessing the symptomatology of the disorder. Secondly, high sample heterogeneity could limit the reliability of the results. Moreover, there is a scarcity of combined studies and high heterogeneity in the diagnostic tools used.

CONCLUSIONS

The presented study is the first of its kind to analyse clinical scales and validity scales able to profile response styles typical of subjects with PTSD and fakers, useful in predicting subjects' vulnerability to PTSD. The results add to current literature assessing the relationship between MMPI-2 and PTSD symptomatology and confirm previous observations, that is, that clinical scales 1 (Hs), 2 (D), 6 (Pa), 7 (Pt), and 8 (Sc), the specific PTSD scales (PK and PS), and the validity scales (L, K, F, FB, FP) are able to discriminate subjects with PTSD from the general population. Furthermore, the usefulness of the F, F_B, and F_P validity scales has been confirmed in discriminating those feigning/exaggerating symptoms from those who actually experience symptoms typical of PTSD.

Implications for practice

An important practical implication of this work is having detected certain specific MMPI-2 clinical scales that tend to elevate in the presence of PTSD symptomatology. This can prove useful in clinical practice to predictively assess PTSD, administering the MMPI-2 longitudinally (for example, upon entry, immediately after a traumatic event) in order to identify which of the clinical scales found to be significant by our meta-analyses are closest to the T score of 65, the ideal level to discriminate the clinical groups from the normative sample of the MMPI-2 (Butcher et al., 1989). It may be advantageous to integrate this assessment with the administration of specific tools for PTSD and any correlated comorbidities, as well as tools capable of assessing the subject's personality structure. This makes it possible to define a profile that is both detailed and useful during the treatment plan.

A secondary reflection suggested by our work regards the possibility of analysing the general trend of validity scales rather than merely considering a single indicator of these scales. By doing so, in fact, the assessment of the subject's response style proves more accurate and allows fakers to be discriminated from those actually experiencing the symptoms of the disorder.

Moreover, given the vast variety of atypical response styles and presentations of "simulated" symptomology, clinics would not need to rely on a single measure and, therefore, a single tool; rather, they would need to use a series of tools and scales with the capacity to detect the various simulation strategies (Boone, 2009; Bush et al., 2005; Bush et al., 2014; Chafetz et al., 2015; Rogers, 2008; Rogers & Bender, 2018). To this end, there are various tools in literature to detect malingering. For example, Smith and Burger (1997) developed the *Structured Inventory of Malingered Symptomatology* (*SIMS*), a 75-item self-reporting tool designed to detect simulated psychopathological conditions and cognitive deficits, including psychosis, neurological disorders, and affective disorders (Widows & Smith, 2004). Moreover, a recent study by Giromini et al. (2019) demonstrated that the joint use of the MMPI-2 and *Inventory of Problems-29* (*IOP-29*; Viglione, Giromini & Landis, 2017) in assessing the credibility of depression-related symptoms can be a useful indicator of incremental validity as compared to exclusively using the MMPI-2 validity scales.

Implications for research

Future research might focus on the content and supplementary scales of the MMPI-2, helpful in defining more accurate PTSD profiles that also take into consideration any subtypes of the disorder and the various comorbidities. A further line of research might examine the use of restructured MMPI-2 clinical scales and evaluate whether they can discriminate PTSD symptomatology in the same way as clinical scales. Additionally, to more accurately discriminate malingering, research studies could be structured to enrol not only students as control subjects but also other subjects so as to examine their different scores and cut-offs. Finally, research models might be designed to associate the various items in the MMPI-2 with the PTSD symptom clusters, according to the DSM-5 diagnostic criteria.

References

ALBRECHT, N.N., TALBERT, F.S., ALBRECHT, J.W., BOUDEWYNS, P.A., HYER, L.A., TOUZE, J. et al. (1994). A comparison of MMPI and MMPI-2 in PTSD assessment. *Journal of Clinical Psychology*, *50*, 578-585.

ALWES, Y.R., CLARK, J.A., BERRY, D.T.R. & GRANACHER, R.P. (2008). Screening for feigning in a civil forensic setting. *Journal of Clinical and Experimental Neuropsychology*, *30*, 133-140.

AMERICAN PSYCHIATRIC ASSOCIATION (2000). DSM-IV-TR. *Diagnostic and statistical manual of mental disorders (4th ed., text revision)*. Washington, DC: American Psychiatric Association (It. ed.: V. Andreoli, GB. Cassano & R. Rossi, *DSM-IV-TR. Manuale diagnostico e statistico dei distubi mentali. Text revision*. 2007, Milano: Elsevier).

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).

ARBISI, P.A., BEN-PORATH, Y.S. & McNULTY, J. (2006). The ability of the MMPI-2 to detect feigned PTSD within the context of compensation seeking. *Psychological Services*, *3* (4), 249-261.

BAGBY, R.M., BUIS, T. & NICHOLSON, R.N. (1995). Relative effectiveness of the standard validity scales in detecting fake-bad and fake-good responding: Replication and extension. *Psychological Assessment*, *7*, 84-92.

BAGBY, R.M. & MARSHALL, M.B. (2005). Assessing response bias with the MCMI modifying indices. In R.J. Craig (Ed.), *New directions in interpreting the Millon Clinical Multiaxial Inventory*. New York: John Wiley.

BAGBY, R.M., MARSHALL, M.B. & BACCHIOCHI, J. (2006). The validity and clinical utility of the MMPI-2 malingering depression scale. *Journal of Personality Assessment*, *85*, 304-311.

BALDRACHI, R., HILSENROTH, M., ARSENAULT, L., SLOAN, P. & WALTER, C. (1999). MMPI-2 assessment of varying levels of posttraumatic stress disorder in Vietnam combat veterans. *Journal of Psychopathology and Behavioral Assessment, 21*, 109-116.

BAREFOOT, J.C., DODGE, K.A., PETERSON, B.L., DAHLSTROM, W.G. & WILLIAMS, R.B. (1989). The Cook-Medley Hostility Scale: Item content and ability to predict survival. *Psychosomatic Medicine*, *51* (1), 46-57.

BECK, A.T. & STEER, R.A. (1987). *The Revised Beck Depression Inventory*. San Antonio, TX: Psychological Corporation.

BECK, A.T., STEER, R.A. & GARBIN, M.G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review*, *8*, 77-106.

BEGIC, D. & JOKIC-BEGIC, N. (2007). Heterogeneity of posttraumatic stress disorder symptoms in Croatian war veterans: retrospective study. *Croatian Medical Journal*, 48, 133-139.

BLAKE, D.D., WEATHERS, F.W., NAGY, L.M., KALOUPEK, D.G., GUSMAN, F.D., CHARNEY, D.S. & KEANE, T.M. (1995). The development of a clinician-administered PTSD scale. *Journal of Traumatic Stress*, 8, 75-90.

BLOUIN, A.G., PEREZ, E.L. & BLOUIN, J.H. (1988). Computerized administration of the Diagnostic Interview Schedule. *Psychiatry Research*, *23*, 335-344.

BOONE, K.B. (2007). A reconsideration of the Slick et al. (1999) criteria for malingered neurocognitive dysfunction. In K.B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective*. New York: Guilford.

BOONE, K.B. (2009). The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *The Clinical Neuropsychologist*, *23* (4), 729-741.

BUSH, S.S., HEILBRONNER, R.L. & RUFF, R.M. (2014). Psychological assessment of symptom and performance validity, response bias, and malingering: Official position of the Association for Scientific Advancement in Psychological Injury and Law. *Psychological Injury and Law*, *7*, 197-205.

BUSH, S.S., RUFF, R.M., TROSTER, A., BARTH, J., KOFFLER, S., PLISKIN, N. et al. (2005). Symptom validity assessment: Practice issues and medical necessity. *Archives of Clinical Neuropsychology*, *20* (4), 419-426.

BUTCHER, J.N. & WILLIAMS, C.L. (2014). *Fondamenti per l'interpretazione del MMPI-2 e del MMPI-A (2a ed.)*. Firenze: Giunti Psychometrics.

BUTCHER, J.N., GRAHAM, J.R., BEN-PORATH, Y.S., TELLEGEN, A., DAHLSTROM, W.G. & KAEMMER, B. (1995, 2011). *MMPI®-2. Minnesota Multiphasic Personality Inventory®-2*. (It. ed.: P. Pancheri & S. Sirigatti. Firenze: Giunti Psychometrics).

CHAFETZ, M.D., WILLIAMS, M.A., BEN-PORATH, Y.S., BIANCHINI, K.J., BOONE, K.B., KIRKWOOD, M.W., LARRABEE, G.J. & ORD, J.S. (2015). Official position of the American Academy of clinical neuropsychology Social Security Administration policy on validity testing: Guidance and recommendations for change. *The Clinical Neuropsychologist*, *29* (6), 723-740.

COOPER, H. & HEDGES, L.V. (1994). *The handbook of research synthesis*. Newbury Park, CA: Russell Sage Foundation.

DAVIDSON, J.R.T., BOOK, S.W., COLKET, J.T., TUPLER, L.A., ROTH, S., DAVID, D., HERTZBERG, M., MELLMAN, T., BECKHAM, J.C., SMITH, R.D., DAVISON, R.M., KATZ, R. & FELDMAN, M.E. (1997). Assessment of a new self-rating scale for posttraumatic stress disorder: The Davidson Trauma Scale. *Psychological Medicine*, *27*, 153-160.

DerSIMONIAN, R. & LAIRD, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7, 177-188.

DI GIORGIO, G., FALCIGLIA, G., VELTRO, F., MALINCONICO, A., MUSTILLO, I., PEDE, V., MUSENGA, E. & BINKIN, N. (2003). Effetti sulla salute del terremoto di San Giuliano, 2002. *Bollettino epidemiologico nazionale*. Retrieved from: https://www.epicentro.iss.it/ben/2003/settembre2003/2.

EGGER, M., SMITH, G.D., SCHNEIDER, M. & MINDER, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*, 629-634.

ELHAI, J.D., FRUEH, B.C., GOLD, P.B., GOLD, S.N. & HAMNER, M.B. (2000). Clinical presentations of posttraumatic stress disorder across trauma populations: A comparison of MMPI-2 profiles of combat veterans and adult survivors of child sexual abuse. *Journal of Nervous and Mental Disease*, *188*, 708-713.

ELHAI, J.D., GOLD, P.B., SELLERS, A.H. & DORFMAN, W.I. (2001). The detection of malingered posttraumatic stress disorder with MMPI-2 fake bad indices. *Assessment*, *8*, 221-236.

ELHAI, J.D., NAIFEH, J.A., ZUCKER, I.S., GOLD, S.N., DEITSCH, S.E. & FRUEH, B.C. (2004). Discriminating malingered from genuine civilian posttraumatic stress disorder: A validation of three MMPI-2 infrequency scales (F, Fp, Fptsd). *Assessment*, *11*, 139-144.

ELHAI, J.D., RUGGIERO, K.J., FRUEH, B.C., BECKHAM, J.C., GOLD, P.B. & FELDMAN, M.E. (2002). The Infrequency Posttraumatic Stress Disorder scale (Fptsd) for the MMPI-2: Development and initial validation with veterans presenting with combat-related PTSD. *Journal of Personality Assessment*, *79*, 531-549.

EPICENTRO, ISTITUTO SUPERIORE DI SANITÀ. *Stress post traumatico. Aspetti epidemiologici*. Retrieved from: https://www.epicentro.iss.it/stress/epidemiologia.

FAIRBANK, J.A., KEANE, T.M. & MALLOY, P.F. (1983). Some preliminary data on the psychological characteristics of Vietnam veterans with posttraumatic stress disorders. *Journal of Consulting and Clinical Psychology*, *51*, 912-919.

FORBES, D., CREAMER, M. & McHUGH, T. (1999). MMPI-2 data for Australian Vietnam Veterans with combat-related PTSD. *Journal of Traumatic Stress*, *12*, 371-378.

FRANKLIN, C.L. & THOMPSON, K.E. (2005). Response style and posttraumatic stress disorder (PTSD): A review. *Journal of Trauma & Dissociation*, *6* (3), 105-123.

FRUEH, B.C., HAMNER, M.B., CAHILL, S.P., GOLD, P.B. & HAMLIN, K.L. (2000). Apparent symptom overreporting in combat veterans evaluated for PTSD. *Clinical Psychology Review*, *20*, 853-888.

FRUEH, B.C., SMITH, D.W. & BARKER, S.E. (1996). Compensation seeking status and psychometric assessment of combact veterans seeking treatment for PTSD. *Journal of Traumatic Stress*, *9* (3), 427-439.

FURUKAWA, T.A., BARBUI, C., CIPRIANI, A., BRAMBILLA, P. & WATANABE, N. (2006). Imputing missing standard deviations in meta-analyses can provide accurate results. *Journal of Clinical Epidemiology*; *59*, 7-10.

GIROMINI, L., CARFORA LETTIERI, S., ZIZOLFI, S., ZIZOLFI, D., VIGLIONE, D.J., BRUSADELLI, E., PERFETTI, B., DI CARLO, D.A. & ZENNARO, A. (2019). Beyond rare-symptoms endorsement: A clinical comparison simulation study using the Minnesota Multiphasic Personality Inventory – 2 (MMPI-2) with the Inventory of Problems-29 (IOP-29). *Psychological Injury and Law*. doi: 10.1007/s12207-019-09357-7

GLENN, D.M., BECKHAM, J.C., SAMPSON, W.S., FELDMAN, M.E., HERTZBERG, M.A. & MOORE, S.D. (2002). MMPI-2 profiles of Gulf and Vietnam combat veterans with chronic posttraumatic stress disorder. *Journal of Clinical Psychology*, *58*, 371-381.

GREEN, D., ROSENFELD, B., BELFI, B., ROHLEHR, L. & PIERSON, A. (2012). Use of measures of cognitive effort and feigned psychiatric symptoms with pretrial forensic psychiatric patients. *International Journal of Forensic Mental Health*, *11*, 181-190.

GREENBLATT, R.L. & DAVIS, W.E. (1999). Differential diagnosis of PTSD, schizophrenia, and depression with the MMPI-2. *Journal of Clinical Psychology*, *55* (2), 217-222.

GREENE, R.L. (2000). *The MMPI-2: An interpretive manual (2nd ed.)*. Boston: Allyn & Bacon.

HEINZE, M.C. & PURISCH, A.D. (2001). Beneath the mask: Use of psychological tests to detect and subtype malingering in criminal defendants. *Journal of Forensic Psychology Practice*, *1*, 23-52.

HIGGINS, J.P.T. & GREEN, S. (Eds.) (2011). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 *[updated March 2011]*. Retrieved from: www.handbook.cochrane.org.

HIGGINS, J.P.T., THOMAS, J., CHANDLER, J., CUMPSTON, M., LI, T., PAGE, M.J. & WELCH, V.A. (Eds.) (2019). *Cochrane Handbook for Systematic Reviews of Interventions version 6.0 [updated July 2019]*. Retrieved from: www.training.cochrane.org/handbook.

HIGGINS, J.P.T. & THOMPSON, S.G. (2002). Quantifying

heterogeneity in a meta-analysis. *Stat Med*; *21* (11), 1539-1558.

HOVENS, J.E., BRAMSEN, I. & van der PLOEG, H.M. (2000). *De Zelfinven-tarisatielijst Post traumatische Stressstoornis: ZIL handleiding [Self-report measure for PTSD symptoms: SIP manual]*. Lisse/Leiden, NL: Swets & Zeitlinger.

KEANE, T.M., CADDELL, J.M. & TAYLOR, K.L. (1988). Mississippi scale for combat-related posttraumatic stress disorder: Three studies in reliability and validity. *Journal of Consulting and Clinical Psychology*, *56*, 85-90.

KEANE, T.M., MALLOY, P.F. & FAIRBANK, J.A. (1984). Empirical development of an MMPI subscale for the assessment of combat-related posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology*, *52*, 888-891.

KEANE, T.M., WOLFE, J. & TAYLOR, K.L. (1987). Posttraumatic stress disorder: Evidence for diagnostic validity and methods of psychological assessment. *Journal of Clinical Psychology*, *43*, 32-43.

LANGE, R.T., SULLIVAN, K.A. & SCOTT, C. (2010). Comparison of MMPI-2 and PAI validity indicators to detect feigned depression and PTSD symptom reporting. *Psychiatry Research*, *176*, 229-235.

LARRABEE, G.J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist*, *22*, 666-679. doi: 10.1080/13854040701494987

LEES-HALEY, P.R. (1992). Efficacy of the MMPI-2 validity scales and the MCMI-II modifier scales for detecting spurious PTSD claims: F, F-K, Fake Bad Scale, Ego Strength, Subtle-Obvious subscales, DIS, and DEB. *Journal of Clinical Psychology*, 48, 681-688.

LIBERATI, A., ALTMAN, D.G., TETZLAFF, J., MULROW, C., GØTZSCHE, PC, IOANNIDIS, J.P.A. et al. (2015). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Evidence*; *7* (6), e1000115.

LITZ, B.T., PENK, W.E., WALSH, S., HYER, L., BLAKE, D. D., MARX, B. et al. (1991). Similarities and differences between MMPI and MMPI-2 applications to the assessment of posttraumatic stress disorder. *Journal of Personality Assessment*, *57*, 238-253.

LYONS, J.A. & WHEELER-COX, T. (1999). MMPI, MMPI-2 and PTSD: Overview of scores, scales, and profiles. *Journal of Traumatic Stress, 12*, 175-183.

MARSHALL, M.B. & BAGBY, R.M. (2006). The incremental validity and clinical utility of the MMPI-2 Infrequency Posttraumatic Stress Disorder Scale. *Assessment*, *13* (4), 417-429.

MODESTI, P.A., REBOLDI, G., CAPPUCCIO, F.P., AGYEMANG, C., REMUZZI, G., RAPI, S., PERRUOLO, E. & PARATI, G. (2016). Panethnic differences in blood pressure in Europe: A systematic review and meta-analysis. *PLoS One*, *11* (1), e0147601.

MOODY, D.R. & KISH, G.B. (1989). Clinical meaning of the Keane PTSD Scale. *Journal of Clinical Psychology, 45*, 542-546.

MUNLEY, P.H., BAINS, D.S., BLOEM, W.D. & BUSBY, R.M. (1995). Post-traumatic stress disorder and the MMPI-2. *Journal of Traumatic Stress*, *8*, 171-178.

ORR, S.P., CLAIBORN, J.M., ALTMAN, B., FORGUE, D.F., DE JONG, J.B., PITMAN, R.K. et al. (1990). Psychometric profile of posttraumatic stress disorder, anxious, and healthy Vietnam veterans: Correlations with psychophysiologic responses. *Journal of Consulting and Clinical Psychology*, *58*, 329-335.

R CORE TEAM (2019). R: *A language and environment for statistical computing*. Wien, A: R Foundation for Statistical Computing.

RADEMAKER, A.R., KLEBER, R.J., MEIJER, M.E. & VERMETTEN, E. (2009). Investigating the MMPI-2 trauma profile in treatment-seeking peacekeepers. *Journal of Personality Assessment*, *91* (6), 593-600.

RESNICK, P.J. (1997). Malingering of posttraumatic disorders. In R. Rogers (Ed.), *Clinical Assessment of Malingering and Deception (2nd ed.)*. New York: Guilford.

ROGERS, R. (2008). Detection strategies for malingering and defensiveness. In R. Rogers (Ed.), *Clinical assessment of malingering and deception*. New York: Guilford.

ROGERS, R. & BENDER, D. (2018). Clinical assessment of malingering and deception. New York: Guilford.

ROGERS, R., SALEKIN, R.T., SEWELL, K.W., GOLDSTEIN, A.M. & LEONARD, K. (1998). A comparison of forensic and nonforensic malingerers: A prototypical analysis of explanatory models. *Law and Human Behavior*, *22*, 353-367.

ROGERS, R., SEWELL, K.W., MARTIN, M.A. & VITACCO, M. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2 and malingering. *Assessment*, *10*, 160-177.

ROGERS, R.R., SEWELL, K.W. & SALEKIN, R.T. (1994). A meta-analysis of malingering on the MMPI-2. Assessment, 1, 227-237.

ROMA, P., GIROMINI, L., BURLA, F., FERRACUTI, S., VIGLIONE, D.J. & MAZZA, C. (2020). Ecological validity of the Inventory of Problems-29 (IOP-29): An Italian study of court-ordered, psychological injury evaluations using the Structured Inventory of Malingered Symptomatology (SIMS) as criterion variable. *Psychological Injury and Law*, *13*, 57-65.

ROYSTON, P. (1995). Remark AS R94: A remark on algorithm AS 181: The W-test for normality. *Appl Stat*, *44* (4), 547-551.

RÜCKER, G., KRAHN, U., KÖNIG, J., EFTHIMIOU, O. & SCHWARZER, G. (2019). *Netmeta: Network meta-analysis using frequentist methods, R package version 1.1-0*.

SCHEIBE, S., BAGBY, R., MILLER, L. & DORIAN, B. (2001).

Assessing posttraumatic stress disorder with the MMPI-2 in a sample of workplace accident victims. *Psychological Assessment*, *13*, 369-374.

SCHLENGER, W.E. & KULKA, R.A. (1987). *Performance of the Keane-Fairbank MMPI scale and other self-report measures in identifying posttraumatic stress disorder*. Paper presented at the American Psychological Association meetings, New York.

SCHWARZER, G. (2007). Meta. An R package for meta-analysis. *R News*, *7* (3), 40-45.

SMITH, G.P. & BURGER, G.K. (1997). Detection of malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of American Academic Psychiatry and the Law*, 25, 183-189.

SPIELBERGER, C.D. (1983). *Manual of the State-Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

SPITZER, R.L., WILLIAMS, J.B.W. & GIBBON, M. (1987). *Structured clinical interview for DSM-III-R (SCID)*. New York: New York State Psychiatric Institute.

STERNE, J.A.C., SUTTON, A.J., IOANNIDIS, J.P.A, TERRIN, N., JONES, D.R., LAU, J. et al. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, *343*, d4002.

TALBERT, F.S., ALBRECHT, N.N., ALBRECHT, J.W., BOUDEWYNS, P.A., HYER, L.A., TOUZE, J.H. et al. (1994). MMPI profiles in PTSD as a function of comorbidity. *Journal of Clinical Psychology*, 50, 529-537.

TOLIN, D.F., MALTBY, N., WEATHERS, F.W., LITZ, B.T., KNIGHT, J. & KEANE, T.M. (2004). The use of the MMPI-2 Infrequency-Psychopathology Scale in the assessment of posttraumatic stress disorder in male veterans. *Journal of Psychopathology and Behavioral Assessment*, *26* (1), 23-29.

van IMPELEN, A., MERCKELBACH, H., JELICIC, M. & MERTEN, T. (2014). The Structured Inventory of Malingered Symptomatology (SIMS): A systematic review and meta-analysis. *The Clinical Neuropsychologist*, *28*, 1336-1365.

VIGLIONE, D.J., GIROMINI, L. & LANDIS, P. (2017). The development of the Inventory of Problems-29: A brief self-administered measure for discriminating bona fide from feigned psychiatric and cognitive complaints. *Journal of Personality Assessment*, *99* (5), 534-544.

WEATHERS, F.W., LITZ, B.T., HERMAN, D.S., HUSKA, J.A. & KEANE, T.M. (1993). *The PTSD checklist: Reliability, validity, & diagnostic utility*. Paper presented at the Annual Meeting of the International Society for Traumatic Stress Studies, San Antonio, TX.

WELLS, G.A., SHEA, B., O'CONNELL, D., PETERSON, J., WELCH, V., LOSOS, M. & TUGWELL, P. (2012). *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses*. Retrieved from: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp.

WETTER, W.M., BAER, R.A., BERRY, D.T.R., ROBINSON, I.H. & SUMPTER, J. (1993). MMPI-2 profiles of motivated fakers given specific symptom information: A comparison to matched patients. *Psychological Assessment*, *5*, 317-323.

WEYERMANN, A.G., NORRIS, F.H. & HYER, L.A. (1996). Examining comorbidity and posttraumatic stress disorder in a Vietnam veteran population using the MMPI-2. *Journal of Traumatic Stress*, *9*, 353-360.

WIDOWS, M.R. & SMITH, G.P. (2004). *SIMS: Structured Inventory of Malingered Symptomatology professional manual*. Odessa, FL: Psychological Assessment Resources.

WILSON, J. & WALKER, A. (1990). Toward an MMPI trauma profile. *Journal of Traumatic Stress*, 3, 151-168.

WISE, E.A. (1996). Diagnosing posttraumatic stress disorder with the MMPI clinical scales: A review of the literature. *Journal of Psychopathology and Behavioral Assessment*, *18*, 71-82.

APPENDIX 1

Study quality analysis

Table A1-1 – Newcastle-Ottawa Quality Assessment Scale (adapted for cross-sectional studies)

Studies	Selection				Comparability		Outcome		Total Quality Score
Autor, year	Representativeness of the sample	Sample size	Non-respondents	Ascertainment of exposure	Study controls for the most important factor	Study controls for any addictional factor	Assessment of the outcome	Statistical test	
Albrecht et al. (1994)	*			**			*		4
Baldrachi et al. (1999)	*			**			*	*	5
Forbes et al. (1999)	*			**			*		4
Glenn et al. (2002)	*		*	**			*	*	6
Greenblatt & Davis (1999)	*			**			*	*	5
Munley et al. (1995)	*			**			*		4
Weyermann et al. (1996)	*			**			*		4

Table A1-2 – Newcastle-Ottawa Quality Assessment Scale

Studies	Selection				Comparability		Exposure			Total Quality Score
Autor, year	Case definition adequate	Representativeness of the cases	Selection of Controls	Definition of Controls	Study controls for the most important factor	Study controls for a second important factor	Ascertain-ment of exposure	Same method of ascertainment for cases and controls	Non-Response rate	
Arbisi & McNulty (2006)	*				*			*		3
Eakin et al. (2006)	*	*	*	*	*	*		*	*	8
Elhai et al. (2000)	*	*	*	*	*			*	*	7
Elhai et al. (2001)	*	*	*		*			*	*	6
Elhai et al. (2004)	*	*	*		*			*	*	6
Lange et al. (2010)			*		*			*		3
Lees-Haley (1992)			*		*			*		3
Litz et al. (2010)	*	*	*	*	*			*		6
Marshall et al. (2006)	*	*	*	*	*	*		*		7
Rademaker et al. (2009)	*			*	*			*		4
Scheibe et al. (2001)	*		*		*			*		4
Tolin et al. (2004)	*		*		*			*		4
Wetter et al. (1993)		*	*	*	*			*		5

APPENDIX 2

Participant characteristics

Table A2-1

Studies	Country	Experimental condition	Population	Gender	Age	Education (in years)	Marital status	War
Albrect et al. (1994)	USA	PTSD	47	M	42.6 (2.2)	13.2 (1.8)	–	Vietnam
Arbisi & McNulty (2006)	USA	PTSD	55	M	61.3 (14.8)	13.3 (3.3)	65.5% married 18.2% divorced 5.5% widowed 9.1% single	38.2% WWII 36.4% Vietnam 7.3% Korea 3.6% Gulf
	USA	Fakers	35	M	57.7 (15.2)	13.4 (2.9)	64.7% married 20.6% divorced 11.8% single 2.9% widowed	26.5% WWII 32.4% Vietnam 8.8% Korea 17.6% Gulf
Baldrachi et al. (1999)	USA	PTSD	36	M	45 (3.7)	12.1 (2.2)	18 married 15 divorced 1 single 2 not classified	Vietnam
Eakin et al. (2006)	USA	PTSD, Control, Fakers	85	M = 26 F = 59	19 (1.2)	80% first or second year of university	96 unmarried	–
Elhai et al. (2000)	USA	PTSD	124	M	45.71 (6.54)	12.28 (2.71)	8 single 62 married 30 previously married	2 WWI 4 Korea 92 Vietnam 2 Gulf
	USA	Fakers	84	M = 27 F = 57	29.77 (9.81)	13.57 (1.81)	55 single 39 married 6 previously married	–

continued on next page

continued

Studies	Country	Experimental condition	Population	Gender	Age	Education (in years)	Marital status	War
Elhai et al. (2001)	USA	CSA PTSD	64	M = 9 F = 55	31.21 (8.73)	12.21 (2.03)	37 single 35 married, in a relationship, or cohabitating 27 divorced or widowed	–
	USA	Fakers	85	M = 25 F = 58	29.67 (9.80)	13.56 (1.83)	38 single 56 married, in a relationship, or cohabitating 6 divorced or widowed	–
Elhai et al. (2004)	USA	CSA PTSD	41	M = 3 F = 38	33.7 (10.3)	12.5 (2.4)	60% single 20% married 20% divorced	–
	USA	Fakers	39	M = 19 F = 20	19.5 (3.1)	95% first, second, or third year of university	–	–
Forbes et al. (1999)	Australia	PTSD	100	M	49 (2.6)	–	59% married	Vietnam
Gleen et al. (2002)	USA	PTSD	134	M	47.7 (3.8)	12.7 (2.0)	–	Vietnam
	USA	PTSD	38	M	32.1 (7.6)	13.3 (1.6)	–	Gulf
Greenblatt & Davis (1999)	USA	PTSD	27	M	44.4 (12.0)[a]	12.8 (2.1)[a]	–	Not specified
Lange et al. (2010)	Australia	Fakers Control	15 20	95% F	22.7 (8.2)	–	–	–
Lees-Haley (1992)	USA	Fakers	55	M = 32 F = 23	38.9 (10.0)	–	–	–
	USA	Control	64	M = 27 F = 37	39.1 (11.3)	–	–	–

continued on next page

continued

Studies	Country	Experimental condition	Population	Gender	Age	Education (in years)	Marital status	War
Litz et al. (1991)	USA	PTSD	29	M	43.4 (2.6)	12.8 (2.1)	–	Vietnam
	USA	Control	32	M = 21 F = 11	34.4 (9.9)	–	–	–
Marshall et al. (2006)	Canada	PTSD	186	M = 137 F = 49	40.54 (9.15)[b]	–	–	–
	Canada	Control	73	–	–	–	–	–
	Canada	Fakers	67	M = 43 F = 24	22.53 (5.8)[b]	–	–	–
Munley et al. (1995)	USA	PTSD	27	M	44	11.65	–	Not specified
Rademaker et al. (2009)	Germany	PTSD	90	M	35 (7.0)	–	–	34 Lebanon 40 Yugoslavia 16 Angola, Iraq, and Cambodia
	Germany	Control	30	M	35 (8.7)	–	–	–
Scheibe et al. (2001)	Canada	PTSD	25	M	39.08 (10.62)[c]	–	–	–
		Control	21	M				
Tolin et al. (2004)	USA	PTSD	295	M	47.3 (9.0)	–	–	341 during Vietnam 54 before Vietnam 28 after Vietnam
	USA	Control	128	M		–	–	
Weyermann et al. (1996)	USA	PTSD	31	M	45 (2.76) d	–	–	Vietnam
Wetter et al. (1993)	USA	PTSD	20	M = 14 F = 6	39.4 (7.1)	12.9 (2.0)	–	Not specified
	USA	Fakers	20	M = 8 F = 12	34.8 (7.5)	11.8 (0.6)		–

[a] Mean and Standard Deviation referring to a total sample of 102 subjects, of whom 18 were diagnosed with schizophrenia, 57 with depression, and 27 with PTSD.

[b] Mean and Standard Deviation referring to the general patient sample of 199 subjects and university students of 77, before exclusion.

[c] Mean and Standard Deviation referring to the general patient sample of 54 subjects, before exclusion.

[d] Mean and Standard Deviation referring to a total sample of 88 subjects.

Table A2-2

Studies	Experimental condition	Sample specifications
Albrect et al. (1994)	P3TSD	Sample with only PTSD diagnoses
Arbisi & McNulty (2006)	PTSD	Sample with only PTSD diagnoses
	Fakers	Veterans coached to exaggerate PTSD symptoms. Of these, 67% made use of PTSD-related treatment services and only 6.7% for other types of medical or psychological conditions
Baldrachi et al. (1999)	PTSD	Sample with only PTSD diagnoses
Eakin et al. (2006)	PTSD	Sample with only PTSD diagnoses
	Control	University students. In order to be included in the control group or in the "experimental fakers", they were required to have experienced symptoms typical of PTSD diagnosis Criteria A and have low PCL scores (no possible PTSD)
	Fakers	
Elhai et al. (2000)	PTSD	Sample with only PTSD diagnoses
	Fakers	Non-clinical university students coached to exaggerate PTSD symptoms
Elhai et al. (2001)	CSA PTSD	Sample with only PTSD diagnoses
	Fakers	Non-clinical university students coached to exaggerate PTSD symptoms
Elhai et al. (2004)	CSA PTSD	Sample with only PTSD diagnoses
	Fakers	Non-clinical university students coached to exaggerate PTSD symptoms
Forbes et al. (1999)	PTSD	Sample with only PTSD diagnoses
Gleen et al. (2002)	PTSD	Sample with only PTSD diagnoses
	PTSD	Sample with only PTSD diagnoses
Greenblatt & Davis (1999)	PTSD	Sample with only PTSD diagnoses
Lange et al. (2010)	Fakers	Non-clinical university students coached to exaggerate PTSD symptoms
	Control	Non-clinical university students
Lees-Haley (1992)	Fakers	Patients who reported psychological suffering but that such suffering could not be equated to PTSD Criteria A but who scored T=65 in the PK and PS scales
	Control	Patients seeking compensation for psychological injuries sustained with low scores in the PK and PS scales

continued on next page

continued

Studies	Experimental condition	Sample specifications
Litz et al. (1991)	PTSD	Sample with only PTSD diagnoses
	Control	Non-clinical sample
Marshall et al. (2006)	PTSD	Sample with only PTSD diagnoses
	Control	Non-clinical sample
	Fakers	Non-clinical university students coached to exaggerate PTSD symptoms
Munley et al. (1995)	PTSD	Sample with only PTSD diagnoses
Rademaker et al. (2009)	PTSD	Sample with only PTSD diagnoses
	Control	Non-clinical sample
Scheibe et al. (2001)	PTSD	Patients with PTSD, of whom 19 (68%) with various comorbidities: 12 (64%) diagnosed with major depression, 4 (21%) with an anxiety disorder other than PTSD, and 3 (18%) with panic attacks
	Control	Patients without PTSD of whom 8 (36%) have not been diagnosed with any disorders, 7 (32%) diagnosed with major depression, 3 (14%) with an anxiety disorder other than PTSD, 4 (18%) with panic attacks, and 3 (14%) with adjustment disorders
Tolin et al. (2004)	PTSD	Veterans with PTSD with various comorbidities: 48% with a depressive disorder, 23% with one or more anxiety disorders other than PTSD
	Control	Veterans without PTSD, of whom 28% with a depressive disorder, 16% with one or more anxiety disorders other than PTSD
Weyermann et al. (1996)	PTSD	Veterans, of whom 31 diagnosed solely with PTSD, 33 diagnosed with PTSD and a mood disorder, 8 diagnosed with PTSD and other anxiety disorders, and 16 diagnosed with PTSD, anxiety and mood disorders
Wetter et al. (1993)	PTSD	Sample with only PTSD diagnoses
	Fakers	Non-clinical sample coached to exaggerate PTSD symptoms

DOI: 10.26387/bpa.286.4

Heavy study investment in college students: Studyholism and Study Engagement prevalence

Yura Loscalzo

School of Psychology, Department of Health Sciences, University of Florence

yura.loscalzo@gmail.com

ABSTRACT. Lo *Studyholism* è una nuova potenziale condizione clinica definita come un disturbo correlato al disturbo ossessivo-compulsivo che può essere associato sia ad alti che a bassi livelli di *Study Engagement*. Questo studio ha l'obiettivo di valutare la prevalenza di Studyholism, Study Engagement e quattro tipi di studente (*Disengaged Studyholic*, *Engaged Studyholic*, *Engaged student*, *Detached student*) in un campione ampio ed eterogeneo di studenti universitari italiani. Inoltre, ha l'obiettivo di determinare se ci sono differenze tra i tipi di studente per quanto riguarda il genere, l'area di studio e l'età. I partecipanti (n = 5159) frequentavano diversi anni di corso e diversi corsi di studio in varie città italiane. Sono stati utilizzati test di Pearson del chi-quadro e ANOVA. Lo *Studyholism* è diffuso (15.4%) e gli *Engaged Studyholic* (3.2%) sono il tipo di studente con maggior prevalenza. Il tipo meno diffuso è l'*Engaged student* (1.6%). Inoltre, ci sono più femmine tra i *Detached student* e gli *Engaged Studyholic*. Data la sua prevalenza, studi futuri dovrebbero analizzare ulteriormente lo *Studyholism* in bambini, pre-adolescenti, e adolescenti, così come in altri paesi.

SUMMARY. Studyholism is a new potential clinical condition defined as an obsessive-compulsive related disorder, which might be associated with either high or low levels of Study Engagement. This study aimed to evaluate the prevalence of Studyholism, Study Engagement, and four types of student (i.e., Disengaged Studyholics, Engaged Studyholics, Engaged students, Detached students) on a wide and heterogeneous sample of Italian college students. Moreover, it aims to discern if there are any differences between types of student concerning gender, academic major, and age. The participants (n = 5159) were in different years and studied different academic majors in various Italian cities. Pearson's chi-squared tests and an ANOVA have been performed. Studyholism is widespread (15.4%) and Engaged Studyholics (3.2%) are the most prevalent type of student. The least prevalent type is the Engaged student (1.6%). Moreover, there are more females in the Detached student and the Engaged Studyholic types. Given its prevalence, future studies should further analyze Studyholism in children, pre-adolescents, and adolescents, as well as in other countries.

Keywords: *OCD, Workaholism, Study addiction*

INTRODUCTION

Workaholism, that is, problematic overworking, has been widely analyzed and the research showed that it is associated with psychological and physical adverse outcomes, with family functioning problems and issues at the organization level (Clark, Michel, Zhdanova, Pui & Baltes, 2016; Loscalzo & Giannini, 2017a). However, only recently, it has been suggested that a similar clinical condition could be evident in school context. More specifically, Atroszko, Andreassen, Griffiths and Pallesen (2015) introduced the construct of Study Addiction, while Loscalzo and Giannini (2017b) coined the term Studyholism.

Atroszko et al. (2015) defined problematic overstudying as a pure addiction characterized by the seven core components of substance addictions (i.e., salience, tolerance, mood modification, relapse, withdrawal, conflict, and problems). Moreover, they developed the *Bergen Study Addiction Scale* (*BStAS*), which is an adaptation of the *Bergen Work Addiction Scale* (*BWAS*; Andreassen, Griffiths, Hetland & Pallesen, 2012) and that comprises seven items, one for each of the core components of addictions. However, the BStAS Italian version showed only acceptable psychometric properties and potential issues in distinguishing between Study Addiction and Study Engagement (Loscalzo & Giannini, 2018a).

Loscalzo and Giannini (2017b), instead, proposed a conceptualization of problematic overstudying that goes beyond the addiction model and that highlights several critical theoretical differences as compared to Atroszko et al. (2015) (Loscalzo & Giannini, 2017b, 2018b, 2018c). First, Loscalzo and Giannini (2017b, 2018b, 2018c, 2019) defined Studyholism as an obsessive-compulsive related disorder (OCD-related disorder) made up by two components: i) obsessive-compulsive symptoms related to study; ii) high or low study engagement, which also includes inner motivation toward study. Hence, they have also taken into account the possibility that Studyholism might be associated with a positive attitude toward study, namely study engagement. Therefore, they suggested two types of Studyholics: i) Engaged Studyholics, that is students with high levels of both Studyholism and Study Engagement; ii) Disengaged Studyholics, namely students with high levels of Studyholism and low levels of Study Engagement. In addition, using the Heavy Study Investment framework (see Snir & Harpaz, 2011), Loscalzo and Giannini (2017b) pointed out that Engaged and Disengaged Studyholics are two types of Heavy Study Investor (HSI). However, they underlined that not all the students who spent lot of time and energy in study are Studyholics, as there is also a positive type of HSI, that is a student who is characterized by high levels of study engagement and low levels of Studyholism (i.e., Engaged Student). In sum, Loscalzo and Giannini (2017b) suggested crossing the high/low levels of Studyholism and Study Engagement to define four kinds of student: Disengaged Studyholics, Engaged Studyholics, Engaged students, and Detached students. This last type of student is not an HSI, since they have low levels of both Studyholism and Study Engagement. However, Detached Students are a negative type as they are detached from one of their most important daily activities, which is studying (Loscalzo & Giannini, 2017b).

Loscalzo and Giannini (2017b) also developed a comprehensive model including potential antecedents and outcomes of Studyholism, and they distinguished between individual and situational ones. More specifically, in order to suggest these antecedents and outcomes, they referred to the wide workaholism literature - taking into account that some differences might be present between workaholism and Studyholism. They listed, among individual antecedents, personality traits, perfectionism, motivation, cognitive factors, and inability to down-regulate negative emotion. Concerning situational antecedents, they referred primarily to the overstudy climate, which might be spread in the family and at school (including the area of study). About the outcomes, for the individual ones, they suggested low well-being at school, poor academic performance, physical and health impairment (including psychological impairment), and family functioning problems; for the situational ones, they listed aggressive behaviors and few positive relationships in class. Then, Loscalzo and Giannini (2017b) stressed the importance of distinguishing between Engaged and Disengaged Studyholics when studying the relationships between Studyholism and its antecedents and outcomes, suggesting that the first type could be less impaired and that the two types could have different relationships with the same variables.

Hence, Loscalzo and Giannini (2018c), based on a critical comparison between the DSM-5 diagnostic criteria for OCD and substance use disorder affirmed that problematic overstudying could better be conceptualized as an OCD-related disorder, even if they pointed out that the literature about the specific features of this construct is too scant now to arrive to any conclusion. Therefore, they suggested that future

studies about cerebral correlates and psychological aspects specifically linked to OCD and addictions will be critical to shed light on the real nature of problematic overstudying.

In line with this, Loscalzo and Giannini (2019) conducted a study to test many of the suggested antecedents and outcomes of Studyholism, with a specific focus on Worry as an antecedent that could provide support to their definition of problematic overstudying as an OCD-related disorder. Worry is indeed an internalizing feature that contributes to OCD (Comer, Kendall, Franklin, Hudson & Pimentel, 2004). Moreover, they aimed to analyze if there were differences between the two types of Studyholics, as they speculated in their model. The results of their path analysis provided support for the conceptualization of Studyholism as an OCD-related disorder, since worry is a strong predictor ($\beta = .67$, $p<.001$). Moreover, they found that Engaged and Disengaged Studyholics should be distinguished between as they have different relationships with some variables. As compared to Engaged Studyholics, Disengaged Studyholics have more impairment in the academic and affective areas, while they have less impairment in the social area despite the two types of Studyholics not differing in physical well-being and aggressive behaviors at school. However, even if Engaged and Disengaged Studyholics showed some differences, Disengaged Studyholics are not the most impaired type of student. Therefore, Loscalzo and Giannini (2019) suggested conceptualizing both Disengaged and Engaged Studyholics as clinical types of Studyholism that differ for their level of Study Engagement and for the area in which they are most impaired. From this evidence, Loscalzo and Giannini (2019) suggested a tentative proposal for Studyholism DSM-like criteria.

On Studyholism prevalence, Loscalzo and Giannini (in press) found a high frequency in Italian college students; a higher number of Engaged Studyholics as compared to Disengaged Studyholics; higher Studyholism and Study Engagement in females; no relationship between age and Studyholism/Study Engagement; and just a few differences concerning the area of study (i.e., Humanities and Educational students have higher Studyholism than Psychology and Health Professional students, as well as higher Study Engagement than Social Science students). Hence, this brief report aims to analyze further the prevalence of Studyholism and Study Engagement, as well as of the four types of student, in a wide and heterogeneous sample of Italian college students. Moreover, it aims to analyze if there are differences related to the gender, the area of study, and the age among the four types of student. Therefore, this study will help to determine if Studyholism deserves to be studied further in other countries and populations (i.e., children, pre-adolescents, and adolescents).

METHOD

Participants

The total sample is composed of 5159 Italian college students (73.9% females) aged between 18 and 60 years (*M* age = 23.20±4.26). They attended their courses in many different Italian cities, although Florence is the most represented (32.4%). Regarding their major of study, which have been coded in macro-groups, the following are the percentages (2.5% are missing): Technology (Engineering, Architecture, Informatics), 11.2%; Social Sciences (Psychology, Sociology, Economy, Law, Educational studies, …), 36.2%; Humanities (Literature, Language, Art, Philosophy, History, …), 15.5%; Medical studies, 13.8%; Sciences (Maths, Physics, Biology, Statistics, Chemistry), 13.1%; Helping professions (Nursing, Obstetrics, …), 4.5%; Para-Medical studies (Biotechnology, Veterinary medicine, Pharmacy, …), 3.2%. The proportions of students in years 1 to 5 were 20.8%, 17.6%, 24.5%, 15.1%, and 19.3% respectively. 2.7% of the students were said to be in their sixth year, however, it is not possible to know if the sixth year students are Medical students who are actually in their sixth year, or if they are Medical or non-Medical students who are off course and hence indicated being in the sixth year, since this information was not gathered at the time of the studies. Concerning their self-reported Grade Point Average (GPA), it ranges between 18 and 31 (where 31 stands for 30 with praise), and the *Mean* value is 26.50±2.24.

Materials

The participants filled many different self-report questionnaires, based on the specific study in which they took part. However, for this study, only the data gathered with the *Studyholism Inventory* (*SI-10*; Loscalzo & Giannini, in press; Loscalzo, Giannini & Golonka, 2018) is used. The SI-10 is a 10-item self-report instrument that has been created from a pool of 68 items and that, in its final version, is made up by

two scales, that is Studyholism and Study Engagement (with two filler items, one for each scale). The participants answer by indicating how much they agree with each item by means of a 5-point Likert scale ranging between 1 (*Strongly Disagree*) to 5 (*Strongly Agree*). In addition, the first sheet of the instrument includes some optional questions about study habits (e.g., GPA). This instrument has good psychometric properties in its Italian version (Loscalzo & Giannini, in press).

Loscalzo and Giannini (in press), in order to define the SI-10 cut-off scores for high and low Studyholism/ Study Engagement, calculated on a sample of 1296 Italian college students the T scores for the two SI-10 scales, and next selected the raw scores corresponding to the 40^{th} and 60^{th} T score. Hence, they suggested using the cut-off scores that arose from these calculations to distinguish between high and low Studyholism and Study Engagement in Italian college students, as well as to screen for the four types of student proposed by Loscalzo and Giannini (2017b), namely Disengaged Studyholics, Engaged Studyholics, Engaged students, and Detached students.

Procedure

The participants that took part in this research were recruited by means of studies for which the approval from the ethical committee of the University of Florence was obtained. More specifically, the participants come from: i) Confirmatory Factor Analysis of the SI-10 (n = 956; Loscalzo & Giannini, in press); ii) path analysis aimed to test Loscalzo and Giannini's (2017b) Studyholism model (n = 1958; Loscalzo & Giannini, 2019); iii) preliminary analyses to select the variables to introduce in Loscalzo and Giannini (2019)'s path analysis model (n = 300; this data have not been published in Loscalzo & Giannini, 2019); iv) other samples gathered for research about Studyholism, whose results have not been published nor submitted to other journals yet. From the merged total sample, the participants who did not answer to all the SI-10 items have been removed, in order to avoid replacing the missing values and using instead only the data from participants who filled all the items of the scale.

Most of the participants (78.5%) filled the questionnaire online, along with the other instruments used for the specific study they took part in. The participants who filled the paper-and-pencil version are the ones gathered for Loscalzo and Giannini's (2019) preliminary analysis and for two other studies that have not been published yet. These students gave their written informed consent before participating in the research. Students who filled out the questionnaire online were provided instead with a first page explaining the research purposes, the anonymity of their responses, and their right to stop filling the questionnaire (and hence not submitting their response) at any time. Moreover, they had to check a box indicating that by continuing to fill out the questionnaire on the following pages they were giving their informed consent to take part to the research.

Data analysis

The analyses have been performed through SPSS. First, the descriptive statistics of Studyholism and Study Engagement, as well as their Pearson's correlation, have been analyzed. Next, Studyholism and Study Engagement have been categorized in three levels each, high, intermediate, and low Studyholism/Study Engagement. In order to do this coding, reference has been made to Loscalzo and Giannini's (in press) cut-off scores: high Studyholism (scores between 19 and 20), low Studyholism (scores between 4 and 9), high Study Engagement (scores between 19 and 20), and low Study Engagement (scores between 4 and 10). Intermediate levels have been defined for scores ranging between 10 and 18 (Studyholism) and between 11 and 18 (Study Engagement). Next, Pearson's chi-squared test has been performed to calculate the percentages of students for each of the four student types, as well as statistically significant differences in their prevalence. Also, Pearson's chi-squared tests have been performed to analyze if there are statistically significant differences among the types of student concerning gender and area of study (for the area of study, Helping Professions and Para-Medical groups have not been included in order to respect the assumption about the minimum count of 5 for each cell of the contingency table). Finally, an ANOVA has been performed to analyze age-related differences among the types of student.

RESULTS

The descriptive statistics highlighted that both Studyholism and Study Engagement range between 4 and 20, with Studyholism having a *Mean* value of 14.04±3.98

and Study Engagement having a *Mean* value of 14.50±3.54. Moreover, the correlation between Studyholism and Study Engagement is of .12 (p<.001), in line with their low factor correlation (Loscalzo and Giannini, in press).

As concerns the prevalence of high/low Studyholism and Study Engagement, as well as of the four types of student, Table 1 shows the results of the contingency table in which Studyholism and Study Engagement have been entered as variables. Moreover, Pearson's chi-squared analyses [$\chi^2(4) = 53.44$, p<.001] showed that there is a statistically significant difference between the prevalence of Engaged Studyholics (3.2%) and Disengaged Studyholics (2.2%), as well as between Detached students (2.8%) and Engaged students (1.6%). The Disengaged Studyholic type has a statistically significant lower prevalence than the Detached student type, and the Engaged student type has a lower prevalence than the Engaged Studyholic. Finally, the results showed that the 15.4% of the sample has high Studyholism.

In order to analyze if there are differences in gender, age, and area of study, two other Pearson's Chi-squared tests and one ANOVA have been performed (see Table 2). The results showed that Detached students and Engaged Studyholics are more prevalent in females as compared to males, while there is no difference concerning gender for Disengaged Studyholics and Engaged students: $\chi^2(3) = 21.58$, p<.001 (see Table 1 for the contingency table arisen from this analysis). For the area of study, there are not statistically significant differences: $\chi^2(12) = 16.77$, $p = .16$. Finally, the ANOVA showed that there is not a difference in the age among the four types of student: $F_{(3, 497)} = 2.05$, $p = .11$, $\chi^2 = .012$.

DISCUSSION AND CONCLUSIONS

This study analyzed the prevalence of the two forms of Heavy Study Investment, that is Studyholism and Study Engagement, and of the four types of student proposed by Loscalzo and Giannini (2017b) through the crossing of high/low levels of Studyholism/Study Engagement on a wide and heterogeneous sample of Italian college students.

Table 1 – Contingency table between Studyholism and Study Engagement, and prevalence of the four types of student on Italian college students

			Study Engagement			
			Low	**Intermediate**	**High**	**Total**
Studyholism	Low	n	142	539	84	765
		%	2.8[a]	10.4	1.6[b]	14.8
	Intermediate	n	461	2649	492	3602
		%	8.9	51.3	9.5	69.8
	High	n	112	514	166	792
		%	2.2[c]	10.0	3.2[d]	15.4
	Total	n	715	3702	742	5159
		%	13.9	71.8	14.4	100

Note. [a] = Detached student; [b] = Engaged student; [c] = Disengaged Studyholic; [d] = Engaged Studyholic. High and low levels of Studyholism and Study Engagement have been defined trough Loscalzo and Giannini's (in press) cut-off scores.

Table 2 – Contingency table between types of student and gender

			Gender		
			Male	**Female**	**Total**
Type of student	Detached student*	n	57	85	142
		%	40.1	59.9	100
	Disengaged Studyholic	n	28	84	112
		%	25.0	75.0	100
	Engaged student	n	29	55	84
		%	34.5	65.5	100
	Engaged Studyholic*	n	29	137	166
		%	17.5	82.5	100
	Total	n	143	361	504
		%	28.4	71.6	100

Note. * = The difference between males and females is statistically significant.

The results showed that in this sample there is a high prevalence of students with high levels of Studyholism, which is even higher than the prevalence found by Loscalzo and Giannini (in press) in their Italian sample. This suggests that this new potential clinical condition should be further analyzed in order to develop effective preventive and clinical interventions, especially considering the negative health outcomes and the higher intention to drop out of these students (Loscalzo & Giannini, 2019). Moreover, in line with Loscalzo and Giannini (in press), there are more Engaged Studyholics than Disengaged Studyholics, providing further support to Loscalzo and Giannini's (2017b) speculation that Studyholism and Study Engagement might co-occur and highlighting the need to distinguish between these two types of Studyholism. Engaged students are the least prevalent type. This provides additional support to Loscalzo and Giannini's (in press) suggestion to develop interventions aimed at fostering Study Engagement in Italian college students, since this positive attitude toward study leads to positive health and academic outcomes, in contrast to Studyholism, as showed by Loscalzo and Giannini (2019). In this study, the most prevalent type is not the Detached student, as in Loscalzo and Giannini (in press) study, but the Engaged Studyholic. This result might be due to the fact that the current sample comprehends a higher prevalence of students with high Studyholism. Anyway, the prevalence of the Detached student is still higher as compared to Engaged students and Disengaged Studyholics.

Finally, concerning differences among the four types of student, females have a statistically significant higher probability of being Engaged Studyholics and Detached students compared to males, but there is not a gender difference for the Disengaged Studyholic and the Engaged student type. In addition, there are no statistically significant differences in the types of student as far as age and major of study is concerned. Hence, in line with Loscalzo and Giannini (2019), preventive interventions should be implemented across all the areas of study by the first year of College, as each student could potentially be a Studyholic, no matter the major of study or the age.

One of the limits of this research is related to the sample, which is large, but most of the participants are females. Also, the sample is made up by college students only, while Studyholism could have its onset at a younger age (Loscalzo & Giannini, 2017b). Besides these limits, the present report has the merit of having analyzed the prevalence of a new clinical condition associated with overstudying that, as shown by the results, is quite widespread in college students and should be prevented through interventions across all the majors and years of study. Also, it has highlighted that Study Engagement, or a positive attitude toward study, should be fostered by means of preventive interventions, since the Engaged student (or the positive type of Heavy Study Investor) is the least prevalent type.

Future studies could analyze the prevalence of Studyholism, Study Engagement, and the four types of student in Italian children, pre-adolescents, and adolescents, aiming to evaluate the developmental trend of this new potential clinical disorder, especially through longitudinal studies. Moreover, it would be interesting to analyze the prevalence of Studyholism and the four types of student in other countries to understand if there are some culture-related differences. Given that there are the Italian, Polish, Croatian, and Spanish versions of the SI-10 currently available, these potential studies could be done in the near future.

References

ANDREASSEN, C.S., GRIFFITHS, M.D., HETLAND, J. & PALLESEN, S. (2012). Development of a work addiction scale. *Scandinavian Journal of Psychology*, *53*, 265-272.

ATROSZKO, P.A., ANDREASSEN, C.S., GRIFFITHS, M.D. & PALLESEN, S. (2015). Study addiction – A new area of psychological study: Conceptualization, assessment, and preliminary empirical findings. *Journal of Behavioral Addiction*, *4* (2), 75-84.

CLARK, M.A., MICHEL, J.S., ZHDANOVA, L., PUI, S.Y. & BALTES, B.B. (2016). All work and no play? A meta-analytic examination of the correlates and outcomes of workaholism. *Journal of Management*, *42* (7), 1836-1873.

COMER, J.S., KENDALL, P.C., FRANKLIN, M.E., HUDSON, J.L. & PIMENTEL, S.S. (2004). Obsessing/worrying about the overlap between obsessive-compulsive disorder and generalized anxiety disorder in youth. *Clinical Psychology Review*, *24* (6), 663-683.

LOSCALZO, Y. & GIANNINI, M. (2017a). Clinical conceptualization of workaholism: A comprehensive model. *Organizational Psychology Review*, *7* (4), 306-329.

LOSCALZO, Y. & GIANNINI, M. (2017b). Studyholism or study addiction? A comprehensive model for a possible new clinical condition. In A.M. Columbus (Ed.), *Advances in Psychological Research*. New York: Nova Publisher.

LOSCALZO, Y. & GIANNINI, M. (2018a). The Bergen study addiction scale: Psychometric properties of the Italian version. A pilot study. *Psychiatria i Psychologia Kliniczna*, *18* (3), 271-275.

LOSCALZO, Y. & GIANNINI, M. (2018b). Response to: Theoretical and methodological issues in the research on study addiction with relevance to the debate on conceptualising behavioural addictions: Atroszko (2018). *Psychiatria i Psychologia Kliniczna*, *18* (4), 426-430.

LOSCALZO, Y. & GIANNINI, M. (2018c). Problematic overstudying: Studyholism or study addiction? Commentary on: Ten myths about work addiction. *Journal of Behavioral Addictions*, *7* (4), 867-870.

LOSCALZO, Y. & GIANNINI, M. (2019). Heavy study investment in Italian college students. An analysis of Loscalzo and Giannini's (2017) Studyholism comprehensive model. *Frontiers in Psychiatry*, *10*, 489.

LOSCALZO, Y. & GIANNINI, M. (2020). *When studying become an obsession: The Studyholism inventory – Extended version (SI-15)*. Manuscript submitted for publication.

LOSCALZO, Y. & GIANNINI, M. (in press). Studyholism Inventory (SI-10): A short instrument for evaluating study obsession within the heavy study investment framework. *European Journal of Psychology*.

LOSCALZO, Y., GIANNINI, M. & GOLONKA, K. (2018). Studyholism Inventory (SI-10): Psychometric properties of the Italian and Polish versions. In T.M. Ostrowski, B. Piasecka & K. Gerc (Eds.), *Resilience and health. Challenges for an individual, family and community*. Krakow: Jagiellonian University Press.

SNIR, R. & HARPAZ, I. (2011). Beyond workaholism: Towards a general model of heavy work investment. *Human Resource Management Review*, *22*, 232-243.