vol. LXXX May-August 2022
N° 294

2/3

Four-monthly Journal
ISSN 0006-6761

BPA

Bollettino di Psicologia Applicata
APPLIED PSYCHOLOGY BULLETIN
Indexed in PsycINFO® – Scopus Bibliographic Database

Research

Experiences & Tools

GIUNTI PSYCHOMETRICS

CONTENTS

⬢ Research

▲ Experiences & Tools

DOI: 10.26387/bpa.2022.00003

Food and alcohol disturbance in Italy: A reality? An epidemiological study on under 40

Ilaria Consolo[1], Tjuana Foffo[2]

[1] Psychologist, Psychotherapist, Expert in Addiction, Sexologist, Istituto Italiano di Sessuologia Scientifica, Rome, Italy
[2] Psychologist, Sexologist, Expert in Addiction, Istituto Italiano di Sessuologia Scientifica, Rome, Italy

ilariaconsolo@tiscali.it

ABSTRACT. Lo studio ha indagato la presenza del *food and alcohol disturbance* (*FAD*) in un campione di 716 abitanti della penisola italiana, tra i 14 e i 40 anni di età. Si tratta del primo studio epidemiologico in Italia sul FAD. I risultati indicano come i predittori dello sviluppo del fenomeno in oggetto siano sia un disturbo del comportamento alimentare che un abuso di alcol; l'attività fisica non sembra invece essere un fattore predittivo, mentre sembra esserlo la limitazione nell'assunzione di calorie. Il FAD pare essere concentrato in specifiche aree geografiche dell'Italia.

SUMMARY. The aim of the present study is to study the presence of food and alcohol disturbance (FAD) in Italian participants aged between 14 and 40 years old, and to examinate the relationship between FAD and eating disorders (EDs), alcohol abuse, and the intensity of physical exercise. A sample of 716 (74% women, mean age = 23 years old) participants rated a series of self-report measures by using an online survey. Measures included: a questionnaire to obtain information about age, gender, alcohol use, geographical provenience, and education; EAT-26 test; AUDIT test; and an ad hoc measure to obtain information on FAD. 84.1% of the sample had a middle school license or higher levels of education. 11.9% had significantly high scores at the EAT-26 test, 8.4% at the AUDIT. FAD behaviors were found in .85 % of the sample, who obtained significant scores in all three measures. Pearson's correlation analysis showed a significant relationship between EAT-26 and DQ, and between AUDIT and DQ. ANOVA and Factor Analysis, used to verify the validity of the DQ, showed positive results. Results show that both an ED and alcoholic abuse predict FAD, while physical activity does not seem to be a predictive factor. Further studies are necessary in order to diagnose, prevent and treat FAD.

Keywords: *Eating disorder, Alcohol-related disorders, Addiction, FAD, Physical exercise*

INTRODUCTION

New York Times media in 2008 have developed the term drunkoressia to indicate a phenomenon which includes the specific characteristics of an eating disorder (ED) related with alcohol abuse (CBS News, 2008; Kershaw, 2008; Smith, 2008; Stoppler, 2008). This term refers to the phenomenon of restricting food intake and starving before using alcohol, which is quite common among young adults. More recently, Choquette et al. (2018) have stated that the

term drunkorexia should not be considered a diagnostic label, but should instead indicate specific eating behaviors. The authors have suggested the use of the term food and alcohol disturbance (FAD), which refers to a clinical phenomenon related to a series of actions which include calories restriction, excessive physical exercise and various compensatory behaviors adopted before, during and after using alcohol to compensate the calories intake or to maximize the intoxication (Choquette et al., 2018; Horvath, Shorey & Racine, 2020). According to scientific literature, those who have been labeled as FAD are mostly university-aged people who drink heavily and don't eat much during the day to compensate the alcohol-related calories intake that they know they will have in the evening (Barry & Piazza-Gardner, 2012). People who limit their caloric intake and/or practice excessive physical exercise to compensate the calories related to alcohol use may be suffering from an eating disorder (ED), an alcohol-related disorder, or both. However, scientific literature has suggested how those who eat less to drink more without gaining weight tended to have higher scores in relation to disordered eating behaviors, while those who intoxicate more rapidly had higher scores in measures related to alcohol abuse (Roosen & Mills, 2015).

There is an interesting association between FAD, and binge eating with EDs: high levels of food restriction, which are typical of both binge eating and FAD, are associated with higher alcohol-related problematics (Stewart, Angelopoulous, Baker & Boland, 2000), and previous studies have enlightened an association between the use of alcohol and dysfunctional eating behaviors in university students (Krahn, Kurth, Gomberg & Drewnowski, 2004). Those who limit their caloric intake or use compensatory behaviors in order to compensate the calories related to the alcohol intake may suffer from EDs, substance related disorders or both. Disordered eating and the use and abuse of alcohol are frequent in university students, and are often in comorbidity (Horvath et al., 2020). Some authors have also enlightened an association between binge drinking and FAD, and EDs: higher levels of food intake restriction, which are typical of both FAD and binge drinking, are associated with higher alcohol intake (Barry & Piazza-Gardner, 2012). The main difference between FAD and anorexia nervosa is that in FAD body weight issues are related to the alcoholic intake: starving is the precondition for drinking. However, as time goes by, the FAD motivation may become an anorexic motivation, losing weight may become the main purpose and patients may feel powerful when they overcome hunger. Starving can become important in relation to avoid weight gain and to maximize the effects of alcohol by drinking heavily on an empty stomach, with all the related physical consequences. According to Thompson-Memmer, Glassman and Diehr (2018), FAD behavior, which the authors have labeled alcoholimia, is part of an ED with alcohol related problematics that must be addressed during therapy.

Although the term FAD has been used both in scientific and popular literature, the diagnosis and treatment of this disorder is still problematic, and there is no fixed parameter in relation to the quantity or intensity of the pathological behaviors (Thompson-Memmer et al., 2018). Moreover, research on this topic is still scarce, and there is a lack of consensus on the psychological consequences of these behaviors and on the (potential) differences related to gender, which however so far have not been found (Horvath et al., 2020).

This disorder has mostly been described in the university population of the USA, however more recent research is also investigating different populations (e.g., older people) (Moeck & Thomas, 2021). Further research has also been conducted in Europe, and studies have demonstrated that FAD is a transcultural problematic, which predictors however may change in relation to different cultures. Therefore, culture should be considered when tailoring a specific intervention (Choquette et al., 2018). Moreover, there is a need for studies which are focused on potential gender differences in FAD behaviors, since data on this topic are still missing. The present study aims at investigating the presence of FAD behaviors in Italy and the relationship between FAD, EDs, alcohol intake and the amount of physical activity.

METHOD

Participants and procedures

The present study represents an observational prospective study which investigated the characteristics of FAD in Italy by using self-report questionnaires. The participants were people aged between 14 and 40 years old, recruited on the Internet by a survey conducted online. The only exclusion criterion was related to age. All participants provided written informant consent before starting the survey. The study was conducted by ethical standards, following the 1964 Helsinki declaration (WMO, 1964) and its later amendments.

Measures

Self-report measures included a questionnaire to obtain personal information (i.e. age, sex, height, geographic origin, education) and the following measures.

- The *Eating Attitude Test – 26* (*EAT-26*; Garner, Olmsted, Bohr & Garfinkel, 1982; Marronaro, Rossi, Aquilio & Scacchioli, 2009), is one of the most widely used measures for the assessment of symptoms and concerns related to EDs, and is a reliable measure for the screening of high risk populations. It is composed by 26 items, with a cut-off of 20. Scores which are higher than 20 indicate an ED risk. In the present study alpha was .87 indicating good psychometric properties.
- The *Alcohol Use Disorders Identification Test* (*AUDIT*; Saunders, Aasland, Babor, de la Fuente & Grant, 1993; Struzzo, Faccio, Moscatelli, Scafato & PRISMA Gruppo, 2006), rates the frequency and amount of the alcohol intake and can identify an excess in the alcoholic intake or an alcohol-related disorder. AUDIT is a 10-item questionnaire which covers the domains of alcohol consumption, drinking behaviour, and alcohol-related problems. Responses to each question are scored from 0 to 4, giving a maximum possible score of 40. Scores ranging between 1 and 7 are considered related to a low risk, scores comprised between 8 and 14 indicate alcohol-related problematics and possible risky behaviors, scores comprised between 15 and 40 correlate with an alcohol-related disorder. In the present study alpha was .85 indicating good psychometric properties.
- An ad-hoc measure for the assessment of FAD (the DQ), composed by four questions on the use and abuse of alcohol in the last 30 days, in which the authors aim at having information on whether the person has reduced the calories' intake before drinking or whether after drinking there have been inappropriate compensatory behaviors or an increase in physical exercise. The afore mentioned questions were developed from the existing scientific literature on the topic, which has documented a positive relationship between the level of physical activity and alcohol intake (Pate, Heath, Dowda & Trost, 1996; Piazza-Gardner & Barry, 2012; Westerterp, Meijer, Goris & Kester, 2004), the recurring restriction of food intake before ingesting alcohol (Burke, Cremeens, Vail-Smith & Woolsey, 2010) and the association between EDs and the increase of physical exercise.

Analytic plan

Descriptive analyses were calculated for the whole sample. We used the tables of frequency to analyze scoring data. Spearman's correlations were calculated between the measures. ANOVA was used to test the possible relationship between sex and results. Cronbach's alphas were calculated for each measure. As for the DQ, explorative factor analysis was uses to test the reliability of the measure. We analyzed the demographic characteristics of the participants who scored above the cut-offs of the adopted measures (i.e., who had alcohol related or eating disorders).

RESULTS

The final sample was composed by 716 participants, with an age ranging between 14 and 40 years (M = 26.72). 74% of the sample was composed by women. The only exclusion criterion was related to the age of the participants, which had to be comprised between 14 and 40 years old.

The sample was composed by 716 valid cases, of whom 186 were men (26%), and 530 women (74%). The average age of the sample was 26.72 years old (SD = 6.77; min = 14; max = 40; median = 27.00; mode = 27.00). The average weight was 64.51 kg (SD = 14.15), ranging from 39 to 166 kg (median = 62.5; mode = 50). The average height was 1.68 mt (SD = .09), ranging from 1.45 to 1.94 mt (median = 1.67; mode = 1.60).

The average BMI was 22.84 (SD = 4.09), ranging from 14.50 to 51.23. According to Italian guidelines, the sample was composed by 478 persons with an average weight (66.8%), 133 overweight participants (18.6%), 18 obese (first degree obesity; 2.5%) and 16 severely obese participants (2.2%), 66 underweight (9.2%), and 4 severe underweight persons (.6%). One participant did not indicate his weight.

The sample was composed mostly by Italian (98.3%), and by Albanian (.8%) participants. The .1% if the sample came from Romania, Bolivia, China, Colombia and Philippines. The 38.8% of the sample lives in a village (278 cases), 49.6% in a town (355 cases) and 11.6% in a city (83 cases).

32.4% of the sample came from Lazio (n = 232), then Basilicata (17.3%, n = 124), Abruzzo (12.2%, n = 87), Apulia (6.8%, n = 49) and Campania (6.0%, n = 43). Sardinia (5.2%, n = 37), Lombardy (4.1%, n = 29), Tuscany (2.4%, n = 17), Emilia Romagna (2.2%, n = 16), Marche (2.1%, n = 15), Umbria and Sicily (both 2.0%, n = 14 each), Calabria (1.8%,

n = 13), Veneto and Molise (1.1%, *n* = 8 each), Piedmont (.4%, *n* = 3), Valle D'Aosta (.3%, *n* = 2), Liguria (.1%, *n* = 1); .3% came from abroad (*n* = 2).

7% of the sample were secondary school students (50 cases), while 8.9% of the sample had completed secondary school (64 cases), 20.9% had graduated from high school (150 cases), 21.4% had a bachelor's degree (153 cases), 29.6% a master's degree (212 cases), 12.2% had a higher education degree (e.g., PhD, master, post-graduation school; 87 cases).

EAT-26 results (716 valid cases) showed an average score of 8.96 (*SD* = 10.52; variance = 110.72; min = 1; max = 68). There were 85 positive cases which scored higher than 20 (cut-off) (11.9%) (see Figure 1).

AUDIT (716 cases) average score was 2.41 (*SD* = 3.74, variance = 13.98; min = 0; max = 23). 91.6% of the sample were in at low-risk (*n* = 656) while 6.6% had possible alcohol-related problems (*n* = 47), and 1.8% had a probable alcohol-related substance use disorder (*n* = 13) (see Figure 2).

DQ results (716 valid cases) indicated a mean score of 17.31 (*SD* = 2.37); variance = 5.64; min = 7; max = 20). The final items of the questionnaire have an opposite rating, therefore higher scores indicate a low risk, while low score indicate a high risk (see Figure 3).

Five-hundred and seventy-five cases (80.3% of the sample; 153 men [21.4%]; 422 women [58.9%]) did not reach the cut-off for any of the measures.

Figure 1 – EAT-26 results

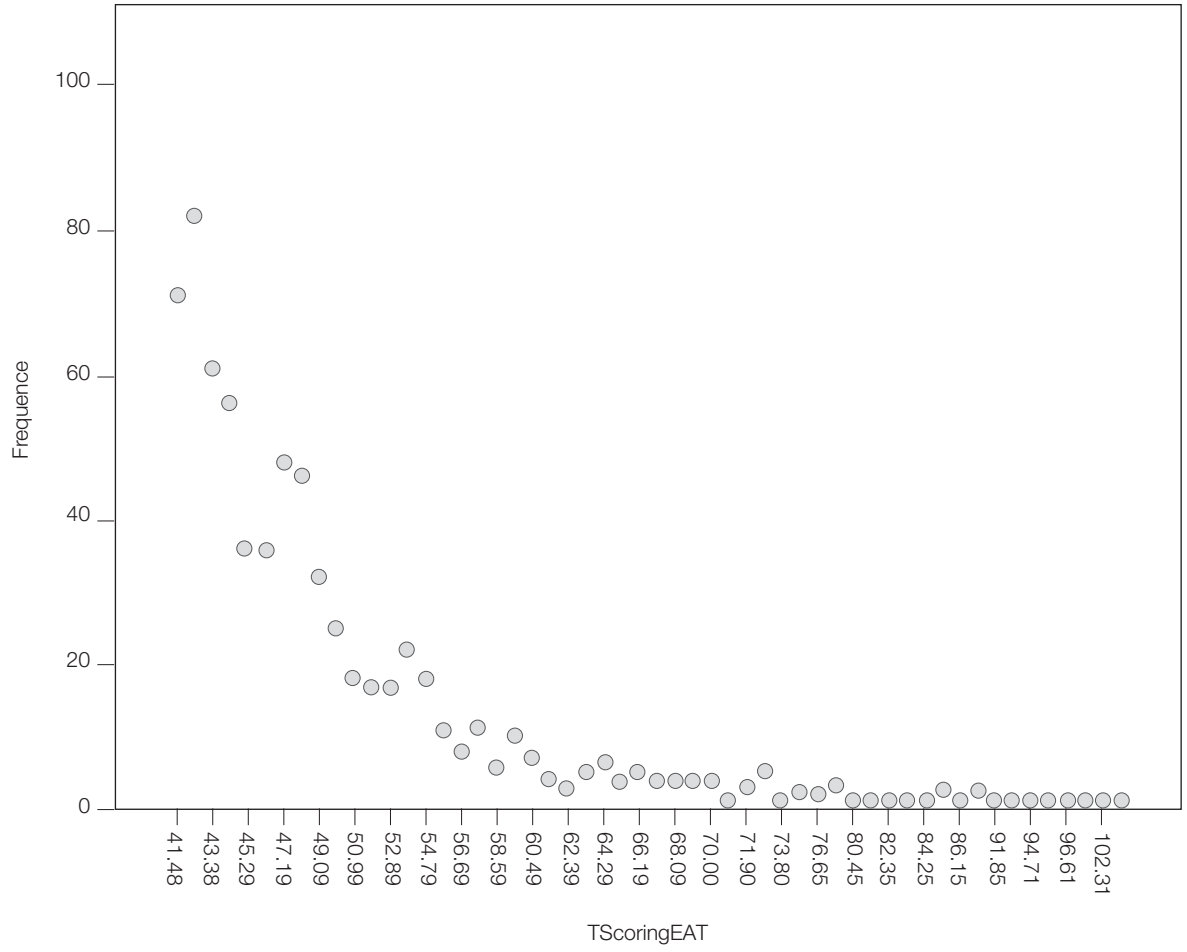
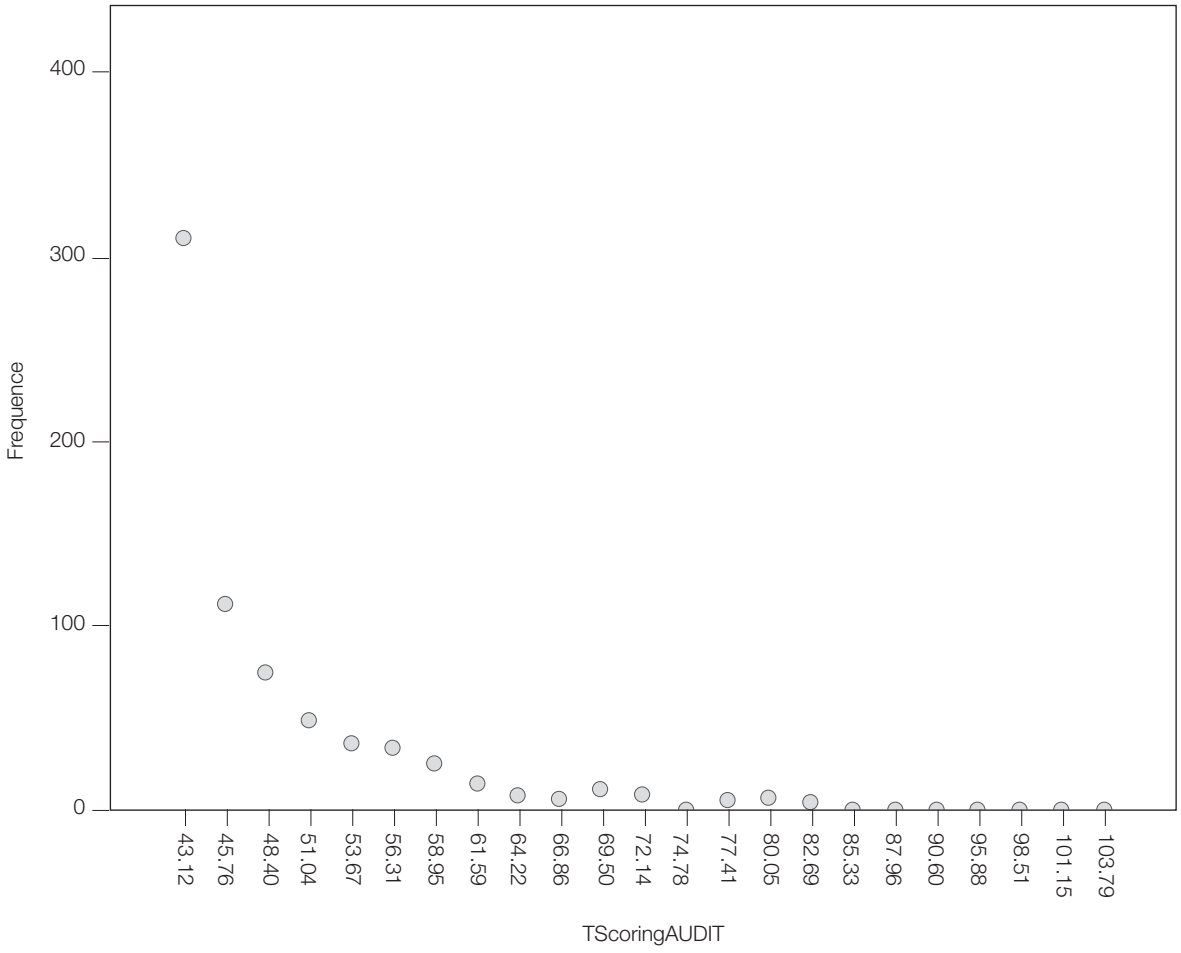


Figure 2 – AUDIT results



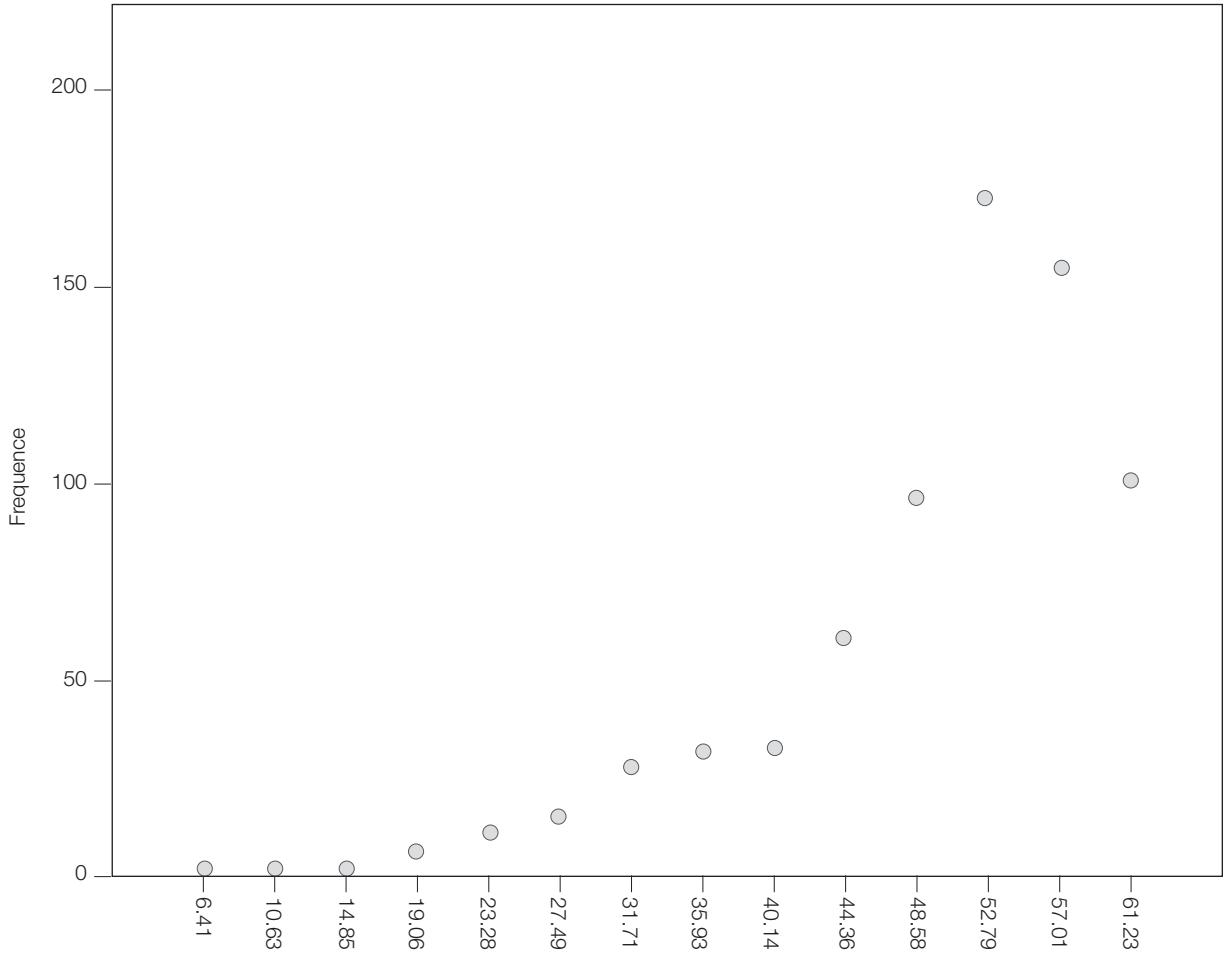
Data analysis enlightened the presence of significant answers in .98% of the population (7 cases), while 1.81% had significantly high scores at EAT-26 and DQ (13 cases) and 1.12% (8 cases) at the AUDIT and the DQ (see Figure 4).

Descriptive analysis of this specific sub-sample of 7 cases indicated a mean age of 24.6 years old (min = 14; max = 34; mode = 14). The BMI of the sample ranged from 17.58 to 24.97, with one person being underweight and the other 6 having an average weight (underweight = 16≤BMI≤18.49; average weight = 18.50≤BMI≤24.99 according to Italian Ministry of Health). In relation to the geographical provenience of the participants of this subsample descriptive analysis showed that 5 came from Basilicata, 1 from Lazio, and 1 from Abruzzo which are the second, the first and the third most represented regions of the whole sample; 6 lived in a village, 1 in a city, 3 were attending secondary school, one had graduated and one had had a master's degree.

Correlational analysis (Spearman's correlation) showed a significant negative correlation between EAT-26 and DQ and a negative correlation between the AUDIT and DQ (see Table 1 and Table 2).

Negative Spearman's correlation must be interpreted considering the difference in the scoring between the two measures, with the scores of the DQ being opposite from the two different measures. Factor analysis and Spearman's coefficient analysis were conducted to verify DQ's validity

Figure 3 – DQ results



and showed a discrete validity of the measure (see Table 3).

ANOVA showed that there were statistically significant differences related to gender at the EAT-26 and AUDIT tests, with higher means in male subjects in the AUDIT (alcohol use), and higher means in females in relation to the EAT-26 (eating attitudes).

DISCUSSION

The results of the present study show that FAD behavior was found in .85% of the sample. The FAD subsample is prevalently composed by women in relation to EDs, in line with the scientific literature and the epidemiological data, which indicate an incidence of 10:1 (women:men) in Italy (Ministero della Salute, 2020). ED patients are not exclusively women, however the prevalence in women is high. There are psychological and cultural possible explanations for this data, and biological factors could also have an impact on the tight relationship between psychopathology and its consequences for the female body.

The higher incidence of EDs in western countries could be related to the widespread presence of a stereotyped image and to the excess of consideration related to physical appearance. The predominant model, promoted by most social and cultural ways of communication, is related to physical perfection. More specifically, female body is considered desirable only in relation to impossible-to-reach standards of beauty and thinness. It is important to underline that, even if our sample was prevalently composed by women (65.5%), previous studies had enlightened how up to 70% of those who restrict food intake before drinking were women (Burke et

Figure 4 – A comparison of EAT-26, AUDIT and DQ standardized scores

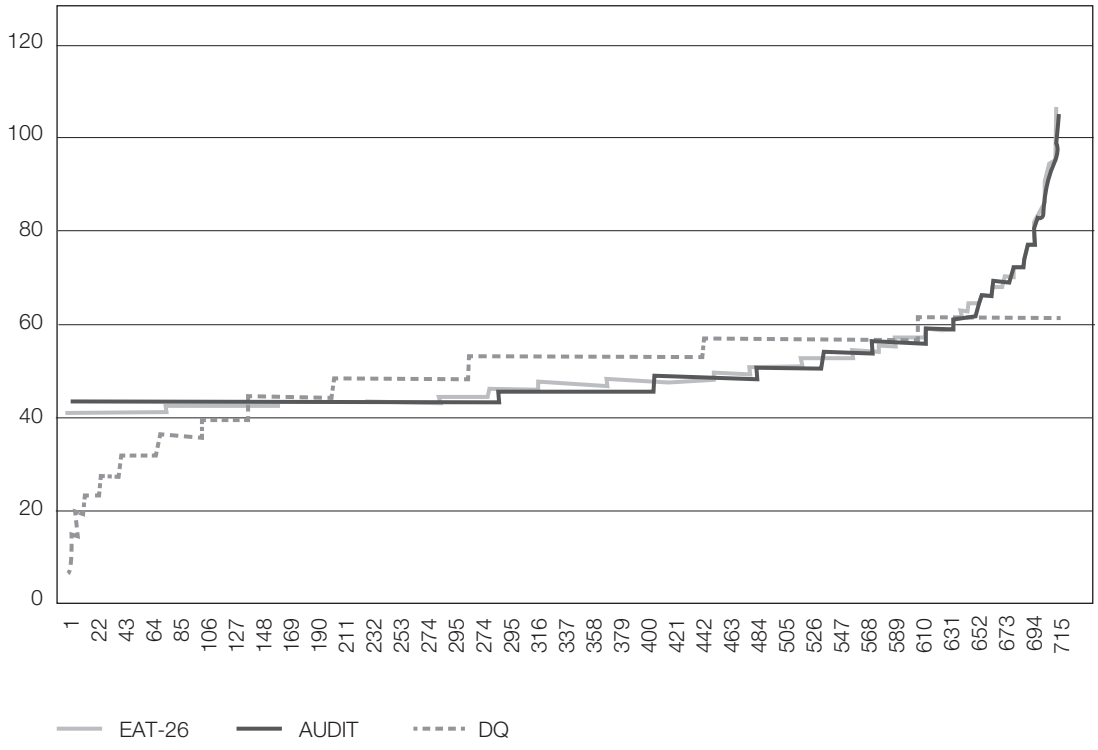


Table 1 – Correlation analysis between the EAT-26 and DQ scores

			EAT-26	DQ
Spearman's Rho	EAT-26	Correlation Coefficient	1.000	–.167**
		Sig. (2-tailed)		.000
		N	716	716
	DQ	Correlation Coefficient	–.167**	1.000
		Sig. (2-tailed)	.000	.
		N	716	716

** $p \leq .01$

Table 2 – Correlation analysis between the AUDIT and DQ scores

			AUDIT	DQ
Spearman's Rho	AUDIT	Correlation Coefficient	1.000	−.554**
		Sig.		.000
		N	716	716
	DQ	Correlation Coefficient	−.554**	1.000
		Sig. (2-code)	.000	.
		N	716	716

** $p \leq .01$

Table 3 – Factor analysis

Factor analysis – Component matrix[a]	
	Component 1
DQ 30 days	.340
DQ Food restriction	.837
DQ Compensatory behaviors	.628
DQ Physical activity	.777
Extraction method: principal component analysis.	

Legenda. [a] = 1 extracted component.

al., 2010). In 2015, Rosen and Mills (2015) found that women who restricted food intake before drinking alcohol to avoid weight gain had higher levels of EDs, while women who did not eat in order to reach faster higher levels of intoxication had more alcohol-related problematics.

Our data seem to suggest a prevalence of the phenomenon in the South of Italy and in rural areas, in which marginality and the poverty of territorial resources may contribute to the development of high-risk behaviors in young adult, as found in a study funded by the European Union on gender inequalities. The historical gap between South and North of Italy is deep and structural. There are also different geographical differences, related to rural areas and small villages and cities, which are important. In this scenario, we also must face the historical problem of the gap between the North and South of Italy. However, territorial differences can be found in the whole country. Those who are mostly affected by it are people living in the suburbs, small towns and internal rural areas, with a reciprocal influence between social and environmental problematics. Those who live in the suburbs, in rural areas or in small towns may feel like there is no future perspective for their hometowns (EEB, ENIGM, GCAP, ITALIA, 2018). There are still persistent cultural stereotypes as a consequence of a differentiated gender education, with marked inequalities and problematics for the South of Italy and the small villages.

Our study seems to suggest that an ED or an alcohol-related problematic may predict FAD behavior, as indicated also by previous studies (Pompili & Laghi, 2020). The most significant relationship was found between the FAD behavior and alcohol abuse, as found by Lupi, Martinotti and Di Giannantonio (2017) and Simons, Hansen, Simons, Hovrud and Hahn (2020). Physical activity did not seem to predict FAD behaviors, in line with previous studies on this topic (Booker, Novik, Galloway & Holmes, 2020; Palermo, Choquette, Ahlich & Rancourt, 2020) while the restriction of food intake seemed to predict FAD behaviors (Roosen & Mills, 2015).

Since 2000 alcohol abuse and EDs have been considered two of the most problematic issues in American universities, and several studies have investigated these phenomena. In 2002 O'Malley and Johnson (Kumar, O'Malley & Johnston, 2002) collected and commented data from different national surveys and studies which examined the use of alcohol in university students, such as the College Alcohol Study, the Core Institute, Monitoring the Future (an epidemiological study conducted on the basis of a research project of the institute for social research of Michigan University) and the National College Health Risk Behavior Survey. The results of the aforementioned studies showed that about 70% of university students had ingested alcohol in the previous month, and 40% reported binge drinking. In 2002 Dunn, Larimer and Neighbors found that first year university students were the most vulnerable to alcohol related problematics, particularly during the first months of university, and more recent data confirm these results (Rancourt, Ahlich, Choquette, Simon & Kelley, 2020). Predictive factors were an intensification of the academic scrupulousness in first year students, campus' social norms (which encourage alcohol use) and the lack of control from parents during college (Baer, 2002; Baer & Bray, 1999; Baer, Kivlahan & Marlatt, 1995). Anderson and colleagues (Anderson, Simmons, Martens, Ferrier & Sheehy, 2006) found that the use of alcohol and EDs in university-aged women were part of avoidant coping strategies. Other studies enlightened that a model of impulsivity/sensation seeking was highly correlated with an increased alcohol intake among students (Baer, 2002). This data was also supported by a study conducted in 2013 in the United Kingdom by da Jones, Chryssanthakis and Groom (2013). The authors used a four-factor model to investigate the relationships between alcohol intake, impulsivity, alcoholic motivations, and tendency to engage in alcohol-related problematic behaviors. Moreover, the afore mentioned study on a sample of students ranging between 18 and 25 years old showed that the feelings, the urgency and the lack of premeditations when drinking were related to different motivations, moreover there were specific associations in relation to different substances (beer, wine, liquors) and the tendency to engage in risky alcoholic related behaviors. A different study investigated the presence of EDs in American universities, by studying whether these disorders (or dysfunctional modalities adopted to control weight) were increasing during time in male and female university students. Data from three random sample survey were collected over a period of 13 years in order to study disordered eating and dysfunctional eating behaviors. Data were collected in 2008 on a sample of 641 male and female university students casually selected and were compared with 274 randomly picked university students interviewed in 2002, and with 493 students interviewed in 1995. EDs behaviors significantly increased over the years both in men and women, as did dysfunctional eating behaviors (White, Reynolds-Malear & Cordero, 2011). A 2019 study conducted

in 28 American universities showed a high risk for EDs among students. By using the Healthy Body Image (HBI) internet program, which includes an online screening to identify people who are at low or high risk or have an ED, and then suggests some possible interventions to face this risk and the related clinical problematics, the authors found out that 60% of the participants had a high ED risk or had a diagnosis of ED (Fitzsimmons-Craft, Balantekin & Eichen, 2019).

The present prospective observational study investigated a population which is larger than the ones studied in previous studies, which included only adolescents. We decided to include participants up to 40 years of age in relation to cultural differences (related mostly to the South of Italy) and to Italian socio-economic peculiarities: in Italy, for example, economic independency from the family is often reached later than in other countries because of the high unemployment rates and of the high costs of houses in most cities. The most recent Eurostat data indicate that the mean age in the European Union in which young adults leave their first home ranges between 15 and 34 years old with a mean of 26.6 years, while in Italy the mean age is 30.2. Losing their childhood certainties, the emergency of an unknown and uncontrolled situation, conforming to the group of peers, rule-breaking, rebellion, conflicts with peers, can all lead to distress and pain, especially in a critical phase of the individual development such as is adolescence, and alcohol or substance intake may be modalities through which adolescents try to cope with anxiety and painful emotions. Higher levels of socialization, transgression, being away from the family, and the "risk culture" (which explains the admiration that young people have for risky behaviors, such as drugs, substances and alcohol assumption, extreme sexual behaviors, and sensation seeking behaviors; Bastiani Pergamo & Drogo, 2012) can also affect young adults (18-20 years old) who study in universities which are distant from their hometowns. These behaviors may begin during adolescence, a period in which high risk behaviors, such as the use/abuse of alcohol (Brown et al., 2008) may be common and have a negative impact on health and, considering the socio-cultural issues enlightened above, may persist even after university.

In Italy, the first survey by Italian Ministry of Health was conducted in 2013 and found more than 300000 cases among people between 14 and 17 years old, 80% of whom were women. An Italian study carried out in 2014 on a sample of about 3000 subjects showed that this phenomenon is common among Italian young adults, with a 32.2% prevalence. Moreover, the study showed that FAD behaviors were carried out also by men and not only by women, as the media seem to suggest (Lupi et al., 2017).

The main strengths of this study are the sample size and the use of validated measures. Limitations may include the fact that part of the sample answered to the questionnaire online and not face to face, and part answered to the questionnaire in class and not alone, and this may have influenced the answers provided. Moreover, this is an unpowered study, and no power analysis was conducted, since it is a pilot study.

CONCLUSIONS

This study enlightened the need for further studies on FAD, which may be considered a construct in itself. These data call for further studies on FAD; however, at the same time, recognizing FAD as a possible nosographic construct helps at tailoring a more adapt diagnosis and treatment plan for patients affected by this condition. Moreover, further studies should try to determine the predictive factors which may play a crucial role in FAD behaviors, such as for example socio-cultural stances. The two predictive factors of FAD behaviors that emerged in this study are the presence of an ED in relation to the restriction of food intake before drinking, and alcohol abuse, with alcohol abuse having a stronger relationship (than EDs) with FAD. Recognizing these predictive factors, especially in clinical contexts by healthcare professionals, could help developing early interventions that may prevent the onset of FAD. Surveys should focus on the local differences related to these phenomena. A more detailed comprehension of FAD is advisable, and a strict definition and standardized investigation methods should be developed. Further studies are also necessary in order to be more prepared to face the onset and the short and long-term consequences of these behaviors.

The authors declare no conflict of interest.

References

ANDERSON, D.A., SIMMONS, A.M., MARTENS, M.P., FERRIER, A.G. & SHEEHY, M.J. (2006). The relationship between disordered eating behavior and drinking motives in college-age women. *Eating Behaviors*, *7* (4), 419-22. doi: 10.1016/j.eatbeh.2005.12.001. Epub 2006 Jan 20. PMID: 17056420

BAER, J.S. (2002). Student factors: Understanding individual variation in college drinking. *Journal of Studies on Alcohol Suppl*, *14*, 40-53. doi: 10.15288/jsas.2002.s14.40. PMID: 12022729

BAER, P.E. & BRAY, J.H. (1999). Adolescent individuation and alcohol us. *Journal of Studies on Alcohol Suppl*, *13*, 52-62. doi: 10.15288/jsas.1999.s13.52. PMID: 10225488

BAER, J.S., KIVLAHAN, D.R. & MARLATT, G.A. (1995). High-risk drinking across the transition from high school to college. *Alcohol Clin Exp Res*, *19* (1), 54-61. doi: 10.1111/j.1530-0277.1995.tb01472.x. PMID: 7771663

BARRY, A.E. & PIAZZA-GARDNER, A.K. (2012). Drunkorexia: Understanding the co-occurrence of alcohol consumption and eating/exercise weight management behaviors. *Journal of American College Health*, *60* (3), 236-243. doi: 10.1080/07448481.2011.587487

BASTIANI PERGAMO, A. & DROGO, G.M. (2012). *I giovani e l'alcol.* Armando Editore, 2012.

BOOKER, R., NOVIK, M., GALLOWAY, R. & HOLMES, M.E. (2021). Relationship between physical activity intensities and drunkorexia severity among first-year college students. *Journal of American College Health*, *69* (6), 689-692. doi.org/10.1080/07448481.2019.1710151

BROWN, S.A., McGUE, M., MAGGS, J., SCHULENBERG, J., HINGSON, R., SWARTZWELDER, S., MARTIN, C., CHUNG, T., TAPERT, S.F., SHER, K., WINTERS, K.C., LOWMAN, C. & MURPHY, S. (2008). A developmental perspective on alcohol and youths 16 to 20 years of age. *Pediatrics*, *121* (4), S290-310. doi: 10.1542/peds.2007-2243D

BURKE, C., CREMEENS, J., VAIL-SMITH, K. & WOOLSEY, C.L. (2010). Drunkorexia: Calorie restriction prior to alcohol consumption among college freshman. *Journal of Alcohol and Drug Education*, *54* (2), 17-34.

CBS News (2008). *Drunkorexia: Health danger for women*. Retrieved from: https://www.cbsnews.com/news/drunkorexia-health-danger-for-women/

CHOQUETTE, E.M., ORDAZ, D.L., MELIOLI, T., DELAGE, B., CHABROL, H., RODGERS, R. & THOMPSON. J.K. (2018). Food and Alcohol Disturbance (FAD) in the US and France: Nationality and gender effects and relations to drive for thinness and alcohol use. *Eating behaviors*, *31*, 113-119. doi: 10.1016/j.eatbeh.2018.09.002

DUNN, E.C., LARIMER, M.E. & NEIGHBORS, C. (2002). Alcohol and drug-related negative consequences in college students with bulimia nervosa and binge eating disorder. *International Journal of Eating Disorders*, *32* (2), 171-8. doi: 10.1002/eat.10075. PMID: 12210659

EEB, ENIGM, GCAP, ITALIA (2018). *Le disuguaglianze in Italia*. Retrieved from http://www.gcapitalia.it/sdg-10-online-il-report-sulle-disuguaglianze-in-italia/

FITZSIMMONS-CRAFT, E.E., BALANTEKIN, K.N. & EICHEN, D.M, (2019). Screening and offering online programs for eating disorders: Reach, pathology, and differences across eating disorder status groups at 28 U.S. universities. *International Journal of Eating Disorders*, *52*, 1125-1136. doi.org/10.1002/eat.23134

GARNER, D.M., OLMSTED, M.P., BOHR, Y. & GARFINKEL, P.E. (1982). The eating attitudes test: Psychometric features and clinical correlates. *Psychological Medicine*, *12*, 871-878.

HORVATH, S.A., SHOREY, R.C. & RACINE, S.E. (2020). Emotion dysregulation as a correlate of food and alcohol disturbance in undergraduate students. *Eating Behaviors*, *38*, 101409. doi.org/10.1016/j.eatbeh.2020.101409

JONES, K.A., CHRYSSANTHAKIS, A. & GROOM, M.J. (2014). Impulsivity and drinking motives predict problem behaviours relating to alcohol use in university students. *Addictive Behaviors*, *39* (1), 289-96. doi: 10.1016/j.addbeh.2013.10.024

KERSHAW, S. (2008). *Starving themselves, cocktail in hand. New York Times*. Retrieved from: http://www.nytimes.com

KRAHN, D.D., KURTH, C.L., GOMBERG, E. & DREWNOWSKI, A. (2004). Pathological dieting and alcohol use in college women: A continuum of behaviors. *Eating Behaviors*, *6*, 43-52.

KUMAR, R., O'MALLEY, P.M. & JOHNSTON, L.D. (2002). Effects of school-level norms on student substance use. *Prev Sci*, *3*, 105-124. doi.org/10.1023/A:1015431300471

LUPI, M., MARTINOTTI, G. & DI GIANNANTONIO, M. (2017). Drunkorexia: An emerging trend in young adults. *Eating and Weight Disorders*, *22* (4), 619-622. doi: 10.1007/s40519-017-0429-2

MARRONARO, M., ROSSI, A., AQUILIO, E. & SCACCHIOLI, F. (2009). Eating Attitude Test (EAT-26) in una popolazione di 1809 studenti delle scuole medie della provincia de L'Aquila. *Psicologia clinica dello sviluppo. Rivista quadrimestrale*, *1*, 195-202. doi: 10.1449/29227

MINISTERO DELLA SALUTE (2020). *Disturbi dell'alimentazione.* Retrieved from: https://www.salute.gov.it/portale/donna/dettaglioContenutiDonna.jsp?lingua=italiano&id=4470&area=indennizzo&menu=patologie&tab=1

MOECK, E.K. & THOMAS, N.A. (2021). Food and alcohol disturbance in a broad age-range adult sample. *Eating Behaviors*, *41*, 101510. doi.org/10.1016/j.eatbeh.2021.101510

PALERMO, M., CHOQUETTE, E.M., AHLICH, E. & RANCOURT D. (2020). Food and alcohol disturbance by athlete status: The roles of drive for thinness, drive for muscularity, and sex. *Journal of American College Health*, *29*, 1-8. doi: 10.1080/07448481.2020.1713791

PATE, R.R., HEATH, G.W., DOWDA, M. & TROST, S.G. (1996). Associations between physical activity and other health behaviors in a representative sample of US adolescents. *American Journal of Public Health*, *86* (11), 1577-81. doi: 10.2105/ajph.86.11.1577. PMID: 8916523; PMCID: PMC1380692

PIAZZA-GARDNER, A.K. & BARRY, A.E. (2012). Examining physical activity levels and alcohol consumption: Are people who drink more active? *American Journal of Health Promotion*, *26* (3), 95-104. doi: 10.4278/ajhp.100929-LIT-328. PMID: 22208422

POMPILI, S. & LAGHI, F. (2020). Drunkorexia: Disordered eating behaviors and risky alcohol consumption among adolescents. *Journal of Health Psychology*, *25* (13-14), 2222-2232. doi:10.1177/1359105318791229

RANCOURT, D., AHLICH, E., CHOQUETTE, E.M., SIMON, J. & KELLEY, K. (2020). A comparison of food and alcohol disturbance (FAD) in sorority and non-sorority women. *Journal of American College Health*, *2*, 1-4. doi: 10.1080/07448481.2020.1740233

ROOSEN, K.M. & MILLS, J.S. (2015). Exploring the motives and mental health correlates of intentional food restriction prior to alcohol use in university students. *Journal of Health Psychology*, *20* (6), 875-886. doi: 10.1177/1359105315573436

SAUNDERS, J.B., AASLAND, O.G., BABOR, T.F., DE LA FUENTE, J.R. & GRANT, M. (1993). Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption - II. *Addiction (Abingdon, England)*, *88* (6), 791-804. doi.org/10.1111/j.1360-0443.1993.tb02093.x

SIMONS, R.M., HANSEN, J.M., SIMONS, J.S., HOVRUD, L. & HAHN, A.M. (2021). Drunkorexia: Normative behavior or gateway to alcohol and eating pathology? *Addictive Behaviors*, *112*, 106577. doi: 10.1016/j.addbeh.2020.106577

STEWART, S.H., ANGELOPOULOUS, M., BAKER, J.M. & BOLAND, F.J. (2000). Relations between dietary restraint and patterns of alcohol use in young adult women. *Psychology of Addictive Behaviors*, *14*, 77-82.

STOPPLER, M.C. (2008). *Drunkorexia, manorexia, diabulimia: New eating disorders?* Retrieved on May 25, 2009 from http://www.medicinenet.com/script/main/art.asp?articlekey=88014.

STRUZZO, P., FACCIO, S., MOSCATELLI, E., SCAFATO, E. & PRISMA GRUPPO (2006). Identificazione precoce dei bevitori a rischio in Assistenza primaria in Italia: Adattamento del questionario AUDIT e verifica dell'efficacia d'uso dello short-AUDIT test nel contesto nazionale. *Bollettino per le farmacodipendenze e l'alcoolismo*, *29*, 20-25.

THOMPSON-MEMMER, C., GLASSMAN, T. & DIEHR, A. (2018). Drunkorexia: A new term and diagnostic criteria. *Journal of American College Health*, *67* (7), 620-626. doi: 10.1080/07448481.2018.1500470

WESTERTERP, K.R., MEIJER, E.P., GORIS, A.H. & KESTER, A.D. (2004). Alcohol energy intake and habitual physical activity in older adults. *British Journal of Nutrition*, *91* (1), 149-52. doi: 10.1079/bjn20031013

WHITE, S., REYNOLDS-MALEAR, J. & CORDERO, E. (2011). Disorderly feeding and use of unhealthy methods for weight control in university students: 1995, 2002 and 2008. *Eating Disorders*, *19* (4), 323-334. doi: 10.1080/10640266.2011.584805

Websites

https://www.sdgwatcheurope.org/wp-content/uploads/2019/06/8.3.b-report-IT.pdf

DOI: 10.26387/bpa.2022.00004

Threshold values for significant changes in test-retest difference scores for the Wechsler Intelligence Scale for Children – Fourth Edition

Lina Pezzuti[1], James Dawe[1], Marco Lauriola[2]

[1] Department of Dynamic and Clinical Psychology, and Health Studies, Sapienza University of Rome, Italy
[2] Department of Social and Developmental Psychology, Sapienza University of Rome, Italy

lina.pezzuti@uniroma1.it

ABSTRACT. La presente ricerca si proponeva di studiare l'effetto della pratica a breve termine della *Wechsler Intelligence Scales for Children - Quarta Edizione (WISC-IV)* e di fornire valori soglia che consentano ai professionisti di valutare se esistono vere differenze nelle prestazioni individuali o se tali differenze siano dovute al caso. A un campione di 440 soggetti è stata somministrata la WISC-IV due volte con un intervallo medio di 30 giorni. I risultati mostrano che la pratica è più pronunciata quando si utilizzano i punteggi grezzi dei subtest rispetto a quelli ponderati. Sono stati ricavati dei valori soglia per valutare i cambiamenti significativi nei subtest e negli indici, consentendo così ai professionisti di valutare con maggiore precisione il significato clinico dei cambiamenti osservati durante una doppia somministrazione a breve termine della WISC-IV.

SUMMARY. One of the purposes of administering intelligence scales is to assess changes in cognitive functioning over time, from a few days to several years, to determine whether the examinee has progressed or regressed after treatment or other events. (e.g., an accident, a rehabilitation, etc.). The present research aimed to study the short-term practice effect of the Wechsler Intelligence Scales for Children – Fourth Edition and provide threshold values that allow practitioners to assess whether there are true differences in individual performance or whether these differences are due to chance. A sample of 440 subjects was administered the WISC-IV twice with an average interval of 30 days. The results show that practice is more pronounced when using raw subtest scores than when using weighted scores. Threshold values for assessing significant change in subtests and indices were obtained. For example, for the Full-Scale Intelligence Quotient, a difference between 6 and 27 IQ points between the first and second administration indicates a practice effect. Conversely, if the difference is equal to or less than 5 IQ points, then there was a decline, while if it is equal to or greater than 28 IQ points, there was an increase in performance not due to the practice effect. Therefore, these data should allow practitioners to more accurately assess the clinical significance of observed changes during a short-term dual administration.

Keywords: *WISC-IV, Test-retest, Individual changes*

INTRODUCTION

The first aim of administering one of the Wechsler intelligence scales (WPPSI-IV, WISC-IV, or WAIS-IV) is to assess an individual's cognitive functioning based on observed performance. A second aim is to evaluate change in cognitive functioning over time, from a few days (short term) to several years (long term), to determine whether the examinee has progressed or regressed after treatment or other events (e.g., an accident, a rehabilitation, etc.).

Intelligence is a psychological construct presumed to be relatively stable; thus, intelligence tests must produce similar scores from one time to another (Canivez & Watkins, 1998; Conley, 1984; Deary, Pattie & Starr, 2013; Heilbronner et al., 2010; Hunt, 2010; Mackintosh, 1998; Moffitt, Caspi, Hakness & Silva, 1993; Reeve & Bonaccio, 2011; Revelle, 2010; Simonton, 2011; Strauss, Sherman & Spreen, 2006; Wright, 2011). Accordingly, intelligence test scores were found to be relatively stable from childhood through adulthood (Chen & Siegler, 2000; Deary, Whalley, Lemmon, Crawford & Starr, 2000; Johnson, Gow, Corley, Starr & Deary, 2010) for both average and above-average samples (Reeve & Bonaccio, 2011; Simonton, 2011).

The information concerning the short-term test-retest stability of the *Wechsler Intelligence Scale for Children (WISC*; Wechsler, 1949), *WISC-Revised* (*WISC-R*; Wechsler, 1974), *WISC-Third Edition* (*WISC-III*; Wechsler, 1991), and *WISC-Fourth Edition* (*WISC-IV*; Wechsler, 2003a; 2003b) is available in their respective test manuals and has typically been conducted with nondisabled youths across retest intervals of fewer than 3 months. Likewise, there is extensive literature dealing with the stability of these WISC editions for a variety of test-retest intervals using healthy children, gifted children, and those with learning disabilities, attention deficit disorder, mental retardation, or other handicapping conditions (e.g., Anderson, Cronin & Kazmierski, 1989; Bauman, 1991; Canivez & Watkins, 1998, 1999, 2001; Ellzey & Karnes, 1990; Truscott, Narrett & Smith, 1994).

Most of these studies have indicated significant increases in verbal intelligence quotient (VIQ), performance intelligence quotient (PIQ), and full scale intelligence quotient (FSIQ) scores, with the largest increases in PIQ during short test-retest intervals (e.g., 30 days). The practice effects tend to disappear with longer retest intervals (e.g., 1-3 years). Canivez and Watkins (1999) concluded that the FSIQ of the WISC-III is the only score that possesses sufficient stability for interpreting individual cases.

The *Wechsler Intelligence Scale for Children – Fourth Edition* (*WISC-IV*; Wechsler 2003a; 2003b; Wechsler, 2012) is currently used in Italy for clinical practice with children and adolescents. Because approximately 60% of the items in its core subtests are new or revised (Watkins, 2010), the internal consistency and test-retest reliability of the WISC-IV cannot be assumed to be equivalent to previous editions and must be studied again.

The WISC-IV includes 10 core subtests (Bock design, Similarities, Digit span, Picture concepts, Coding, Vocabulary, Letter-number sequencing, Matrix reasoning, Comprehension and symbol search) and 5 supplemental subtests (Picture completion, Cancellation, Information, Arithmetic, and Word reasoning). The interpretation of the WISC-IV is mainly based on the full-scale intelligence quotient (FSIQ) and four index scores: the verbal comprehension index (VCI); the perceptual reasoning index (PRI); the working memory index (WMI); and the processing speed index (PSI). Two other indices, the general ability index (GAI) and the cognitive proficiency index (CPI), can be derived.

The stability of the WISC-IV scores across time has been investigated in several studies, mainly with long time intervals. The only short-term reliability study dates back to the USA standardization of the WISC-IV (Wechsler, 2003b), in which 243 children (52.3% female and 47.7% male) were tested twice, with a time interval ranging from 13 to 63 days (Mean = 32 days). Observing the results, the stability coefficients were satisfactory for the indices (from .80 to .90). Still, the short-term practice effects were observed with gains ranging from 2.1 points for the verbal comprehension index, to 7.1 points for processing speed index. On average, the increase was 5.6 points for the FSIQ.

All the other studies were carried out with long time intervals for the second administration of WISC-IV and are briefly described in Table 1. Such studies showed that: subtest long-term stability coefficients were consistently lower than the short-term stability coefficients reported for the normative samples; the long test-retest reliability coefficients for the subtests were generally lower than the scores for the four indices and the FSIQ; the FSIQ exhibited a higher long-term stability coefficient respect four WISC-IV indices. Some studies showed differences of 1-2 points in the subtest scores and up to 9-10 points in the four indices and FSIQ in a high

Table 1 – Summary of the long test-retest studies by the WISC-IV

Authors	**Mean time interval**	**Sample and age at first administering**	**Results:** *test-retest stability coefficients*	**Results:** *test-retest mean differences*
Ryan, Glass & Bartels, 2010	11 months	43 elementary and middle school children (mean age = 7.77 years, *SD* = 1.91)	Ranged from .54 for the PSI to .88 for the FSIQ.	42% of the FSIQ scores increased by 5 or more points on retest.
Lander, 2010	36 months	131 students with a learning disability	Ranged from .52 for the PSI to .65 for the FSIQ.	
Watkins & Smith, 2013	34 months	344 students aged 6,1 to 14,3 years	They were .72, .76, .66, .65, and .82 for the VCI, PRI, WMI, PSI, and FSIQ respectively.	The subtest scores did not differ by more than 1 point, and the index scores did not differ by more than 2 points. 44% of the students' VCI, PRI, WMI, and PSI scores increased by 10 or more points.
Bartoi, Issner, Hetterscheidt, January, Kuentzel & Barnett, 2015	22 months	51 children aged 8 to 16 years	Ranged from .58 for the PSI to .86 for the FSIQ.	78.4% of the children had test-retest differences up to 9 points for the FSIQ; similarly, 68.6%, 56.9%, 54.9%, and 54.9% of the children increased up to 9 points for the VCI, PRI, WMI, and PSI, respectively.
Kieng, Kieng & Geistlich, 2017	21 months	277 children aged 7 to 12 years	Ranged from .63 for the WMI to .80 for the FSIQ.	Half of the subjects shift from one intelligence classification category to the higher category.
Okada, Kawasaki, Shinomiya, Hoshino, Ino, Sakai, ... & Niwa, 2021	31 months	138 children with autism spectrum disorder (aged 5,5 to 16,8 years)	.83 for FSIQ, ranged from .62 to .79 for the four WISC-IV indices.	The mean of the FSIQ and VCI scores increased by 3.4 and 4.6 points in the second test.

percentages of the subjects tested. Some authors (Ryan, Glass & Bartels, 2010; Watkins & Smith, 2013) concluded that given this variability, it could not be assumed that the WISC-IV scores are consistent across long test-retest intervals.

In summary, most research has focused on the study of long-term stability, while there are no short-term reliability studies except for the US WISC-IV standardization data. Therefore, assessing the clinical significance of changes in retest performance must be carefully considered for short intervals. The test-retest procedure (from a few weeks to several years) can be used to address this. Indeed, numerous studies have shown that performance on a second test is superior to performance on the first test (Estevis, Basso & Combs, 2012; Salthouse, 2014; Sherman, Brooks, Iverson,

Slick & Strauss, 2011). However, to determine whether the test-retest difference score represents a significant change, threshold values must be obtained with correction for the effect of practice, measurement error, and regression to the mean, as done in a paper by Lecerf, Kieng and Geistlich (2017) on the WISC-IV.

If psychologists had instruments with perfect reliability, the performance observed on the test should be the same as that obtained on their retest. Nevertheless, no score of psychological measures has perfect reliability. Every test has a bias that makes it difficult to interpret the differences in observed scores between test and retest only in terms of cognitive functioning change. Therefore, determining if there has been a true change involves taking into account various psychometric phenomena, such as measurement error (i.e., the source of inaccuracies in test scores), and providing information about the reliability of test scores and practice effects, which reflect changes associated with repeated test administration.

It is worth noting that assessing change requires distinguishing statistical significance from the clinical relevance of a test-retest difference score (Brooks, Sherman, Iverson, Slick & Strauss, 2011; Jacobson & Truax, 1991). The 5% threshold ($p<.05$; for discussion Cohen, 1994; Reuchlin, 1992) is regularly used in psychology to determine whether a difference is statistically significant. However, statistically significant differences in intelligence tests do not mean clinically significant ones. Differences are considered of clinical interest if they are rare in the population. For some authors, this corresponds to a difference observed in less than 5% of the population (Chelune, 2003), for others in less than 10% (Kaufman & Kaufman, 2008), still others in less than 15% (Sattler, 2008). In this paper, we will use the 10% threshold.

As stated above, intelligence is a psychological construct that is assumed to be stable; so intelligence tests should produce similar scores from one time to the other. However, test-retest stability is not only characteristic of the test but may vary depending on: the type of the sample assessed (e.g., clinical or healthy); the size of the sample, i.e., Charter (2003) recommends a minimum of 400 participants for test-retest studies and Watson (2004) suggested a sample of at least 300, and possibly 400; the time interval between test and retest (short or long interval); the statistical methods used to calculate the reliability of the test-retest (e.g., Pearson correlation or intraclass correlations) and the practical effects (e.g., Cohen's *d*, Anova, or reliable change index). Finally, the type of scores examined influences the test-retest results, e.g., in the study by Lemay and colleagues (Lemay, Bedard, Rouleau& Tremblay, 2004), the reliabilities of the raw score on the Wechsler's scales were higher than the reliabilities assessed by demographically adjusted scores (i.e., scaled or weighted scores). Although the conversion from raw score to scaled score may be helpful to compare results gathered from different age groups, it also tends to induce a greater variability in the performance, limiting its usefulness in repeated administration, leading to a slight decrease in reliability (Lemay et al. 2004).

In the present paper, we studied the practice effect of repeated administration of the WISC-IV at a short time interval. We proposed threshold values to assess whether there are true differences in individual performance or whether these differences are due to chance. These threshold values are estimated using a sample of 440 subjects taking the WISC-IV twice. The threshold values reported here consider the effects of practice and measurement error. Therefore, these data allow practitioners to assess the clinical significance of observed changes more accurately during a short-term dual administration.

METHOD

Participants

We recruited 440 children and adolescents (219 girls and 221 boys) equally distribuited into the 11 age groups from 6 to 16 years of age reported in the WISC-IV manual. For each participant, biographical data and the education level of the parents were recorded according to four categories: primary, secondary, high school and university degree. As was done for the WISC-IV Italian standardization, the national data with which to compare the data of the present research were derived from estimates taken from a representative sample of the Italian population carried out on a sample of 7977 households with a total of 19907 individuals (Census of Italy Bank, 2010). The distribution of the individuals by age, gender, and parents' education is comparable to those of the above-mentioned survey. All of the children and adolescents examined had no previous psychological diagnosis and assessment, nor were they undergoing psychological treatment of any kind.

Instrument

The *Wechsler Intelligence Scale for Children – Fourth Edition* (*WISC-IV*; Wechsler 2003a; 2003b) is one of the most frequently used tests to assess the general intellectual functioning of Italian-speaking children. The WISC-IV is an individually administered test of intelligence for children ages 6 years 0 months through 16 years 11 months. It includes 10 core subtests (Bock design, Similarities, Digit span, Picture concepts, Coding, Vocabulary, Letter-number sequencing, Matrix reasoning, Comprehension and Symbol search) and 5 supplemental subtests (Picture completion, Cancellation, Information, Arithmetic, and Word reasoning). Each subtest has a standardized mean of 10 and a standard deviation of 3. The Italian WISC-IV was standardized on a nationally representative sample (N = 2200), closely approximating the 2010 database from the Census of Italy Bank (2010) on gender, parents' socioeconomic status, and parents' professional class. Currently, interpretation of the WISC-IV is mainly based on the full-scale intelligence quotient (FSIQ) and four index scores. The verbal comprehension index (VCI) is derived from the sum of Similarities, Vocabulary, and Comprehension scores; the perceptual reasoning index (PRI) from the sum of Block design, Picture concepts, and Matrix reasoning scores; the working memory index (WMI) from the sum of Digit span, and Letter-number sequencing scores; and the processing speed index (PSI) from the sum of Coding, and Symbol search scores. Two other Indices as the general ability index (GAI: the subtests are those of the VCI and PRI indices) and the cognitive proficiency index (CPI: the subtests are those of the WMI and PSI indices), can be derived. Finally, the FSIQ is obtained by adding the ten core subtest scores.

To interpret the outcomes at the WISC-IV, the raw scores of the subtests are converted to age-weighted scores, and the sum of the age-weighted scores of the subtests belonging to the indices is converted to the standard point IQ.

The reliability study of the Italian edition of the WISC-IV was mainly conducted through the split-half method, which is helpful in studying the homogeneity of the items composing the subtest. On the contrary, the reliability coefficients for 3 of the 15 subtests (Coding, Symbol search, and Deletion) and 2 of the 7 process scores (random deletion strategy and structured deletion strategy) were calculated using test-retest method. The reliability of the composite scores was instead studied using Mosier's (1943) formula.

The Italian WISC-IV standardization manual (Wechsler, 2012) reported average reliability indices varying between .74 for Symbol search and to .90 for Vocabulary, similar to the US edition values, which ranged between .79 for Symbol search and Cancellation to .90 for Letter-number sequencing. The average reliability of the four indices varied between .84 for the processing speed index to .94 for the verbal comprehension index in the Italian standardization sample; similarly, in the US edition, it varies between .88 for the processing speed index to .94 for the verbal comprehension index. Further, the average reliability of IQ was .96 in the Italian standardization and .97 in the US edition.

Procedure

Research participants were approached via primary and secondary schools in central Italy. Informed consent was requested from both parents and also information regarding any previous diagnoses and/or psychological treatment given to their children.

Participants were individually administered all 15 subtests of the Italian WISC-IV version twice, with test-retest intervals ranging from 17 to 38 days (M = 30; SD = 2.8 days).

Data analysis

- *Practice effect of repeated administration with the WISC-IV.* Cohen's d is used to estimate the effect size of practice effects due to repeated administrations for raw and weighted scores, for subtests and total scores, and for indices and FSIQ. An effect size ≥.80 is considered as a large practice effect; .50-.79, medium; .20-.49, small; and <.20, trivial (Cohen, 1988).
- *Reliable change index and threshold values for detecting decline or progression at the second evaluation.* Several statistical procedures exist to assess the significance of changes in test-retest performance and to account for bias and error (Basso, Carona, Lowery & Axelrod, 2002; Brooks, Strauss, Sherman, Iverson & Slick, 2009; Estevis et al., 2012). In this paper, we use the method initially proposed by Jacobson and colleagues (Jacobson, Follette & Revenstrof, 1984; Jacobson & Truax, 1991) and revised by Chelune and colleagues (Chelune, Naugle, Lüders, Sedlak & Awad, 1993), which results in the calculation

of a reliable change index (RCI). This method provides threshold values determining the magnitude of test-retest changes required for significant differences (≤10%). The RCI is calculated using simple descriptive statistics: mean and standard deviation values of the test and retest scores and test-retest correlations (i.e., a reliability coefficient). For each individual, the test-retest difference score is calculated. If the test-retest difference is greater than the RCI, it is considered a significant difference, rarely observed in the population.

In calculating the RCI, we use the standard error of a difference (SEM_{diff}) formula. According to Chelune et al. (1993), SEM_{diff} considers that measurement errors are unlikely to be the same at test and retest administrations (and therefore, reliability could not be the same) and that there might be an effect of practice. Thus, modifying the original formula of Jacobson and Truax (1991), who assumed that the measurement errors were the same at test and retest and that there was no effect of practice, the formula proposed by Chelune et al. (1993) is: SEM_{diff} = Squared root [$(SEM_{test})^2$ + $(SEM_{retest})^2$].

In this formula, the SEM_{test} is the standard error of measurement at the first evaluation (test), and the SEM_{retest} is the standard error of measurement at the second evaluation (retest). The SEM_{diff} value is multiplied by 1.645 and 1.96 to obtain the 90% and 95% confidence intervals respectively. The SEM_{diff} value is 0 for a test with perfect reliability and stability; when reliability decreases, SEM_{diff} increases.

In a later step, Chelune et al. (1993) incorporated practice effects (provided by the mean difference between test and retest) to identify threshold values for significant decline or progression at 90% respectively through the formulae: (SEM_{diff}(90%) - Practice effect) and (SEM_{diff}(90%) + Practice effect).

In summary, this SEM_{diff} procedure corrects the distribution of observed change scores by firstly taking into account measurement error and secondly taking into account the effects of the practice.

RESULTS

Examining Table 2 about the subtests, reveals a more pronounced practice effect for the raw subtest scores (*Mean d* = −.61) than for the weighted scores M*ean d* = −.50). More specifically, a small practice effect emerged for Letter-number sequencing and Arithmetic subtests, and a large practice effect for Coding and Picture completion subtests. For all other subtests, the effect was medium.

Regarding the WISC-IV indices (see Table 3), the same effect was found for both the sum of weighted and standardized points (i.e., IQs), with the impact of practice ranging from .55 (medium) for the working memory index, to 1.39 (very large) for full-scale intelligence quotient.

Tables 4 and 5 show the SEM_{diff} values and thresholds at 90% and 95% confidence intervals for deciding whether a significant change in the direction of decline or progression has occurred at a second short-term administration.

DISCUSSION AND CONCLUSION

Few studies have investigated the short-term stability of one of the most frequently used tests in the field of cognitive administration: the WISC-IV.

Given the distortions in psychological tests, particularly in intelligence scales, the mere difference between the scores observed on the examination and the retest cannot provide information on a possible change in cognitive functioning. The present study shows that correcting the test-retest difference scores for the effects of practice and measurement error present in the two administrations is crucial.

Comment for the first results of the study on the effect of practice in a short-time test-retest WISC-IV with 440 children is that this effect is stronger when using the raw scores of the subtests, which corresponding to the sums of corrected items, with respect to weighted scores that are age-corrected scaled scores transformed using WISC-IV Italian norms; a result that confirms the research of Lemay et al. (2004). The explanation for this result lies in the continuity of the raw scores and the age-weighted scores classification: i.e., if a subject obtains a raw score of 8 on the first administration and a score of 10 on the second administration, the change is evident, but if the raw scores of 8 and 10 correspond to the same age-weighted score, no difference between test and retest will emerge. So, it may be more beneficial for the practitioners to assess changes in subtest by using the first type of scores (raw) than the second (weighted).

A second reflection is on the threshold values found in the present study, which are higher than those found in the research by Lecerf et al. (2017) and Kieng et al. (Kieng, Rossier, Favez, Geistlich & Lecerf, 2015) resulting from the

Table 2 – Mean and SD for the early and late WISC-IV administration, and tests of mean differences, using raw and standard scores for the subtest (n = 440)

Test score	Early administration		Late administration		Test of mean difference		
	M	*SD*	*M*	*SD*	*t*	Diff. (or practice effect)	Cohen's *d*
Subtest				*Raw scores*			
Block design	34.66	13.66	39.10	14.23	−14.19	−4.45	−.68
Similarities	20.60	7.60	22.35	7.73	−12.75	−1.76	−.61
Digit span	17.18	4.07	18.25	4.51	−10.89	−1.07	−.52
Picture concepts	16.07	4.04	17.58	4.03	−12.89	−1.51	−.61
Coding	52.57	14.85	58.56	15.86	−17.12	−5.99	−.82
Vocabulary	36.82	9.44	38.72	9.45	−11.51	−1.91	−.55
Letter–number sequencing	17.75	4.31	18.62	3.94	−8.92	−.88	−.43
Matrix reasoning	19.30	6.09	21.03	6.16	−13.07	−1.76	−.62
Comprehension	19.25	6.07	20.70	6.04	−12.05	−1.46	−.58
Symbol search	28.43	8.13	31.09	8.19	−11.34	−2.67	−.54
Picture completion	21.60	6.38	24.13	6.63	−19.10	−2.53	−.91
Cancellation	83.79	23.34	91.84	24.16	−14.44	−8.06	−.69
Information	18.27	4.83	19.22	5.11	−11.90	−.95	−.57
Arithmetic	22.84	5.57	23.64	5.34	−8.97	−.80	−.43
Word reasoning	13.63	3.46	14.67	3.60	−12.89	−1.04	−.61
Subtest				*Weigthed scores*			
Block design	10.71	3.01	11.88	2.84	−11.16	−1.17	−.53
Similarities	10.49	2.43	11.44	2.52	−11.38	−.95	−.54
Digit span	11.19	2.54	12.03	2.63	−9.99	−.84	−.48
Picture concepts	10.24	2.85	11.69	2.81	−12.29	−1.45	−.59
Coding	10.31	2.60	11.93	2.80	−15.55	−1.62	−.74
Vocabulary	10.53	2.31	11.18	2.37	−8.02	−.65	−.38
Letter–number sequencing	10.97	2.91	11.69	2.79	−6.94	−.73	−.33
Matrix reasoning	10.62	2.69	11.84	2.82	−11.99	−1.22	−.57
Comprehension	9.89	2.72	10.80	2.61	−10.06	−.91	−.48
Symbol search	11.18	3.26	12.54	3.25	−9.37	−1.36	−.45
Picture completion	10.10	2.95	11.64	2.93	−16.68	−1.54	−.79
Cancellation	10.89	2.75	12.50	4.35	−7.98	−1.61	−.38
Information	10.14	2.39	10.88	2.40	−8.84	−.73	−.42
Arithmetic	11.19	2.77	11.77	2.68	−6.65	−.58	−.32
Word reasoning	10.35	2.57	11.35	2.41	−11.10	−1.01	−.53

Note. All comparisons are significant at $p<.00001$.

Table 3 – Mean and SD for the early and late WISC-IV administration, and tests of mean differences, using the sum of weighted scores and standard scores (IQ) for indices (n = 440)

Test score	Early administration		Late administration		Test of mean difference		
	M	*SD*	*M*	*SD*	*t*	Diff. (or practice effect)	Cohen's *d*
Indices			*Sum of weighted scores*				
Verbal comprehension index	30.91	5.89	33.43	6.01	−14.86	−2.52	−.71
Perceptual reasoning index	31.57	6.14	35.40	6.31	−19.52	−3.83	−.93
Working memory index	22.15	4.72	23.72	4.69	−11.52	−1.57	−.55
Processing speed index	21.48	5.05	24.47	5.11	−15.52	−2.98	−.74
Full−scale IQ	106.11	14.56	117.41	15.18	−29.18	−11.30	−1.39
General ability index	62.42	9.83	68.83	10.25	−24.11	−6.41	−1.15
Cognitive proficiency index	43.58	7.35	48.12	7.53	−20.16	−4.54	−.96
Indices			*Standard scores (IQ)*				
Verbal comprehension index	101.81	11.80	106.85	12.02	−14.86	−5.04	−.71
Perceptual reasoning index	103.18	13.34	111.53	13.73	−19.54	−8.35	−.93
Working memory index	106.46	14.15	111.17	14.07	−11.52	−4.71	−.55
Processing speed index	104.36	14.94	113.13	14.97	−15.48	−8.77	−.74
Full−scale IQ	104.84	11.85	114.08	12.34	−29.36	−11.30	−1.40
General ability index	102.72	11.54	110.24	12.04	−24.11	−7.53	−1.15
Cognitive Proficiency Index	106.64	13.72	115.11	14.06	−20.16	−8.48	−.96

Note. All comparisons are significant at $p<.00001$.

Table 4 – Practice effects, SEM_{diff}, and RCI threshold values for significant test-retest differences for the raw and weighted WISC-IV subtest scores

WISC-IV subtest	Practice effect	SEM_{diff}	SEM_{diff} (90%)	RCI SEM_{diff} (90% decline)	RCI SEM_{diff} (90% progression)	RCI threshold (90% decline)	RCI threshold (90% progression)	SEM_{diff} (95%)	RCI SEM_{diff} (95% decline)	RCI SEM_{diff} (95% progression)	RCI threshold (95% decline)	RCI threshold (95% progression)
Subtest raw scores												
Block design	4.45	7.74	12.69	8.24	17.14	8	17	15.16	10.72	19.61	11	20
Similarities	1.76	3.34	5.47	3.72	7.23	4	7	6.54	4.78	8.30	5	8
Digit span	1.07	2.28	3.74	2.67	4.80	3	5	4.47	3.40	5.53	3	6
Picture concepts	1.51	2.79	4.58	3.07	6.09	3	6	5.47	3.96	6.99	4	7
Coding	5.99	9.12	14.95	8.97	20.94	9	21	17.87	11.88	23.86	12	24
Vocabulary	1.91	3.92	6.43	4.52	8.34	5	8	7.68	5.78	9.59	6	10
Letter-number sequencing	.88	2.21	3.62	2.74	4.49	3	4	4.33	3.45	5.20	3	5
Matrix reasoning	1.76	3.26	5.34	3.58	7.10	4	7	6.39	4.63	8.14	5	8
Comprehension	1.46	2.89	4.74	3.28	6.21	3	6	5.67	4.21	7.14	4	7
Symbol search	2.67	5.46	8.95	6.28	11.61	6	12	10.70	8.03	13.36	8	13
Picture completion	2.53	3.62	5.94	3.41	8.47	3	8	7.10	4.57	9.63	5	10
Cancellation	8.06	13.80	22.64	14.58	30.70	15	31	27.06	19.00	35.11	19	35
Information	.95	1.90	3.12	2.17	4.07	2	4	3.73	2.78	4.67	3	5
Arithmetic	.80	2.03	3.33	2.53	4.14	3	4	3.98	3.18	4.79	3	5
Word reasoning	1.04	1.94	3.18	2.14	4.22	2	4	3.80	2.76	4.84	3	5

continued on next page

continued

WISC-IV subtest	Practice effect	SEM_{diff}	SEM_{diff} (90%)	RCI SEM_{diff} (90% decline)	RCI SEM_{diff} (90% progression)	RCI threshold (90% decline)	RCI threshold (90% progression)	SEM_{diff} (95%)	RCI SEM_{diff} (95% decline)	RCI SEM_{diff} (95% progression)	RCI threshold (95% decline)	RCI threshold (95% progression)
Subtest weighted scores												
Block design	1.17	2.39	3.92	2.75	5.09	3	5	4.69	3.52	5.85	4	6
Similarities	.95	1.93	3.16	2.21	4.12	2	4	3.78	2.83	4.74	3	5
Digit span	.84	1.91	3.13	2.29	3.97	2	4	3.74	2.90	4.59	3	5
Picture concepts	1.45	2.69	4.41	2.97	5.86	3	6	5.28	3.83	6.72	4	7
Coding	1.62	2.50	4.10	2.48	5.72	2	6	4.90	3.28	6.52	3	7
Vocabulary	.65	1.79	2.94	2.28	3.59	2	4	3.51	2.86	4.16	3	4
Letter-number sequencing	.73	2.28	3.74	3.01	4.46	3	4	4.46	3.74	5.19	4	5
Matrix reasoning	1.22	2.34	3.84	2.62	5.06	3	5	4.59	3.37	5.81	3	6
Comprehension	.91	2.05	3.36	2.45	4.27	2	4	4.01	3.10	4.92	3	5
Symbol search	1.36	3.20	5.25	3.88	6.61	4	7	6.27	4.91	7.63	5	8
Picture completion	1.54	2.32	3.81	2.27	5.35	2	5	4.55	3.01	6.09	3	6
Cancellation	1.61	4.22	6.92	5.31	8.53	5	9	8.27	6.66	9.88	7	10
Information	.73	1.85	3.03	2.29	3.76	2	4	3.62	2.88	4.35	3	4
Arithmetic	.58	1.89	3.11	2.53	3.68	3	4	3.71	3.13	4.29	3	4
Word reasoning	1.01	2.06	3.38	2.38	4.39	2	4	4.04	3.04	5.05	3	5

Table 5 – Practice effects, SEM_{diff}, and RCI threshold values for significant test-retest differences for raw (sum of weighted scores) and standard (IQs) WISC-IV indices

WISC-IV indices	Practice effect	SEM_{diff}	SEM_{diff} (90%)	RCI SEM_{diff} (90% decline)	RCI SEM_{diff} (90% progression)	RCI threshold (90% decline)	RCI threshold (90% progression)	SEM_{diff} (95%)	RCI SEM_{diff} (95% decline)	RCI SEM_{diff} (95% progression)	RCI threshold (95% decline)	RCI threshold (95% progression)
Sum of weighted scores												
Verbal Comprehension index	2.52	4.17	6.84	4.32	9.36	4	9	8.18	5.66	10.69	6	11
Perceptual reasoning index	3.83	5.16	8.46	4.62	12.29	5	12	10.11	6.28	13.94	6	14
Working memory index	1.57	3.17	5.20	3.63	6.77	4	7	6.22	4.65	7.79	5	8
Processing speed index	2.98	4.63	7.59	4.61	10.57	5	11	9.07	6.09	12.05	6	12
Full-scale IQ	11.30	12.25	20.09	8.79	31.39	9	31	24.01	12.71	35.31	13	35
General ability index	6.41	7.74	12.69	6.28	19.09	6	19	15.16	8.76	21.57	9	22
Cognitive Competency index	4.54	6.01	9.86	5.32	14.40	5	14	11.78	7.24	16.32	7	16
Standard scores (IQ)												
Verbal comprehension index	5.04	8.35	13.69	8.65	18.74	9	19	16.28	11.24	21.32	11	21
Perceptual reasoning index	8.35	11.23	18.41	10.06	26.76	10	27	21.89	13.54	30.24	14	30
Working memory index	4.71	9.52	15.61	10.90	20.32	11	20	18.56	13.85	23.27	14	23
Processing speed index	8.77	13.63	22.36	13.59	31.13	14	31	26.59	17.82	35.35	18	35
Full-scale IQ	11.30	9.98	16.37	5.08	27.67	5	28	19.47	8.17	30.76	8	31
General ability index	7.53	9.09	14.90	7.38	22.43	7	22	17.72	10.19	25.24	10	25
Cognitive competency index	8.48	11.22	18.41	9.93	26.88	10	27	21.89	13.41	30.36	13	30

long-term (more than 1 year) WISC-IV test-retest studies, where the practice effects are lower.

In particular, Tables 4 and 5 show the SEM_{diff} values and thresholds for deciding whether a significant change in the direction of decline or progression has occurred at a second short-term administration. For example, if we administer the Block design subtest twice, a raw score difference between 9-16 points (for 90% confidence interval) indicates that there was only a practice effect between the first and second administrations; conversely, a difference less than or equal to 8 points indicates a decline in performance, whereas a difference greater than or equal to 17 indicates a progression or increase in the ability measured by the subtest. Similarly, thresholds for indices can be used. In particular, for the FSIQ (see the standard score in Table 5), a difference between 6 and 27 IQ points between the first and second administration indicates a practice effect. Conversely, if the difference is equal to or less than 5 IQ points, then there was a decline, while if it is equal to or greater than 28 IQ points, there was an increase in performance not due to the practice effect.

This study is not without limitations: the sample consists only of children with typical development. Test-retest studies should therefore be conducted on other cultural and clinical samples to assess the generalizability of the threshold values proposed here. Another limitation is that these methods assume practice effects are the same for all children, but studies have shown this is not the case. For example, children with better intellectual abilities tend to have more significant practice effects on the retest. Despite these limitations, the threshold values proposed here should help practitioners identify whether the changes observed in a short period are significant or not, i.e., present in less than 10% of the population with typical development.

References

ANDERSON, P.L., CRONIN, M.E. & KAZMIERSKI, S. (1989). WISC-R stability and re-evaluation of learning-disabled students. *Journal of Clinical Psychology*, *45*, 941-944. doi:10.1002/1097-4679(198911)45:6_941: AID-JCLP2270450619_3.0.CO;2-P

BARTOI, M., ISSNER, J.B., HETTERSCHEIDT, L., JANUARY, A.M., KUENTZEL, J.G. & BARNETT, D. (2015). Attention problems and stability of WISC-IV scores among clinically referred children. *Applied Neuropsychology: Child*, *4*, 133-140. http://dx.doi.org/10.1080/21622965.2013.811075

BASSO, M.R., CARONA, F.D., LOWERY, N. & AXELROD, B.N. (2002). Practice effects on the WAIS-III across 3- and 6-month intervals. *The Clinical Neuropsychologist*, *16*, 57-63. http://dx.doi.org/10.1076/clin.16.1.57.8329

BAUMAN, E. (1991). Determinants of WISC-R subtest stability in children with learning difficulties. *Journal of Clinical Psychology, 47*, 430-435. doi:10.1002/1097-4679(199105)47:3_430::AID-JCLP2270470317_3.0.CO;2-N

BROOKS, B.L., SHERMAN, E.M.S., IVERSON, G.L., SLICK, D.J. & STRAUSS, E. (2011). Psychometric foundations for the interpretation of neuropsychological test results. In M.R. Schoenberg & J.G. Scott (Eds.), *The little black book of neuropsychology: A syndrome-based approach*. Springer Science + Business Media.

BROOKS, B.L., STRAUSS, E., SHERMAN, E.M.S., IVERSON, G.L. & SLICK, D.J. (2009). Developments in neuropsychological assessment: Refining psychometric and clinical interpretive methods. *Canadian Psychology*, *50*, 196-209. http://dx.doi.org/10.1037/a0016066

CANIVEZ, G.L. & WATKINS, M.W. (1998). Long-term stability of the Wechsler Intelligence Scale for Children – Third Edition. *Psychological Assessment*, *10* (3), 285-291. https://doi.org/10.1037/1040-3590.10.3.285

CANIVEZ, G.L. & WATKINS, M.W. (1999). Long-term stability of the Wechsler Intelligence Scale for Children – Third Edition

among demographic subgroups: Gender, race/ethnicity, and age. *Journal of Psychological Assessment*, *17*, 300-313. doi:10.1177/073428299901700401

CANIVEZ, G.L. & WATKINS, M.W. (2001). Long-term stability of the Wechsler Intelligence Scale for Children – Third Edition among students with disabilities. *School Psychology Review*, 30, 438-453.

CHARTER, R.A. (2003). Study samples are too small to produce sufficiently precise reliability coefficients. *The Journal of General Psychology*, *130* (2), 117-129.

CHELUNE, G.J. (2003). Assessing reliable neuropsychological change. In R.D. Franklin (Ed.), *Prediction in forensic and neuropsychology: Sound statistical practices*. Erlbaum.

CHELUNE, G.J., NAUGLE, R.I., LÜDERS, H., SEDLAK, J. & AWAD, I.A. (1993). Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology*, *7*, 41-52. https://doi.org/10.1037/0894-4105.7.1.41

CHEN, Z. & SIEGLER, R.S. (2000). Intellectual development in childhood. In R.J. Sternberg (Ed.), *Handbook of intelligence*. Cambridge University Press. doi:10.1017/CBO9780511807947.006

COHEN, J. (1994). The earth is round (p<.05). *American Psychologist*, *4* (12), 997-1003.

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

CONLEY, J.J. (1984). The hierarchy of consistency: A review and model of longitudinal findings on adult individual differences in intelligence, personality and self-opinion. *Personality and Individual Differences*, *5*, 11-25.

DEARY, I.J., PATTIE, A. & STARR, J.M. (2013). The stability of intelligence from age 11 to age 90 years: The Lothian birth cohort of 1921. *Psychological Science*, *24*, 2361-2368. http://dx.doi.org/10.1177/0956797613486487

DEARY, I.J., WHALLEY, L.J., LEMMON, H., CRAWFORD, J.R. & STARR, J.M. (2000). The stability of individual differences in mental ability from childhood to old age: Follow-up of the 1932 Scottish Mental Survey. *Intelligence*, *28*, 49-55. http://dx.doi.org/10.1016/S0160-2896(99)00031-8

ELLZEY, J.T. & KARNES, F.A. (1990). Test-retest stability of WISC-R IQs among gifted students. *Psychological Reports*, *66*, 1023-1026. doi: 10.2466/pr0.1990.66.3.1023

ESTEVIS, E., BASSO, M.R. & COMBS, D. (2012). Effects of practice on the Wechsler Adult Intelligence Scale – IV across 3- and 6-month intervals. *The Clinical Neuropsychologist*, *26*, 239-254. http://dx.doi.org/10.1080/13854046.2012.659219

HEILBRONNER, R.L., SWEET, J.J., ATTIX, D.K., KRULL, K.R., HENRY, G.K. & HART, R.P. (2010). Official position of the American Academy of Clinical Neuropsychology on serial neuropsychological assessments: The utility and challenges of repeat test administrations in clinical and forensic contexts. *The Clinical Neuropsychologist*, *24*, 1267-1278. doi:10.1080/13854046.2010.526785

HUNT, E. (2010). *Human intelligence*. Cambridge University Press.

JACOBSON, N.S., FOLLETTE, W.C. & REVENSTROF, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, *15*, 336-352.

JACOBSON, N.S. & TRUAX, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12-19.

JOHNSON, W., GOW, A.J., CORLEY, S., STARR, J.M. & DEARY, I.J. (2010). Location in cognitive and residential space at age 70 reflects a lifelong trait over parental and environmental circumstances: The Lothian birth cohort 1936. *Intelligence*, *38*, 402-411. doi:10.1016/j.intell.2010.04.001

KAUFMAN, A.S. & KAUFMAN, N.L. (2008). *KABC-II batterie pour l'examen psychologique de l'enfant* (2 éd.). ECPA.

KIENG, S., ROSSIER, J., FAVEZ, N., GEISTLICH, S. & LECERF, T. (2015). Stabilité à long terme des scores du WISC-IV: Forceset faiblesses personnelles. *Pratiques psychologiques*, *21*, 137-154. http://dx.doi.org/10.1016/j.prps.2015.03.002

LANDER, J. (2010). *Long-term stability of scores on the Wechsler Intelligence Scale for Children – Fourth Edition in children with learning disabilities*. Fairleigh Dickinson University.

LECERF, T., KIENG, S. & GEISTLICH, S. (2017). WISC-IV: Valeurs seuils pour des changements significatifs des scores de différence test-retest [WISC-IV: Cutoff values for meaningful test-retest difference scores]. *Pratiques Psychologiques, 23* (4), 345-358. https://doi.org/10.1016/j.prps.2016.07.003

LEMAY, S., BEDARD, M.A., ROULEAU, I. & TREMBLAY, P.L. (2004). Practice effect and test-retest reliability of attentional and executive tests in middle-aged to elderly subjects. *The Clinical Neuropsychologist*, *18*, 284-302. doi:10.1080/13854040490501718

MACKINTOSH, N.J. (1998). *IQ and human intelligence*. New York, NY: Oxford University Press.

MOFFITT, T.E., CASPI, A., HARKNESS, A.R. & SILVA, P.A. (1993). The natural history of change in intellectual performance: Who changes? How much? Is it meaningful? *Child Psychology & Psychiatry & Allied Disciplines*, *34* (4), 455-506. https://doi.org/10.1111/j.1469-7610.1993.tb01031.x

MOSIER, C.I. (1943). On the reliability of a weighted composite.

Psychometrika, *8*, 161-168. https://doi.org/10.1007/BF02288700

OKADA, S., KAWASAKI, Y., SHINOMIYA, M., HOSHINO, H., INO, T., SAKAI, K., ... & NIWA, S.I. (2021). Long-term stability of the WISC-IV in children with autism spectrum disorder. *International Journal of School & Educational Psychology*, 1-12. https://doi.org/10.1080/21683603.2021.1930307

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

REEVE, C.L. & BONACCIO, S. (2011). On the myth and the reality of the temporal validity degradation of general mental ability test scores. *Intelligence*, *39*, 255-272. doi:10.1016/j.intell.2011.06.009

REUCHLIN, M. (1992). *Introduction à la recherche en psychologie*. Nathan.

REVELLE, W. (2010). *An introduction to psychometric theory with applications in R*. Retrieved from http://www.personality-project.org/r/book/

RYAN, J.J., GLASS, L.A. & BARTELS, J.M. (2010). Stability of the WISC-IV in a sample of elementary and middle school children. *Applied Neuropsychology*, *17*, 68-72. http://dx.doi.org/10.1080/09084280903297933Saklofske

SALTHOUSE, T. (2014). Frequent assessments may obscure cognitive decline. *Psychological Assessment*, *26*, 1063-1069. http://dx.doi.org/10.1037/pas0000007

SATTLER, J.M. (2008). Assessment of children. *Cognitive foundations (5th ed.)*. Jerome M. Sattler, Publisher, Inc.

SHERMAN, E.M.S., BROOKS, B.L., IVERSON, G.L., SLICK, D.J. & STRAUSS, E. (2011). Reliability and validity in neuropsychology. In M.R. Schoenberg & J.G. Scott (Eds.), *The little black book of neuropsychology*. Springer.

SHROUT, P.E. & FLEISS, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86* (2), 420-428.

SIMONTON, D.K. (2011). Exceptional talent and genius. In T. Chamorro-Premuzic, S. von Stumm & A. Furnham (Eds.), *Wiley-Blackwell handbook of individual differences.* Blackwell.

STRAUSS, E., SHERMAN, E.M.S. & SPREEN, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary*. New York, NY: Oxford University Press.

TRUSCOTT, S.D., NARRETT, C.M. & SMITH, S.E. (1994). WISC-R subtest reliability overtime: Implications for practice and research. *Psychological Reports, 74*, 147-156. http://dx.doi.org/10.2466/pr0.1994.74.1.147

WATKINS, M.W. & CANIVEZ, G.L. (2004). Temporal stability of WISC-III subtest composite: Strengths and weaknesses. *Psychological Assessment*, *16*, 133-138.

WATKINS, M.W. & SMITH, L.G. (2013). Long-term stability of the Wechsler Intelligence Scale for Children – Fourth Edition. *Psychological Assessment*, *25* (2), 477-483. https://doi.org/10.1037/a0031653

WATSON, D. (2004). Stability versus change, dependability versus error: Issues in the assessment of personality over time. *Journal of Research in Personality*, *38*, 319-350. doi:10.1016/j.jrp.2004.03.001

WECHSLER, D. (1949). *Manual for the Wechsler Intelligence Scale for Children*. Psychological Corporation.

WECHSLER, D. (1974). *Manual for the Wechsler Intelligence Scale for Children – Revised*. Psychological Corporation.

WECHSLER, D. (1991). *WISC-III manual.* Psychological Corporation.

WECHSLER, D. (2003a). *WISC-IV administration and scoring manual.* Psychological Corporation.

WECHSLER, D. (2003b). *WISC-IV technical and interpretive manual*. Psychological Corporation.

WECHSLER, D. (2012). *WISC-IV. Manuale di somministrazione e scoring*. It. ad. A. Orsini & L. Pezzuti (Eds.). Firenze: Giunti O.S. Organizzazioni Speciali.

WRIGHT, A. J. (2011). *Conducting psychological assessment: A guide for practitioners*.

Declaration of conflicting interests. One of the authors (L. Pezzuti) receives royalties from sales of the WISC-IV (Wechsler, 2012, It ad. Orsini & Pezzuti).

DOI: 10.26387/bpa.2022.00005

Potential benefits of human-animal interaction among nurses pet owners and non-pet owners: A pilot study

Daniela Acquadro Maran[1], Maria Michela Gianino[2], Giorgia Serra[3], Claudio Giovanni Cortese[1]

[1] Department of Psychology, Università di Torino, Italy
[2] Department of Public Health Sciences and Pediatrics, Università di Torino, Italy
[3] A.S.L., Presidio Ospedaliero Cardinal Massaia, Asti, Italy

mariola.gianino@unito.it

ABSTRACT. Lo scopo di questo lavoro è quello di analizzare il rapporto tra benessere psicologico, soddisfazione di vita, strategie di coping e stili di attaccamento ad un animale domestico nel personale infermieristico. 147 infermieri hanno compilato un questionario, 79 dei quali erano proprietari di animali domestici. I risultati di questo studio possono essere utili per comprendere l'impatto che la cura di un animale domestico può avere sulla vita personale e professionale, in particolare sulla percezione della qualità della vita e gestione degli eventi stressanti.

SUMMARY. The aim of this work was to analyse the relationship between psychological well-being, life satisfaction, optimism, coping strategies and attachment to a pet in nurses. Data from pet owners were compared with those from non-pet owners working in the same health care facility. A self-reported questionnaire was used. 147 nurses completed the questionnaire, 79 of whom were pet owners. Overall, the results of this study show that the participants perceive a good quality of life, are satisfied with their work and life, and are optimistic about the future. Nurses with and without pets did not differ in perceived quality of life and life satisfaction between pet owners and non-pet owners. However, nurses who did not own a pet tended to use the coping strategy of emotional support more than pet owners. The findings contribute to a better understanding of how nurses who do and do not own a pet assess their quality of life, their satisfaction with life and work, and their coping with stressful events. In addition, the results of this study may be useful in learning first-hand about whether to get, care for, or keep a pet and the impact it may have on personal and professional lives.

Keywords: *Health care professional, Pets, Well-being, Attachment*

INTRODUCTION

Research on pet owners has shown that these individuals experience a variety of physical, psychological, and social benefits resulting from human-animal interactions (Aydin et al., 2012; Beetz, Uvnäs-Moberg, Julius & Kotrschal, 2012; Polheber & Matchock, 2013; Shiloh, Sorek & Terkel, 2003). At the same time, there are studies that have found either no significant relationship or an inverse relationship (Antonacopoulos & Pychyl, 2010; Wells, 2009; Wisdom, Saedi & Green, 2009). These inconsistent results could be due to a number of methodological differences and potential issues that make comparison difficult. These include sample size, homogeneous samples and vulnerable populations, different variables used to assess the human-animal interaction, and different measures used to assess the same constructs.

Much of the literature focuses on the benefits of human-animal interaction for vulnerable populations, for example, psychiatric patients, hospitalised patients, and the elderly (Berget, Ekeberg & Braastad, 2008; Berget, Ekeberg, Pedersen & Braastad, 2011; Cole, Gawlinski, Steers & Kotlerman, 2007; Colombo, Buono, Smania, Raviola & De Leo, 2006; Friedmann, Thomas, Cook, Tsai & Picot, 2007; Kramer, Friedmann & Bernstein, 2009; Pedersen, Martinsen, Berget & Braastad, 2012; Siegel, 1990, 2011; Villalta-Gil et al., 2009). Selecting samples from subgroups of the population, such as elderly residents of a nursing home, limits the generalizability and comparability of results. However, they provide greater insight into the direct effects of human-animal interaction in specific populations (Buckle, 2015; Friedmann & Gee, 2019; Le Roux, Swart & Swart, 2014). In addition, there are a large number of questionnaires that measure a variety of different variables related to human-animal interaction, as well as a number of different ways to assess them. Anderson (2007) provides a comprehensive collection of different questionnaires that assess a variety of the same and different aspects of human-animal interaction. Some assess attachment, bonding, attitudes, behaviours toward animals, responsibilities toward pets, expectations of pets, and quality of life of pets. This makes it difficult to compare previous literature because different variables have been assessed.

A potential problem with research on pet ownership is that much of the research focuses on pet ownership and does not consider participants' attachment or attitude toward the pet (Chur-Hansen, Peacock & Winefield, 2012). Not all pet owners necessarily have a close bond with their pet. The relationship can range from a dysfunctional relationship where people neglect or mistreat their pet to a very supportive, caring relationship (Anderson, Lord, Hill & McCune, 2015). Another possible explanation for these conflicting findings is that these benefits are not directly attributable to a specific cause. Benefits may be attributed to the support and comfort a person receives from their pet or therapy animal (Cohen, Gottlieb & Underwood, 2000). Benefits are also attributed to the satisfaction of an innate desire to connect with other living beings (Bjick, 2013; Fine & Beck, 2010). Another explanation could be due to the fact that not all pet owners necessarily have a close relationship with their pet or have a caregiving relationship that implies a decline in quality of life (Anderson et al., 2015). While research on HAIs is increasing, most of the literature focuses on the benefits of animal-assisted interaction (AAI) (Acquadro Maran et al., 2022; Buckle, 2015; Le Roux & Kemp, 2009; Le Roux et al., 2014; Lubbe & Scholtz, 2013; Odendaal, 2000). There is still very little research on pet ownership and attachment in nurses (Black, Winefield & Chur-Hansen, 2011), their relation to life satisfaction and work outcomes (Stallter & Gustin, 2021). Research on veterinary staff, for example, has found that attachment to the pet is associated with lower job satisfaction (Black et al., 2011). However, research conducted with subjects working in the health care sector has shown that "playing with the pet" is one of the most commonly used coping strategies to increase the workplace well-being (Chipas et al., 2012). Research on coping strategies to manage work-related stress points to the benefits of using multiple adaptive strategies that can impact increased perceptions of well-being, thus the caring for pets could be included in these strategies.

Potential benefit for pet ownership

Previous research has found neurological evidence of HAI benefits (Handlin et al., 2011; Miller et al., 2009; Nagasawa, Kikusui, Onaka & Ohta, 2008; Odendaal, 2000; Odendaal & Meintjjies, 2003). In these studies, significant increases in pet owners' oxytocin levels were found after interaction with a dog and, in general, interaction with a pet can produce positive changes in other hormones such as cortisol, B-endorphin, prolactin, and dopamine (Handlin et al., 2011). Also, the activity with pets has a positive impact. For example, walking dogs increased physical activity and has been associated with both physical (for example, reductions

in blood pressure, risk of obesity, and so on; Chandler et al., 2017) and mental health (for example, higher life satisfaction and lower anxiety, and depression; Gilmour, 2007). Previous research has found that dog walkers are significantly more likely to achieve recommended activity levels than non-dog walkers and non-dog owners (Brown & Rhodes, 2006; Feng et al., 2014; Ham & Epping, 2006; Thorpe, Simonsick & Brach, 2006). In a sample of older participants, dog owners were 12% more active than non-dog owners (Feng et al., 2014). However, conflicting results were reported by Rijken and Beek (2011), in their investigation pet owners did not differ from non-pet owners in terms of physical health. Moreover, research examining loneliness and attachment to a pet has yielded conflicting results in both control group studies and surveys (Antonacopoulos & Pychy, 2010; Buckle, 2015; Wood, Giles-Corti, Bulsara & Bosch, 2007). In an Australian survey, non-pet owners were two times more likely to report being lonely than pet owners (Wood et al., 2007). In contrast, a Canadian study found no difference in loneliness scores between pet owners and non-pet owners (Antonacopoulos & Pychy, 2010). Loneliness is related with the 'hopelessness' that is the perception of loneliness as a constant state (Taube, Jakobsson, Midlöv & Kristensson., 2016). Aside from the physical and psychological benefits of HAI, pets have been found to be a source of social support (Kikusui, Winslow & Mori, 2006; Risley-Curtiss, 2010). Pets have been found to facilitate social interaction between strangers (Cutt, Giles-Corti, Wood, Knuiman & Burke, 2008; Guéguen & Ciccotti, 2008; Hill, Gaines & Wilson, 2008; Knight & Edwards, 2008; Lee, Shepley & Huang, 2009; Wood et al., 2007). Risley-Curtiss (2010) suggested that pets provide us with social support because of the unconditional love and affection received from them and the needs to care for others.

Pet attachment

Humans are biologically predisposed to seek emotional connections with others who provide emotional and physical support (Bowlby, 1977). These emotional bonds are not limited to human relationships, but are also experienced in human-animal relationships (Beck, 2014). These emotional bonds, also referred to as pet bonds, are characterised by similar qualities to those experienced in human relationships, such as love, affection, support (Field, Orsini, Gavish & Packman, 2009; Noonan, 2008). One possible reason is that a pet's love is unconditional, loyal, and free of judgement and pretence (McCune et al., 2014). Like children, pets depend on their caregiver to provide them with food, shelter, and mental stimulation. As a result, pet owners tend to assume a similar caregiving role for their pets (Bowlby, 1982; Mikulincer & Shaver, 2007). Pets have been repeatedly found to be considered part of the family (Adrian, Deliramich & Frueh, 2009; Hunt & Padilla, 2006; Kaufman & Kaufman, 2006).

Chur-Hansen, Winefield and Beckwith (2009) found that the relationship between pet attachment and human health benefits (physical and psychological) is bell-shaped rather than linear. According to this hypothesis, pet owners with moderate attachment to their pet would achieve the most health benefits. It is suggested that an overly dependent relationship with a pet may lead to mental health problems. Conflicting results from Smolkovic et al. (Smolkovic, Fajfar & Mlinaric, 2012) and Winefield et al. (Winefield, Black & Le Roux, 2008) shown that adults had a non-significant relationship between attachment to a pet and social support. One possible explanation for these findings is that people who have problems with attachment to other people become more attached to their pets to compensate for the lack of social support (Kurdek, 2009; Smolkovic et al., 2012). Interesting, some studies have found a nonsignificant difference in attachment to pets between the sexes (Herzog, 2007; 2011; Prato-Previde, Fallani & Valsecchi, 2006).

Satisfaction with life and human-animal interaction

Bao and Schreer (2016) found that pet owners scored significantly higher on life satisfaction than non-pet owners. The same study found a non-significant difference in happiness, positive or negative emotions. Results from a large German study showed a positive relationship between attachment to a pet and life satisfaction (Luhmann & Kalitzki, 2016). The researchers suggested that people who lack close human relationships may compensate by having a deep bond with their pet. Moreover, a strong attachment to a pet might lead the person to invest little time in other interpersonal relationships (Luhmann & Kalitzki, 2016). In contrast, a very large Canadian study found that pet ownership was negatively related to life satisfaction (Himsworth & Rock, 2013). Mháistir (2013) collected data from a sample of Irish older adults recruited by a senior citizens' organisation. The

researcher suggested that while the pets may have buffered their owners' well-being, their financial worries negatively impacted their life satisfaction (Mháistir, 2013). Even more, some research suggested that life satisfaction does not differ between pet and non-pet owners (El-Alayli, Lystad, Webb, Hollingsworth & Ciolli, 2006; Ramirez & Hernandez, 2014). For example, in a sample of Mexican adults, a non-significant difference in life satisfaction was found between pet owners and non-pet owners (Ramirez & Hernandez, 2014).

The purpose of this paper was to analyse the relationship between psychological well-being, life satisfaction, optimism, coping strategies, and pet attachment among nurses. Data from pet owners were compared with non-pet owners working in the same health care organization. It is hypothesized that those who have a pet will experience have higher well-being and satisfaction scores, less perceived hopelessness and use adaptive coping strategies (Hp_1). Within subjects who have a pet, a comparison will be made between those who engage with the pets' care in the first person, both in terms of satisfaction and attachment and those who do not engage with the pets' care themselves. It is hypothesized that who engage with the pet's care in the first person experienced less life satisfaction and high person substitution attachment (Hp_{2a}) while who don't engage in the first person experienced high life satisfaction and general attachment ($Hp_{2b)}$. About coping strategies, our expectation is to find no difference among who is engaged and who isn't engaged with the pets' care in the first person (Hp_3).

METHOD

Ethical statement and procedure

All ethical guidelines required for conducting research projects with human subjects were followed, including compliance with legal requirements in Italy. This research project was approved by the local ethics committee (prot. N. 14520-14/02/2020) After an initial meeting with the management of a healthcare facility located in North-western Italy, the questionnaire was distributed to 275 healthcare workers there. The cover sheet clearly explained the research objective, the voluntary nature of participation, the anonymity of the data, and the elaboration of the results. Thus, return of the questionnaires signified consent. The data were collected by two research assistants who had been previously trained by the researchers. Together with the questionnaire, each participant received an information letter and the consent form. It took approximately 20 minutes to complete the questionnaire. The request was to complete the questionnaire within 15 working days and drop it in a special box in the locker room. Respondents participated in the study voluntarily and received no compensation for their participation.

Measures

The Italian short version of the *Psychological General Well Being Index – A (PGWBI-A)* (Testa et al., 2016) was used to assess the perceived quality of life and psychological well-being. The PGWBI-A can be considered one of the first quality of life assessment instruments. To achieve greater acceptability, a short form reduced to six questions (one for each domain) was validated in 2016. The six questions cover the following domain: anxiety, depression, positive well-being, self-control, general health, vitality. An increasing score, ranging from 0 to 30, or higher scores indicate better mental well-being.

The *Brief COPE scale* (Carver, 1997; Conti, 1999) was used to assess the coping strategies. Each strategy corresponds to a couple of questions, proposed in random order; the strategy to which assigned a higher score, from 2 to 8, will be the most frequently adopted. The choice to use this scale lies in the fact that the adoption of strategies adaptive, that is, constructive and functional, or, on the contrary, maladaptive, is considered an indicator of the level of psychological well-being or malaise. The coping strategies proposed by the tool are the following: venting, denial, substance use, behavioural disengagement, active coping, positive reframing, planning, humour, acceptance, emotional support, religion, instrumental support, self-blame.

The *Satisfaction with Life Scale* (*SWLS*; Diener, Emmons, Larsen & Griffin, 1985) was used to assess the satisfaction for life and for work. The range goes from 10 to 70 (from 5 to 35 for the two items); the degree of satisfaction is directly proportional to the total value obtained.

The *Beck Hopelessness Scale* (*BHS*; Iliceto & Fino, 2015) was used to assess the level of perceived optimism/pessimism for the future expectations. The scores range from 22 to 44. The score of 22 indicates that the individual is not pessimistic, while 44 corresponds to the highest level of pessimism towards the future.

The *Lexington Attachment to Pet Scale* (*LAPS*) was used to assess the attachment of owners to their pets. This scale was developed by Johnson, Garrity and Stallones (1992) to assess people's relationship with their pets by asking participants to indicate the extent to which they agree or disagree with statements such as "Quite often I confide in my pet". The scale was composed by three subscales: general attachment (feelings of attachment to own pets), animal rights and welfare (view the pet's role in the house), and people substituting (how central the pet is to their life).

In the last part of the questionnaire socio-demographic data were collected, including information on the pets (type, possible health issue) and their primary care.

Participants

147 nurses fill-out the questionnaire (response rate 53.5%). Most part of them were female (113, 80.1%). On average, participants aged 42.28 years (*range* = 22-64 years, *SD* = 10.98). Most part of them were married (76, 53.9%), 50 (35.5%) were single, 13 (9.1%) were separated or divorced, 2 (1.4%) were widowed. 78 (55.3%) had one or more children (*range* 1-4). They have been working in the health care sector for 18.31 years on average (*range* = 1-41 years, *SD* = 12.70). In a week, they work on average 39.31 hours (*range* = 25-56 hours, *SD* = 3.53).

Statistical analysis

Statistical analyses were performed using the statistical software SPSS, version 28. Descriptive measures (*Means* ± *SD*) were calculated for all test variables for all groups of participants. t test was used to measure the differences between groups. Differences were considered statistically significant if $p<.05$. Correlations were calculated to examine the relationship between different attachment style, perceived quality of life, life and work satisfaction, pessimism/optimism and coping strategies for pet owners.

RESULTS

79 nurses (53.7%) have a pet, which 61 of them take care in first person (77.2%). In 12 cases (15.2%), the pet needed special care, for example, because it suffered from diseases. The nurse owners of pets aged on average 43.49 years (*SD* = 10.94). Most part of them were married (40, 55.6%), 23 (31.9%) were single, 7 (9.1%) were separated or divorced, 2 (2.8%) were widowed. 45 (62.5%) had one or more children (*range* 1-4). They had on average 19.15 years of working experience (*SD* = 13.02) and working 38.97 hours per week (*SD* = 3.63). Nurse non-owners of pets aged on average 41.01 years (*SD* = 10.96), they had 17.41 years of working experience (*SD* = 12.38) and working 39.68 hours per week (*SD* = 3.42). Most part of them were married (36, 52.2%), 27 (39.1%) were single, 6 (8.6%) were separated or divorced. 33 (47.8%) had one or more children (*range* 1-4) (see Table 1).

About the attachment style, the correlation analysis shown that in pet owner general attachment and welfare and people substituting styles were significantly related with acceptance (respectively, $r = .25$, $p = .035$ and $r = .34$, $p = .003$) and negatively related to religion (respectively, $r = -.23$, $p = .050$ and $r = -.26$, $p = .028$) coping strategies. Animal rights was negatively related with emotional support ($r = -.34$, $p = .004$) coping strategy (see Table 2).

About the attachment style in first person pets' care, the correlation analysis shown that welfare and people substituting style was significantly related with acceptance ($r = .27$, $p = .042$) coping strategy while animal rights were negatively related with emotional support ($r = -.35$, $p = .009$). coping strategy. In not first-person pets' care, general attachment style was negatively related with satisfaction for life ($r = -.53$, $p = .035$). Welfare and people substituting style was significantly related with acceptance ($r = .53$, $p = .037$) and negatively related to instrumental support ($r = -.60$, $p = .014$) coping strategies.

DISCUSSION

The aim of this work was to analyse the relationship between perceived psychological well-being, life satisfaction, optimism, coping strategies and attachment to a pet in nurses. To this end, data from pet owners were compared with those from non-pet owners working in the same healthcare facility. Overall, the results of this study show that participants perceive a good quality of life, are satisfied with their work and life, and are optimistic about the future. It should be noted that the study was conducted before the pandemic, so the results should not be generalised and should be taken with caution.

Table 1 – Perceived quality of life, life and work satisfaction, pessimism/optimism and coping strategies in pet owners and non-owners and attachment style in pet owners (N = 147)

	Pet owner *n* = 79 *M (DS)*	**Not pet-owner** *n* = 68 *M (DS)*	*t*	*p*
PGWBI-A	9.49 (2.78)	9.06 (2.47)	.972	n.s.
Brief COPE:				
Self-distraction	4.46 (1.44)	4.19 (1.37)	1.136	n.s.
Active coping	6.76 (1.29)	6.42 (1.38)	1.528	n.s.
Denial	3.78 (1.45)	3.61 (1.31)	.727	n.s.
Substance use	2.22 (.61)	2.13 (.48)	.988	n.s.
Emotional support	4.40 (1.38)	4.88 (1.53)	−1.963	.026
Instrumental support	5.22 (1.48)	5.52 (1.30)	−1.272	n.s.
Disengagement	3.11 (1.38)	2.92 (1.09)	.875	n.s.
Venting	5.01 (1.57)	4.90 (1.27)	.478	n.s.
Positive reframing	5.99 (1.40)	5.74 (1.48)	1.108	n.s.
Humour	4.26 (1.49)	4.22 (1.30)	.197	n.s.
Acceptance	6.25 (1.28)	6.14 (1.34)	.477	n.s.
Religion	3.86 (1.87)	4.06 (1.88)	−.623	n.s.
Self-blame	5.90 (1.43)	6.03 (1.21)	−.565	n.s.
SWLS Life	23.51 (6.34)	25.14 (6.40)	−1.523	n.s.
SWLF Work	18.06 (4.75)	18.75 (4.17)	−.926	n.s.
Hopelessness	24.29 (3.80)	24.29 (3.80)	1.361	n.s.
Attachment style:				
General attachment	21.83 (6.44)			
Welfare, people sub.	10.58 (5.47)			
Animal rights	10.50 (1.85)			

Table 2 – Perceived quality of life, life and work satisfaction, pessimism/optimism, coping strategies and attachment style in pet owners engaged and not-engaged in first person in the pets' care (N = 79)

	First person pets' care *n* = 59 *M (DS)*	**Not first-person pets' care** *n* = 20 *M (DS)*	*t*	*p*
PGWBI-A	10 (2.80)	7.75 (1.81)	3.034	.002
Brief COPE:				
Self-distraction	4.46 (1.50)	4.37 (1.26)	.197	n.s.
Active coping	6.70 (1.32)	6.94 (1.18)	−.644	n.s.
Denial	3.67 (1.50)	4.06 (1.23)	−.964	n.s.
Substance use	2.28 (.67)	2.00 (.00)	1.655	n.s.
Emotional support	4.40 (1.38)	4.43 (1.36)	−.087	.026
Instrumental support	5.19 (1.40)	5.37 (1.75)	−.433	n.s.
Disengagement	3.09 (1.43)	3.19 (1.17)	−.256	n.s.
Venting	4.95 (1.56)	5.25 (1.57)	−.684	n.s.
Positive reframing	5.98 (1.40)	6.06 (1.44)	−.202	n.s.
Humour	4.35 (1.49)	3.88 (1.45)	1.132	n.s.
Acceptance	6.30 (1.16)	6.06 (1.61)	.655	n.s.
Religion	3.98 (1.84)	3.38 (1.93)	1.156	n.s.
Self-blame	5.84 (1.45)	6.00 (1.41)	−.387	n.s.
SWLS Life	23.44 (6.82)	23.81 (4.08)	−.208	n.s.
SWLF Work	17.98 (4.75)	18.50 (4.77)	−.385	n.s.
Hopelessness	24.35 (3.82)	24.12 (3.72)	.210	n.s.
Attachment style:				
General attachment	21.96 (6.26)	21.13 (7.09)	.461	n.s.
Welfare, people sub.	10.63 (5.46)	10.06 (5.67)	.365	n.s.
Animal rights	10.54 (1.77)	10.19 (2.20)	.673	n.s.

The first hypothesis of our work was that pet owners would have higher well-being and satisfaction, lower perceived hopelessness, and a higher propensity to use adaptive coping strategies than non-pet owners. This hypothesis was not confirmed: we found a nonsignificant difference in perceived quality of life and life satisfaction between pet owners and non-pet owners (see Table 1). These results are consistent with those of Le Roux and Wright (2020), Mháistir (2013), El-Alayli et al. (2006), and Ramirez and Hernandez (2014), who also found a nonsignificant difference in life satisfaction between pet owners and non-pet owners. However, these results differ from those of Bao and Schreer (2016), and Luhmann and Kalitzki (2016), who found that pet owners have significantly higher life satisfaction than non-pet owners. As Le Roux and Wright (2020) noted, the positive impact of leisure activities is not necessarily associated with pet care (Kellert, 1993). Other activities, such as gardening or nature walks, could also have a calming effect. Furthermore, consistent with the buffer hypothesis of social support theory, the benefits of pet ownership could account for the lack of other affections.

Interestingly, we found that non-pet owners tended to use the coping strategy of emotional support more than pet owners. Emotional support is a strategy by which a person seeks empathy and understanding from others. For example, the person receives emotional support by telling someone who offers sympathy when they have received bad news. Consistent with social support theory, pets have been found to meet the social needs of people who care for them by acting as friends and providing unconditional love and acceptance (Hill et al., 2008; Nebbe, 2001). Thus, one possible explanation is that pet owners are less likely to need to rely on others in moments of discouragement.

Among pet owners, attachment data indicated that general attachment was the most frequently reported general attachment style. These data are consistent with previous research, see for example Branson and colleagues (Branson, Boss, Cron & Turner, 2017). In addition, both the general attachment style and the welfare and people substituting styles were significantly related to acceptance and negatively related to religious coping strategies. Acceptance means acknowledging that something is the way it is, that it happened, and that it cannot be changed. Using this coping strategy can help prevent severe pain from escalating into suffering. Religion is a coping strategy that includes aspects of faith and spirituality understood as meaning, purpose, and hope (Damiano, Lucchetti & Peres, 2021) and is associated with a connection to nature, a life force, or god(s). In addition, animal rights were negatively related to emotional support, suggesting that there is less need to share emotions with others and that emotions may be better processed.

Among nurse pet owners, a comparison was made between those who cared for pets in the first person and those who did not care for pets in the first person. The second hypothesis was that those who provide first-person care for the pet will experience lower life satisfaction and high attachment to the surrogate, while those who do not provide first-person care will experience high life satisfaction and overall attachment. Contrary to what we expected, owners who cared for their pet in the first person experienced higher well-being than owners who did not care for their pet in the first person. One possible explanation could be that caring for pets is associated with more happiness and meaning compared to other activities (Kalenkoski & Korankye, 2021). Regarding coping strategies, we expected to find no difference between those who engaged in first-person pet care and those who did not. Contrary to our expectation, the coping strategy of emotional support was more frequently reported by owners who did not engage in first-person care of their pet than by those who engaged in first-person care of their pet. A possible explanation could be the need to feel closer to others, to share emotions, and to receive support to cope with negative feelings at work. Regarding the attachment style, the results showed no difference between the pet owners who cared in the first person and those who did not care in the first person. However, the correlation analysis showed that general attachment style was negatively related to life satisfaction in non-first-person cared pets. In our opinion, this finding is very interesting and could be considered in future research to better understand how pet care might affect life satisfaction and perceived quality of life. The activities associated with caring for a pet are varied, such as cleaning the bedding, playing with the pet, having the pet do activities, taking the pet to the veterinarian. A more detailed examination of the activities that constitute pet care might be useful in understanding the types of activities that - in this working population - might influence perceived well-being. Another finding concerned the style of welfare and people substituting, which was significantly related to acceptance and negatively related to instrumental coping strategies. The use of instrumental support coping strategy is the tendency to rely on others for help in difficult situations. Thus, this could mean that these nurses try to

solve the difficult situation by themselves by accepting the situation as it is.

This study has several limitations. First, this is a cross-sectional study, so the results should not be generalised. The sample consisted of health professionals, a particular occupational group that is at higher risk of stress than other occupational groups (Bennett, Lowe, Matthews, Dourali & Tattersall, 2001; Foster, Roche, Giandinoto, Platania-Phung & Furness, 2021). Second, we did not consider the marital status of animal owners and nonanimal owners. Due to the size of the sample, no in-depth analysis was performed. In further research, marital status could be taken into account, as was intended by Mháistir (2013) and Himsworth and Rock (2013). For example, Himsworth and Rock (2013) found an inverse relationship between pet ownership and life satisfaction, but not among those who were divorced and lived alone. Third, work-related stress was not examined. Future research could examine perceived stress among pet owners and non-pet owners, taking into account the contrasting results of the studies by McConnell and colleagues (McConnell, Brown, Shoda, Stayton & Martin, 2011) and Ramirez and Hernandez (2014), for example. In addition, the study was conducted before the pandemic. As Applebaum and colleagues noted (Applebaum, Tomlinson, Matijczak, McDonald & Zsembik, 2020) pet owners had difficulty meeting the needs of the behavioural pet during the pandemic, which could impact the owners' personal and professional quality (for example, the pet could interrupt work from home). A qualitative study could be used to examine perceived quality of life, job and life satisfaction, and attachment during the pandemic to gain a deeper understanding of the impact of forced home living among pet owners.

To optimize future implementation of programs and proposals for coping strategies to increase perceived quality of life, further research is needed to assess the individual and organizational factors that influence well-being. From an organizational perspective, facilities that promote activities that have a calming effect could have a positive impact on organizational climate. A review of facilities that have been used successfully in various healthcare settings or other organizations could be helpful in planning opportunities for nurses in their workplace. Future work could also explore the feasibility of implementing animal-assisted therapy programs for staff in facilities that do not currently have such programs for patients, the factors that contribute to the positive impact of the program on aspects of staff well-being, and how to optimize the impact.

Despite its limitations, this pilot study provides implications for professional practice and HRM in healthcare organizations. Indeed, the results contribute to a better understanding of how nurses who do or do not own a pet assess their quality of life, their satisfaction with life and work, and their coping with stressful events. The results of this study may be useful in gaining first-hand knowledge about whether or not to get, care for, or keep a pet and the impact this may have on one's personal and professional life.

References

ACQUADRO MARAN, D., CAPITANELLI, I., CORTESE, C.G., ILESANMI, O.S., GIANINO, M.M. & CHIRICO, F. (2022). Animal-assisted intervention and health care workers psychological health: A systematic review of the literature. *Animals*, *12* (3), 383. doi:10.3390/ani12030383

ADRIAN, J.A., DELIRAMICH, A.N. & FRUEH, B.C. (2009). Complicated grief and post-traumatic stress disorder in humans' response to the death of pets/animals. *Bulletin of the Menninger Clinic*, *73*, 176-187. doi:10.1521/bumc.2009.73.3.176

ANDERSON, D.C. (2007). *Assessing the human-animal bond: A compendium of actual measures*. West Lafayette, IN: Purdue University Press.

ANDERSON, K.A., LORD, L.K., HILL, L.N. & McCUNE, S. (2015). Fostering the human-animal bond for older adults: Challenges and opportunities. *Activities, Adaptation & Aging*, *39* (1), 32-42. doi:10.1080/01924788.2015.994447

ANTONACOPOULOS, D. & PYCHYL, T.A. (2010). Examining the role of pet ownership, human social support, pet attachment in psychological health of individuals living alone. *Anthrozoös*, *23*, 37-54. doi:10.2752/175303710X12627079939143

APPLEBAUM, J.W., TOMLINSON, C.A., MATIJCZAK, A., MCDONALD, S.E. & ZSEMBIK, B.A. (2020). The concerns, difficulties, and stressors of caring for pets during COVID-19: Results from a large survey of US pet owners. *Animals*, *10* (10), 1882. doi:10.3390/ani10101882

AYDIN, N., KRUEGER, J.I., FISCHER, J., HAHN, D., KASTENMÜLLER, A., FREY, D. & FISCHER, P. (2012). Man's best friend: How the presence of a dog reduces mental distress after social exclusion. *Journal of Experimental Social Psychology, 48*, 446-449. doi:10.1016/j.jesp.2011.09.011

BAO, K.J. & SCHREER, G.S. (2016). Pets and happiness: Examining the association between pet ownership and wellbeing. *Anthrozoös*, *29* (2), 283-296. doi:10.1080/08927936.2016.1152721

BECK, A.M. (2014). The biology of the human-animal bond. *Animal Frontiers*, *4* (3), 32-36. doi:10.2527/af.2014-0019

BEETZ, A., UVNÄS-MOBERG, K., JULIUS, H. & KOTRSCHAL, K. (2012). Psychosocial and psychophysiological effects of human-animal interactions: The possible role of oxytocin. *Frontiers in Psychology*, *3*, 234. doi:10.3389/fpsyg.2012.00234

BENNETT, P., LOWE, R., MATTHEWS, V., DOURALI, M. & TATTERSALL, A. (2001). Stress in nurses: Coping, managerial support and work demand. *Stress and Health: Journal of the International Society for the Investigation of Stress*, *17* (1), 55-63. doi:10.1002/1532-2998(200101)

BERGET, B., EKEBERG, O. & BRAASTAD, B.O. (2008). Animal-assisted therapy with farm animals for persons with psychiatric disorders: Effects on self-efficacy, coping ability and quality of life, a randomized controlled trial. *Clinical Practice and Epidemiology in Mental Health*, *4*, 4-9. doi:10.1186/1745-0179-4-9

BERGET, B., EKEBERG, O., PEDERSEN, I. & BRAASTAD, B. (2011). Animal-assisted therapy with farm animals for persons with psychiatric disorders: Effects on anxiety and depression. A randomized controlled trial. *Clinical Practice and Epidemiology in Mental Health*, *27*, 50-64. doi:10.1080/0164212X.2011.543641

BJICK, M. (2013). *The effects of a therapy animal on college students stress and arousal*. Master's thesis, St. Catherine University, St. Paul, Minnesota. Retrieved from http://sophia.stkate.edu/msw_papers/152

BLACK, A.F., WINEFIELD, H. & CHUR-HANSEN, A. (2011). Occupational stress in veterinary nurses: Roles of the work environment and own companion animal. *Anthrozoös*, *24* (2), 191-202. doi:10.2752/175303711X12998632257503

BOWLBY, J. (1977). The making and breaking of affectionate bonds. I. Etiology and psychopathology in the light of attachment theory. *The British Journal of Psychiatry*, *130* (3), 201-210. doi:10.1192/bjp.130.3.201

BOWLBY, J. (1982). *Attachment and loss: Attachment*. New York, NY: Basic Books.

BRANSON, S.M., BOSS, L., CRON, S. & TURNER, D.C. (2017). Depression, loneliness, and pet attachment in homebound older adult cat and dog owners. *Journal of Mind and Medical Sciences*, *4* (1), 38-48. doi:10.22543/7674.41.P3848

BROWN, S.G. & RHODES, R.E. (2006). Relationships among dog ownership and leisure-time walking in Western Canadian adults. *American Journal of Preventive Medicine*, *30* (2), 131-136. doi:10.1016/j.amepre.2005.10.007

BUCKLE, C.J. (2015). *Effects of an animal visitation intervention on the depression, loneliness, and quality of life of older people: A randomised controlled study*. Master's thesis, Stellenbosch University, Stellenbosch, South Africa. Retrieved from https://scholar.sun.ac.za

CARVER, C.S. (1997). You want to measure coping but your protocol'too long: Consider the brief cope. *International Journal of Behavioral Medicine*, *4* (1), 92-100. doi: 10.1207/s15327558ijbm0401_6

CHANDLER, M., CUNNINGHAM, S., LUND, E.M., KHANNA, C., NARAMORE, R., PATEL, A. & DAY, M.J. (2017). Obesity and associated comorbidities in people and companion animals: A

one health perspective. *Journal of Comparative Pathology*, *156*, 296-309. doi: 10.1016/j.jcpa.2017.03.006

CHIPAS, A., CORDREY, D., FLOYD, D., GRUBBS, L., MILLER, S. & TYRE, B. (2012). Stress: Perceptions, manifestations, and coping mechanisms of student registered nurse anesthetists. *American Association of Nurse Anesthetists Journal*, *80* (4), 49-55.

CHUR-HANSEN, A., PEACOCK, J. & WINEFIELD, H. (2012). Mental health implications of human attachment to companion animals. *Journal of Clinical Psychology*, *68* (3), 292-303. doi:10.1002/jclp.20866

CHUR-HANSEN, A., WINEFIELD, H. & BECKWITH, M. (2009). Companion animals for elderly women: The importance of attachment. *Qualitative Research in Psychology*, *6* (4), 281-293. doi:10.1080/14780880802314288

COHEN, S., GOTTLIEB, B.H. & UNDERWOOD, L.G. (2000). Social relationships and health. In S. Cohen, L.G. Underwood & B.H. Gottlieb (Eds.), *Social support measurement and intervention: A guide for health and social scientists*. New York, NY: Oxford University Press.

COLE, K.M., GAWLINSKI, A., STEERS, N. & KOTLERMAN, J. (2007). Animal-assisted therapy in patients hospitalized with heart failure. *American Journal of Critical Care*, *16*, 575-585.

COLOMBO, G., BUONO, M.D., SMANIA, K., RAVIOLA, R. & DE LEO, D. (2006). Pet therapy and institutionalized elderly: A study on 144 cognitively unimpaired subjects. *Archives of Gerontology and Geriatrics*, *42*, 207-216. doi:10.1016/j.archger.2005.06.011

CONTI, L. (1999). *Repertorio delle scale psicometriche di valutazione*. Firenze, Italy: SEE Firenze.

CUTT, H.E., GILES-CORTI, B., WOOD, L.J., KNUIMAN, M.W. & BURKE, V. (2008). Barriers and motivators for owners walking their dog: Results from qualitative research. *Health Promotion Journal of Australia*, *19* (2), 118-124. doi:10.1071/he08118

DAMIANO, R.F., LUCCHETTI, G. & PERES, M.F.P. (2021). Defining spirituality in healthcare: A systematic review and conceptual framework. *Frontiers in Psychology*, *12*. doi:10.3389/fpsyg.2021.756080

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

EL-ALAYLI, A., LYSTAD, A.L., WEBB, S.R., HOLLINGSWORTH, S.L. & CIOLLI, J.L. (2006). Reigning cats and dogs: A pet-enhancement bias and its link to pet attachment, pet self-similarity, self-enhancement, and well-being. *Basic and Applied Social Psychology*, *28*, 131-143. doi:10.1207/s15324834basp2802_3

FENG, Z., DIBBEN, C., WITHAM, M.D., DONNAN, P.T., VADIVELOO, T., SNIEHOTTA, F., CROMBIE, I.K. & McMURDO, M.E. (2014). Dog ownership and physical activity in later life: A cross-sectional observational study. *Preventive Medicine*, *66*, 101-106. doi:10.1016/j.ypmed.2014.06.004

FIELD, N.P., ORSINI, L., GAVISH, R. & PACKMAN, W. (2009). Role of attachment in response to pet loss. *Death Studies*, *33*, 334-355. doi:10.1080/07481180802705783.

FINE, A.H. & BECK, A. (2010). Understanding our kinship with animals: Input for health care professionals interested in the human/animal bond. In A.H. Fine (Ed.), *Handbook on animal-assisted therapy: Theoretical foundations and guidelines for practice (3rd ed.)*. London, UK: Elsevier Inc.

FOSTER, K., ROCHE, M., GIANDINOTO, J.A., PLATANIA-PHUNG, C. & FURNESS, T. (2021). Mental health matters: A cross-sectional study of mental health nurses' health-related quality of life and work-related stressors. *International Journal of Mental Health Nursing*, *30* (3), 624-634. doi:10.1111/inm.12823

FRIEDMANN, E. & GEE, N.R. (2019). Critical review of research methods used to consider the impact of human-animal interaction on older adults' health. *The Gerontologist*, *59* (5), 964-972. doi:10.1093/geront/gnx150

FRIEDMANN, E., THOMAS, S.A., COOK, L.K., TSAI, C.C. & PICOT, S.J. (2007). A friendly dog as potential moderator of cardiovascular response to speech in older hypertensives. *Anthrozoös*, *20*, 51-63. doi:10.2752/089279307780216605

GILMOUR, H. (2007). Physically active Canadians. *Health Reports*, *18*, 45-65. Retrieved from http://www.statcan.gc.ca/pub/82-003-x/2006008/article/phys/10307-eng.htm

GUÉGUEN, N. & CICCOTTI, S. (2008). Domestic dogs as facilitators in social interaction: An evaluation of helping and courtship behaviors. *Anthrozoös*, *21*, 339-349. doi:10.1371/journal.pone.0046362

HAM, H. & EPPING, J. (2006). Dog walking and physical activity in the United States. *Preventing Chronic Disease*, *3* (2), 1-7.

HANDLIN, L., HYDBRING-SANDBERG, E., NILSSON, A., EJDEBÄCK, M., JANSSON, A. & UVNÄSMOBERG, K. (2011). Short-term interaction between dogs and their owners: Effects on oxytocin, cortisol, insulin and heart rate - an exploratory study. *Anthrozoös*, *24*, 301-316. doi:10.2752/175303711X13045914865385

HERZOG, H.A. (2007). Gender differences in human-animal Interactions: A review. *Anthrozoös*, *20* (1), 17-21. doi:10.2752/089279307780216687

HERZOG, H.A. (2011). The impact of pets on human health and psychological well-being: fact, fiction, or hypothesis? *Current Directions in Psychological Science*, *20*, 4. doi:10.1177/0963721411415220

HILL, R.P., GAINES, J. & WILSON, R.M. (2008). Consumer behavior, extended-self, and sacred consumption: An alternative perspective from our animal companions. *Journal of Business Research*, *61*, 553-562. doi:10.1016/j.jbusres.2006.11.009

HIMSWORTH, C.G. & ROCK, M. (2013). Pet ownership, other domestic relationships, and satisfaction with life among seniors: Results from a Canadian national survey. *Anthrozoös*, *26* (2), 295-305. doi: 10.2752/175303713X13636846944448

HUNT, M. & PADILLA, Y. (2006). Development of the pet bereavement questionnaire. *Anthrozoös*, *19*, 308-324. doi:10.2752/089279306785415493

ILICETO, P. & FINO, E. (2015). Beck Hopelessness Scale (BHS). *European Journal of Psychological Assessment*, *31* (1), 31-37. doi:10.1027/1015-5759/a000201

JOHNSON, T.P., GARRITY, T.F. & STALLONES, L. (1992). Psychometric evaluation of the Lexington attachment to pets scale (LAPS). *Anthrozoös*, *5* (3), 160-175. doi:10.2752/089279392787011395

KALENKOSKI, C.M. & KORANKYE, T. (2021). Enriching lives: how spending time with pets is related to the experiential well-being of older Americans. *Applied Research in Quality of Life*, *8*, 1-22. doi: 10.1007/s11482-020-09908-0

KAUFMAN, K.R. & KAUFMAN, N.D. (2006). And then the dog died. *Death Studies*, *30*, 61-76. doi:10.1080/07481180500348811

KELLERT, S.R. (1993). Introduction. In S.R. Kellert & E.O. Wilson (Eds.), *The biophilia hypothesis*. Washington, DC: Island Press.

KIKUSUI, T., WINSLOW, J.T. & MORI, Y. (2006). Social buffering: Relief from stress and anxiety. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*, 2215-2228. doi:0.1098/rstb.2006.1941

KNIGHT, S. & EDWARDS, V. (2008). In the company of wolves: The physical, social, and psychological benefits of dog ownership. *Journal of Aging and Health*, *20* (4), 437-455. doi:10.1177/0898264308315875

KRAMER, S.C., FRIEDMANN, E. & BERNSTEIN, P.L. (2009). Comparison of the effect of human interaction, animal-assisted therapy, and AIBO- assisted therapy on long-term care residents with dementia. *Anthrozoös*, *22*, 43-57. doi:10.2752/175303708X390464

KURDEK, L.A. (2009). Pet dogs as attachment figures for adult owners. *Journal of Family Psychology*, *23*, 439-446. doi:10.1037/a0014979.

LEE, H., SHEPLEY, M. & HUANG, C. (2009). Evaluation of off-leash dog parks in Texas and Florida: A study of use patterns, user satisfaction, and perception. *Landscape and Urban Planning*, *92*, 314-324. doi:10.1016/j.landurbplan.2009.05.015

LE ROUX, M.C. & KEMP, R. (2009). Effect of a companion dog on depression and anxiety levels of elderly residents in a long-term care facility. *Psychogeriatrics*, *9* (1), 23-26. doi: 10.1111/j.1479-8301.2009.00268.x

LE ROUX, M.C., SWART, L. & SWART, E. (2014). The effect of an animal-assisted reading program on the reading rate, accuracy and comprehension of grade 3 students: A randomized controlled study. *Child Youth Care Forum*, *43*, 655-673. doi:10.1007/s10566-014-9262-1

LE ROUX, M.C. & WRIGHT, S. (2020). The relationship between pet attachment, life satisfaction, and perceived stress: Results from a South African online survey. *Anthrozoös*, *33* (3), 371-385. doi:10.1080/08927936.2020.1746525

LUBBE, C. & SCHOLTZ, S. (2013). The application of animal-assisted therapy in the South African context: A case study. *South African Journal of Psychology*, *43* (1), 116-129. doi:10.1177/0081246312474405

LUHMANN, M. & KALITZKI, A. (2016). How animals contribute to subjective well-being: A comprehensive model of protective and risk factors. *The Journal of Positive Psychology*, *13* (2), 200-214. doi:10.1080/17439760.2016.1257054

McCONNELL, A.R., BROWN, C.M., SHODA, T.M., STAYTON, L.E. & MARTIN, C.E. (2011). Friends with benefits: On the positive consequences of pet ownership. *Journal of Personality and Social Psychology*, *101* (6), 1239-1252. doi:10.1037/a0024506

McCUNE, S., KRUGER, K.A., GRIFFIN, J.A., ESPOSITO, L., FREUND, L.S., HURLEY, K.J. & BURES, R. (2014). Evolution of research into the mutual benefits of human-animal interaction. Animal *Frontiers*, *4* (3), 49-58. doi:10.2527/af.2014-0022

MHÁISTIR, G. (2013). *Influence of pet ownership on self-esteem, life satisfaction and loneliness among over 65s in Ireland*. Master's thesis, DBS School of Arts, Dublin. Retrieved from http://esource.dbs.ie/bitstream/handle/10788/1607/hdip_mhaistir_g_2013.pdf?sequence=1

MIKULINCER, M. & SHAVER, P.R. (2007). Boosting attachment security to promote mental health, prosocial values, and inter-group tolerance. *Psychological Inquiry*, *18*, 139-156. doi:10.1080/10478400701512646

MILLER, S.C., KENNEDY, C., DEVOE, D., HICKEY, M., NELSON, T. & KOGAN, L. (2009). An examination of changes in oxytocin levels in men and women before and after interaction with a bonded dog. *Anthrozoös*, *22*, 31-42. doi:10.2752/175303708X390455

NAGASAWA, M., KIKUSUI, T., ONAKA, T. & OHTA, M. (2008). Dog's gaze at its owner increases owner's urinary oxytocin during social interaction. *Hormones and Behaviour*, *55*, 434-441. doi:10.1016/j.yhbeh.2008.12.002

NEBBE, L. (2001). The elementary school counsellor and the HCAB. In P. Salloto (Ed.), *Pet assisted therapy: A loving intervention and an emerging profession: Leading to a friendlier, healthier, and more peaceful world*. Norton, MA: D.J. Publications.

NOONAN, E. (2008). People and pets. *Psychodynamic Counselling*, *4* (1), 17-31. doi:10.1080/14753630802492722

ODENDAAL, S.J.S. (2000). Animal-assisted therapy - magic or medicine? *Journal of Psychosomatic Research*, *49*, 275-280. doi:10.1016/s0022-3999(00)00183-5

ODENDAAL, S.J.S. & MEINTJIES, R.A. (2003). Neurophysiological correlates of affiliative behaviour between humans and dogs. *The Veterinary Journal, 165* (3), 296-301. doi:10.1016/S1090-0233(02)00237-X

PEDERSEN, I., MARTINSEN, E.W., BERGET, B. & BRAASTAD, B.O. (2012). Farm animal-assisted intervention for people with clinical depression: A randomized controlled trial. *Anthrozoös*, *25* (2), 149-160. doi:10.2752/175303712X13316289505260

POLHEBER, J. & MATCHOCK, R. (2013). The presence of a dog attenuates cortisol and heart rate in the trier social stress test compared to human friends. *Journal of Behavior Medicine*, *37*, 860-867. doi:10.1007/s10865-013-9546-1

PRATO-PREVIDE, E., FALLANI, G. & VALSECCHI, P. (2006). Gender differences in owners interacting with pet dogs: An observational study. *Ethology*, *112* (1), 64-73. doi:10.1111/j.1439-0310.2006.01123.x

RAMÍREZ, M.T.G. & HERNANDEZ, R. (2014). Benefits of dog ownership: Comparative study of equivalent samples. *Journal of Veterinary Behavior*, *9*, 311-315. doi:10.1016/j.jveb.2014.08.002

RIJKEN, M. & BEEK, S. (2011). About cats and dogs: Reconsidering the relationship between pet ownership and health related outcomes in community-dwelling elderly. *Social Indicators Research*, *102* (3), 373-388. doi:10.1007/s11205-010-9690-8

RISLEY-CURTISS, C. (2010). Social work practitioners and the human-companion bond: A national study. *Social Work*, *55* (1), 38-46. doi:10.1093/sw/55.1.38

SHILOH, S., SOREK, G. & TERKEL, J. (2003). Reduction of state-anxiety by petting animals in a controlled laboratory experiment. *Anxiety Stress Coping*, *16*, 387-395. doi:10.1080/1061580031000091582

SIEGEL, J.M. (1990). Stressful life events and use of physician services among the elderly: The moderating role of pet ownership. *Journal of Personality and Social Psychology*, *58*, 1081. doi:10.1037/0022-3514.58.6.1081

SIEGEL, J.M. (2011). Pet ownership and health. In C. Blazina, G. Boyra & D. Shen-Miller (Eds.), *The psychology of the human-animal bond*. New York, NY: Springer USA.

SMOLKOVIC, I., FAJFAR, M. & MLINARIC, V. (2012). Attachment to pets and interpersonal relationships. *Journal of European Psychology Students*, *3*, 15-23. doi:10.5334/jeps.ao

STALLTER, C. & GUSTIN, T.S. (2021). Evaluating advanced practice nurses' burnout and potential helping modalities. *Journal of Nurse Practitioners*, *17*, 1297-1299. doi:10.1016/j.nurpra.2021.07.003

TAUBE, E., JAKOBSSON, U., MIDLÖV, P. & KRISTENSSON, J. (2016). Being in a bubble: The experience of loneliness among frail older people. *Journal of Advanced Nursing*, *72* (3), 631-640. doi:10.1111/jan.12853

TESTA, S., CIVILOTTI, C., DI FINI, G., ROSSETTO, C., BONCINELLI, V. & VEGLIA, F. (2016). Development of two equivalent short forms of the Psychological General Well-Being Index: PGWBI-A and PGWBI-B. *TPM Testing, Psychometrics, Methodology in Applied Psychology*, *23* (2), 149-166. doi:10.4473/TPM23.2.2

THORPE, R., SIMONSICK, E. & BRACH, J. (2006). Dog ownership, walking behavior, and maintained mobility in late life. *Journal of the American Geriatrics Society*, *5*, 154-168. doi:10.1111/j.1532-5415.2006.00856.x

VILLALTA-GIL, V., ROCA, M., GONZALEZ, N., DOMENEC, E., CUCA, B., ESCANILLA, A. & HARO, J.M. (2009). Dog-assisted therapy in the treatment of chronic schizophrenia inpatients. *Anthrozoös*, *22*, 149-159. doi:10.2752/175303709X434176

WELLS, D.L. (2009). The effects of animals on human health and well-being. *Journal of Social Issues*, *65* (3), 523-543. doi:10.1111/j.1540-4560.2009.01612.x

WINEFIELD, H.R., BLACK, A. & LE ROUX, A. (2008). Health effects of ownership of and attachment to companion animals, in an older population. *International Journal of Behavioral Medicine*, *15* (4), 303-310. doi:10.1080/10705500802365532

WISDOM, J.P., SAEDI, G.A. & GREEN, C.A. (2009). Another breed of service animals: STARS study findings about pet ownership and recovery from serious mental illness. *American Journal of Orthopsychiatry*, *79*, 430-436. doi:10.1037/a0016812.

WOOD, L.J., GILES-CORTI, B., BULSARA, M.K. & BOSCH, D.A. (2007). More than a furry companion: The ripple effect of companion animals on neighbourhood interactions and sense of community. *Society & Animals*, *15* (1), 43-56. doi:10.1163/156853007X169333

DOI: 10.26387/bpa.2022.1

Italian adaptation of the Calling and Vocation Questionnaire - Short Form

Anna Dalla Rosa[1], Michelangelo Vianello[1], Ryan D. Duffy[2]

[1] Department of Philosophy, Sociology, Education and Applied Psychology, University of Padova, Padova, Italy
[2] Department of Psychology, University of Florida, USA

anna.dallarosa@unipd.it

ABSTRACT. Una delle misure di chiamata professionale più utilizzate in letteratura è il *Calling and Vocation Questionnaire (CVQ)*. In questo articolo si presenta la validazione di una versione breve in italiano dello strumento, svolta su un ampio campione di studenti universitari (*N* = 5886). Le analisi confermano la struttura fattoriale originale a sei fattori e supportano l'invarianza della scala nel tempo, fra i generi e fra le diverse discipline di studio. I punteggi presentano buona attendibilità e validità convergente. Il CVQ-I risulta quindi un valido strumento di misura della chiamata professionale, che può essere adottato per la conduzione di ricerche cross-culturali e longitudinali.

SUMMARY. The Calling and Vocation Questionnaire (CVQ) is one of the most widely used multidimensional scales of career calling. We developed a short version of the CVQ in Italian (CVQ-I) and investigated its validity in a large sample of college students (N = 5886). Confirmatory factor analyzes supported a six-factor structure, which was shown to be invariant over time, gender, and study domains using multi-group measurement invariance analyzes. Scale scores were found to possess adequate internal consistency and convergent validity. These results demonstrate validity for the use of the CVQ-I in cross-cultural and longitudinal research and interventions.

Keywords: *Calling, Scale validation, Italian college students*

INTRODUCTION

People may feel called to a particular profession, while others may see work as an instrument to gain power or money. Calling has been defined as an attitude toward work and as a consuming and meaningful passion (Thompson & Bunderson, 2019). According to the most cited definition, calling is a transcendent summons to approach a life role perceived as meaningful and oriented toward helping others (Dik & Duffy, 2009). Dik and Duffy (2009) suggested that people can perceive that they currently have a calling, whereas others may be actively looking for a calling but do not currently have one. Therefore, their conceptualization of calling distinguished between two states: the 'presence of' and the 'search for' a calling. Although agreement on its definition is still missing, research has been growing and studies have provided clear evidence that perceiving work as a calling contributes to improving the well-being of both individuals and organizations (Duffy, Dik, Douglass, England & Velez, 2018). Calling has been mainly studied with samples from the United States, and more intercultural research on calling has recently been advocated (Thompson & Bunderson, 2019).

This paper proposes a step forward in this direction by presenting the Italian validation of a short form of the most widely used multidimensional measure of calling: the *Calling and Vocation Questionnaire* (*CVQ*; Dik, Eldridge, Steger & Duffy, 2012). The CVQ is based on the definition of calling proposed by Dik and Duffy (2009). Although other measures of calling have been proposed, the CVQ possess the largest empirical evidence, distinguishes between having and searching for a calling, and its conceptual model is the basis of an influential theory on calling (WCT; Duffy et al., 2018). This paper also presents the first measurement invariance analysis of the CVQ scores across gender, domain, and time. Establishing measurement invariance allows researchers to interpret comparisons between groups as true differences in calling and not as a measurement artifact. Evidence of longitudinal invariance would support the utility of CVQ in monitoring changes in calling over time and across different stages of individuals' life and career.

The CVQ consists of 24 items divided into six subscales: Presence of - and the Search for - Transcendent summons ($\alpha_{presence}$ = .85, α_{search} = .86; test-rest: $r_{presence}$ = .67, r_{search} = .62), Presence of - and the Search for - Purposeful work ($\alpha_{presence}$ = .88, α_{search} = .88; test-rest: $r_{presence}$ = .63, r_{search} = .60), Presence of - and the Search for - Prosocial orientation ($\alpha_{presence}$ = .88, α_{search} = .92; test-rest: $r_{presence}$ = .66, r_{search} = .67). Dik et al. (2012) presented evidence of good overall reliability ($\alpha_{presence}$ = .89, α_{search} = .90; test-rest: $r_{presence}$ = .75, r_{search} = .67), convergent and discriminant validity. The presence and search scores correlated more strongly with each other than with less conceptually similar criterion variables (career decision self-efficacy, life satisfaction, intrinsic and extrinsic work motivation, work hope, prosocial work motivation, and meaning in life) suggesting convergent and discriminant validity (Dik et al., 2012).

METHODS

Participants and procedure

Data were collected by means of a non-experimental three-wave online survey with a 12-month time lag. The dataset is composed of 5886 Italian college students enrolled in 24 different programs; 1700 participated in the second data collection (T_2) and 881 in the third data collection (T_3). There were 36.2% males (1954 out of 5391) and 63.8% females, with a mean age of 23.37 (*SD* = 5.39). Participants' age at T_1 ranged between 18 and 69. The full list of items and their English translation is provided here: https://osf.io/5hdv3/?view_only=7316855961d240a5820882014e946c54

Measures

– *Calling and Vocation Questionnaire*. To develop a short version of the CVQ, we retained the items that cover the key facets of the construct and showed the highest loadings on their respective factors (Dik et al., 2012, p. 250). The scale examined in this study consists of 18 items (3 items per facet) rated on a scale of 1-4, with 1 being 'not at all true for me' and 4 being 'totally true for me'. The wording of the items was adapted for a sample of college students. We asked them to evaluate their calling towards their major (presence of a calling) and their future professional career (search for a calling). Examples of items include: "I am pursuing this line of study because I believe I have been called to do so" and "I'm searching for my career calling", respectively measuring presence of and search for a transcendent summons.

– *Calling as passion*. The scale developed by Dobrow and Tosti-Kharas (2011) was administered as an alternative measure of calling. This scale measures calling as "a consuming, meaningful passion people experience toward a domain" (Dobrow & Tosti-Kharas, 2011, p. 1005). Items were answered on a 7-point likert scale with 1 being 'strongly disagree' and 7 being 'strongly agree'. The original scale demonstrated high internal consistency (α>.88 across samples) and moderate stability in the short and long term (at 2 months, 3.5 and 7 years). The measure has good convergent and discriminant validity. Exploratory and confirmatory factor analyzes supported a one-dimensional structure that accounts for approximately 50% of the overall variance. The reliability of scale scores for the current study was α = .90. A single factor CFA model provided a good fit to the present data: χ^2 (*df* = 49) = 1972.73, *p*<.001, CFI = .95, RMSEA = .08, SRMR = .04.
– *Job, career, and calling work orientations*. To assess individuals' job, career, and calling work orientations we used three single items developed by Wrzesniewski, McCauley, Rozin and Schwartz (1997). Respondents were presented with three paragraphs describing three different workers: 1) a worker with a job orientation who was interested mainly in monetary compensation and was motivated by extrinsic incentives; 2) a worker with a career orientation who was interested in gaining power and achievement; 3) a worker with a calling orientation who works for the sense of fulfillment that job brings to him/her. Participants were asked to read the paragraphs and then rate the degree to which they identified with each of the three workers (exemplifying job, career, and calling work orientations). The three items were rated on a scale from 1, being 'not at all similar', to 4, being 'totally similar'. Validity of the three single items was established by Wrzesniewski et al. (1997) by correlating the scores at these three single items with the scores obtained at 18 true/false items that were developed to measure the three orientations. For instance, the single-item calling work orientation score correlated on average .41 with six true-false calling items and –.38 with seven true-false job items (Wrzesniewski et al., 1997). In addition, respondents who see their work as a calling ranked their job as relatively more important in comparison to hobbies and friends, and were more satisfied in job and life than respondents who see their work as a job or a career.
– *Single item measure of calling*. We assessed the presence of a calling with a yes-no question ("Do you have a calling?") and the extent to which participants perceive a calling in their life with the single item "How much do you feel a vocation for a specific line of study/work?" answered on a 4-point likert scale with 1 being 'not at all' and 4 being 'extremely'.
– The *Integrated Calling Scale* (*ICS*) developed by Dobrow and Tosti-Kharas (2011) was administered as an alternative measure of calling.

RESULTS

Confirmatory factor analysis

The factor structure of the scale was assessed using three nested CFA models estimated in MPlus 7.0. Fit statistics were evaluated as acceptable on the basis of the following criteria: CFI≥.90; RMSEA≤.08; SRMR≤.10 (Brown, 2015). In the first model, all items are loaded onto a single factor of calling. The second is a two-factor model in which the nine presences of calling items loaded on a global "presence of calling" factor, and the remaining nine items, measuring "search for a calling", loaded on a second latent factor. Finally, a six-factor model was estimated representing the presence of and search for the three dimensions of calling. Fit indexes are reported in Table 1.

Model comparisons suggest that the six-factor solution fits the data better than all other models. Item loadings were all above .53. See Figure 1 for a graphic representation of the final CFA model. The correlations between the presence and search components of calling are similar to those observed in Dik et al. (2012) and are reported in Table 2.

Measurement invariance

Configural (equivalence of model form), metric (equivalence of factor loadings), scalar (equivalence of item intercepts) and strict (equivalence of item residuals) measurement invariance were evaluated across time, gender, and study domain. A change between the more constraint and the less constraint models ≤–.010 in CFI, supplemented by a change ≥.015 in RMSEA or a change ≥.010 in SRMR were used as indicators of noninvariance (Chen, 2007). The results of the nested model comparisons are reported in Table 3.

Table 1 – Fit indexes of alternative first-order factor models

Number of factors	χ^2	*df*	CFI	RMSEA	95% CI	SRMR
1	21958.87	135	.50	.17	[.168; .17]	.13
2	20889.03	134	.53	.17	[.16; .17]	.13
6	3668.74	120	.92	.07	[.07; .08]	.07

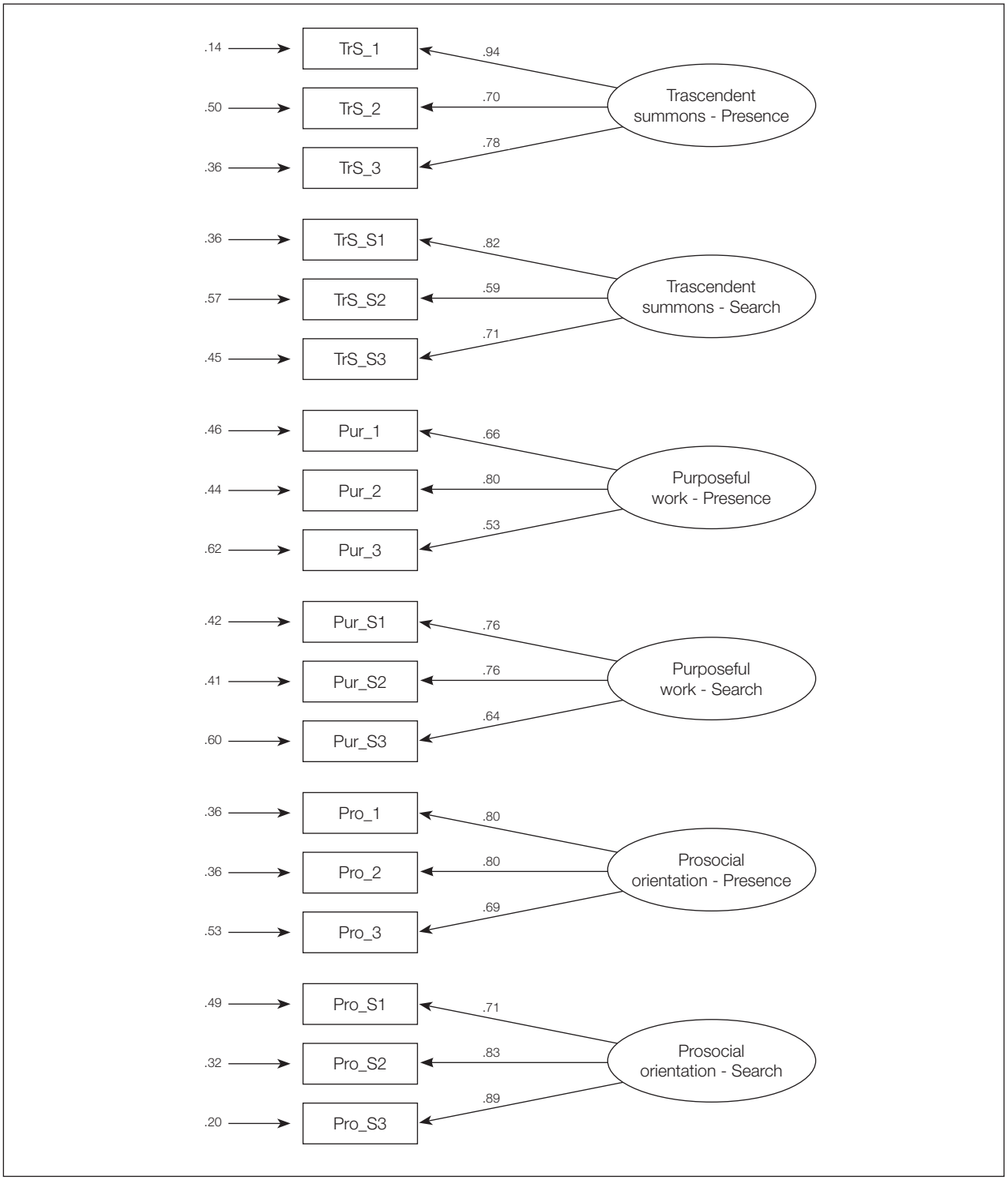
Legenda. df = degree of freedom; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual.
Note. N = 5626.

The fit of the configural invariance model was adequate in T_1 (*N* = 5626), T_2 (*N* = 1699) and T_3 (*N* = 878), which means that the basic organization of the structure (the same loading pattern) is supported on all measurement occasions, $\chi^2_{(974)}$ = 5424.55, CFI = .94, RMSEA = .027, 95% CI [.027, .028], SRMR = .049. The model with constrained loadings (metric invariance) fit the data equally well than the configural model; hence each item contributes to its corresponding latent construct to a similar degree across waves. Models testing the invariance of item intercept (scalar invariance) fit the data equally well as the previous, less parsimonious, metric invariance model. Hence, we can assume that mean differences in the latent construct capture all mean differences in the shared variance of the items. Equity constraints on indicators error variances (strict invariance) produced a slight loss in fit in term of CFI, based on the highest modification index, the error variance of item TrS_Search2 ("I yearn for a sense of calling in my professional career") was found to be non-invariant (larger at Time 1 than at Time 2 and 3). Partial strict invariance was obtained by freely estimating the variance of the item. Hence, the variance of the items that are not shared with the factor and the error variance (measurement error) are similar over time for all items except one.

Using the same procedure, the model was evaluated for invariance between genders. Models estimated in the female (*N* = 3431) and male (*N* = 1947) subsamples at T_1 showed an acceptable fit to the data, which supported configural invariance, $\chi^2_{(240)}$ = 3597.06, CFI = .92, RMSEA = .072, 95% CI [.07, .07], SRMR = .07. Models testing the invariance of loadings, item intercepts, and error variances fit the data equally well as previous, less parsimonious models. Hence, we can assume that they are all invariant between men and women.

Finally, we tested the measurement invariance of the model across study domains composed of more than 300 participants, specifically: Economy (*N* = 475), Engineering (*N* = 702), Education (*N* = 410), Psychology (*N* = 644), Art (*N* = 331), and Medical Science (*N* = 595). The fit of the configural invariance model was acceptable: $\chi^2_{(720)}$ = 2817.43, CFI = .91, RMSEA = .07, 95% CI [.07, .08], SRMR = .07. Models testing the invariance of loadings fit the data equally well as the previous, less parsimonious model. The equality constraints on the indicator intercepts produced a slight loss of fit in terms of CFI. Based on the highest modification index, TrS_Search2 item ("I yearn for a sense of calling in my professional career") intercept was found to be non-invariant and the equality constraint was released. The intercept is slightly higher in the educational (2.71) and art (2.62) domains than in the engineering domain (2.35). Comparison of subgroup Search -Transcendent Summon means might be biased and should be interpreted considering the non-invariance of this item. The model with equality constraints on error variances (strict) fits the data equally well as the partial-scalar model. Loadings, intercept (except for one) and error variances are all invariant across study domains.

Figure 1 – The CFA measurement model



Note. Parameter estimates are standardized. The correlations between factors, reported in Table 2, are omitted for clarity. $\chi^2_{(120)} = 3668.74$, $p<.001$, CFI = .92, RMSEA = .07, 95% CI [.07, .08], SRMR = .07.

Table 2 – Estimates of relations among the six latent factors model from confirmatory factor analyses and internal reliabilities of scale scores

	1	2	3	4	5	6
1 Transcendent summons - Presence	.85, .86					
2 Transcendent summons - Search	.03	.75, .75				
3 Purposeful work - Presence	.38	.07	.71, .71			
4 Purposeful work - Search	.38	.43	.66	.76, .77		
5 Prosocial orientation - Presence	.39	.05	.36	.36	.81, .81	
6 Prosocial orientation - Search	.32	.14	.25	.39	.83	.85, .85

Note. $N = 5626$. All coefficients are statistically different from zero ($p<.001$). Cronbach's alpha and McDonald's omega (in this order) are reported on the main diagonal.

Internal consistency and convergent validity

Table 2 reports the internal consistencies of the six CVQ-I subscales and Table 4 their intercorrelations and correlations with two alternative measures of calling, career, and job orientations. Internal consistencies, evaluated with Cronbach's alpha and McDonald's omega, of the scale scores were higher than .71. The three presence of calling factors had stronger relations with the single item measure of having a calling and the ICS than the three search for a calling factors. The strongest relations were between alternative measures of calling with the presence of transcendent summons and purposeful work.

Calling orientation toward work was positively related to the six CVQ factors (except for the non-significant correlation with search for transcendent summons), whereas job and career orientation were negatively or weakly related to the CVQ scale scores.

To provide further evidence of validity, we compared the means of all dimensions of calling between students who reported having (vs. not) a calling in life. Students with a calling scored higher than students who declared they do not have a calling, in all dimensions of presence and search for a calling (t-tests were significant at $p<.001$, effect size greater than $d = .32$), except for search for transcendent summons, which was higher in students who declared not having a calling ($t = -5.80$, $d = -.22$).

CONCLUSIONS

This is the first study to examine the factor structure of an Italian version of the CVQ. The results of the CFA conducted on a sample of college students identified a six-factor model that was found to have good internal consistency. The high correlation between the search and presence of prosocial orientation (.83) and purposeful work (.66) indicates a potential overlap between the presence of and search for a calling components that might complicate the interpretation of the corresponding test scores. A similar overlap between those calling components was observed in the original validation study (Dik et al., 2012), in which the correlations between the presence of and search for purposeful work

Table 3 – Model fit comparisons for measurement invariance tests by time, gender and study domain

Model	$\Delta\chi^2$	Δdf	ΔCFI	ΔRMSEA	ΔSRMR
Time [a]					
Metric	270.99	20	−.004	−.001	.002
Scalar	303.42	20	−.004	.00	.00
Strict	1290.94	32	−.018	.00	.007
Partial Strict [b]	749.49	30	−.01	<.001	.005
Gender					
Metric	35.78	12	−.001	−.001	.001
Scalar	96.61	12	−.002	−.00	.00
Strict	136.97	18	−.002	−.001	.002
Study domain					
Metric	158.52	60	−.004	−.001	.006
Scalar	321.34	60	−.011	.002	.003
Partial scalar [c]	254.05	55	−.008	.001	.002
Strict	198.14	85	−.004	−.002	.005

Legenda. df = degree of freedom; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual.

Note. Models were compared with the previous and less parsimonious invariance model (Metric vs Configural, Scalar vs Metric, and Strict vs Scalar). All chi-square difference tests are significant at $p<.001$. The means of the latent factors were constrained to zero to achieve identification in the configural and metric invariance models. In the scalar invariance models, latent means were set at 0 at T_1, in the female subsample and in the psychology subsample.

[a] The two items were modified at T_2 and T_3 (TrS_2, Pro_3) and were set as missing in the longitudinal invariance analysis. Presence of transcendent summons and prosocial orientation at T_2 and T_3 were saturated by two items instead of three.

[b] Partial strict invariance was reached, releasing the equality constraint on the error variance of the TrS_Search2 observed variable.

[c] Partial scalar invariance was reached releasing the equality constraint on the intercept of the TrS_Search2 observed variable.

Table 4 – Correlations among CVQ-I scores and measures of calling, job, and career orientations

	Single item	ICS	Calling or.	Job or.	Career or.
Transcendent summons - Presence	.43**	.40**	.23**	–.13**	–.11**
Purposeful work - Presence	.34**	.51**	.33**	–.24**	–.03*
Prosocial orientation - Presence	.24**	.30**	.27**	–.18**	–.12**
Transcendent summons - Search	.01	–.02	–.02	.09**	.08**
Purposeful work - Search	.25**	.31**	.24**	–.14**	–.03*
Prosocial orientation - Search	.19**	.25**	.24**	–.13**	–.08**

Legenda. ICS = *Integrated Calling Scale* developed by Dobrow & Tosti-Kharas (2011).
Note. N = 5886.
** $p<.001$; * $p<.05$

ranged between .71 and .79, and the correlations between the presence of and search for prosocial orientation ranged between.78 and .84. College students are in the middle of their career self-exploration and discernment, and the development of their calling is an ongoing process. Students can have a calling and perceive a prosocial and meaningful purpose in their study domain, but they might also keep seeking for other meaning and value in their academic and professional career. The high correlations between the presence and search for these two components of calling might be a feature of our sample, perhaps generalizable to other people who are at the very beginning of the process of building a career.

In addition to examining the factor structure of the scale and the reliability of the scores, we examined its measurement invariance. The scale was found to be invariant across gender and partially invariant across time and study domain. The measurement and structural invariance of the model across gender and study domain is crucial for the study of calling because it indicates that the measurement model of the latent calling construct, the composite score of the six factors, and the correlations among them can be compared in both sexes and across study domains. These results strengthen the validity of CVQ in settings where the heterogeneity of the population is substantial. To precisely measure the true change and inter-individual differences in career calling, it is critical to examine if the CVQ consistently measures the same construct over time. To our knowledge, this is the first study to examine the longitudinal invariance of CVQ. In our study, we found support for the measurement and structural invariance of the CVQ over time. Therefore, changes in CVQ scores over time reflect true changes in the underlying latent constructs rather than changes in the measurement properties of the scale.

Finally, the scale was found to possess adequate convergent validity supported by positive and moderate to strong relations between CVQ scores with alternative measures of calling. Correlations between CVQ presence subscales and alternative measures of calling were stronger than with CVQ search subscales, which is reasonable given that the alternative measure of calling we adopted is conceptually more similar to the presence of calling rather

than the experience of searching for a calling. In addition, we observed mean differences at the level of the five dimensions of calling between people with internal summons compared to people who experience an external summons.

A limitation of the current study regards the use of a sample of college students. We cannot assume that adult workers approach the construct in the same way as college students do. Future researchers are encouraged to examine the validity of the CVQ-I in samples of Italian adult workers.

The current study suggests that CVQ is a reliable instrument that can be applied in the Italian student population. The multidimensional structure of the scale allows researchers and practitioners to analyze at a fine-grained level the relations between components of calling and other variables. For instance, practitioners can investigate whether individuals experience their calling more as a transcendent summons or as a purposeful work. Composite scores can be compared over time and across gender without specific adjustments.

This research has been supported by a Departmental Grant (VIAN_AFAR18_01) awarded to the second author.
The authors would like to apply for the Open Data Badges. The dataset analyzed during the current study are available in the Open Science Framework repository at the following link: https://osf.io/v56du/?view_only=7316855961d240a5820882014e946c54.

References

BROWN, T.A. (2006). *Confirmatory factor analysis for applied research*. (1st ed.). New York, NY: The Guilford Press.

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: 10.1080/10705510701301834

DIK, B.J. & DUFFY, R.D. (2009). Calling and vocation at work: Definitions and prospects for research and practice. *The Counseling Psychologist*, *37* (3), 424-450. doi: 10.1177/0011000008316430

DIK, B.J., ELDRIDGE, B.M., STEGER, M.F. & DUFFY, R.D. (2012). Development and validation of the Calling and Vocation Questionnaire (CVQ) and Brief Calling Scale (BCS). *Journal of Career Assessment*, *20* (3), 242-263. doi: 10.1177/1069072711434410

DOBROW, S.R. & TOSTI-KHARAS, J. (2011). Calling: The development of a scale measure. *Personnel Psychology*, *64* (4), 1001-1049. doi: 10.1111/j.1744-6570.2011.01234.x

DUFFY, R.D., DIK, B.J., DOUGLASS, R.P., ENGLAND, J.W. & VELEZ, B.L. (2018). Work as a calling: A theoretical model. *Journal of Counseling Psychology*, *65* (4), 423-439.

THOMPSON, J.A. & BUNDERSON, J.S. (2019). Research on work as a calling… and how to make it matter. *Annual Review of Organizational Psychology and Organizational Behavior*, *6*, 421-443. doi: 10.1146/annurev-orgpsych-012218-015140

WRZESNIEWSKI, A., McCAULEY, C., ROZIN, P. & SCHWARTZ, B. (1997). Jobs, careers, and callings: People's relations to their work. *Journal of Research in Personality*, *31* (1), 21-33. doi: 10.1006/jrpe.1997.2162

DOI: 10.26387/bpa.2022.2

Job Satisfaction Scale of Warr, Cook and Wall (1979): The psychometric properties of the Portuguese version

Joana Vieira dos Santos, Cátia Sousa, Gabriela Gonçalves

Psychology Research Centre (CIP), University of Algarve

jcsantos@ualg.pt

ABSTRACT. La *Job Satisfaction Scale* di Warr, Cook e Wall (1979) è ampiamente utilizzata negli studi di psicologia dell'organizzazione ma la struttura fattoriale non è stata sufficientemente esplorata. Il presente studio, composto da 632 partecipanti, si propone di analizzare l'adattamento di questa scala alla lingua portoghese. L'analisi fattoriale confermativa supporta un modello gerarchico a tre fattori di soddisfazione lavorativa, in termini di coefficienti di adeguatezza del modello. Disporre di uno strumento validato in lingua portoghese ne consente un uso molto diffuso.

SUMMARY. The Job Satisfaction Scale of Warr, Cook and Wall (1979) is widely used in studies within the scope of organizational psychology. However, the factor structure was not sufficiently explored (solutions: one factor, two factors, three factors). The present study aims to analyse the adaptation of this scale to the Portuguese language. The sample of the present study consists of 632 participants. Confirmatory factor analysis supports a hierarchical model of three factors of job satisfaction, in terms of adequacy coefficients of the model. The results support the use three factor model. Job satisfaction is widely important concept commonly used by investigators in different scientific areas. It's important to have a very used questionnaire available in Portuguese language, that could be practically used by Portuguese native speakers. A way to empirically have a real notion of job satisfaction levels of the employees is measuring it; having a validated instrument allow its use in Portuguese language.

Keywords: *Professional satisfaction, Validation, Factorial structure, Psychometric*

INTRODUCTION

Professional satisfaction is one of the most studied concepts in the field of organizational behavior since it is one of the most important human results of the work. Some studies pointed out a strong connection between being satisfied at work and performance (Bota, 2013; Gu & Siu, 2009; Wening & Choerudin, 2015). In other ways, there are studies in which no significant relationship was apparent (Mohr & Puck, 2007). Theoretical and practical studies have not guaranteed an unambiguous answer to the nature and strength of the connection between job satisfaction and organizational performance (Bakotic, 2016). Although, organizations with more satisfied employees tended to be more effective than organizations with dissatisfied employees (Ostroff, 1992). So, the real importance of professional satisfaction is reiterated empirically, as well as the relevance of its relation (antecedent or result) to other concepts, namely absenteeism (Schaumberg, & Flynn, 2017), turnover (Lusine, Jianfang, Jingjing & Thomas, 2017; Romeo, Yepes-Baldó & Lins, 2020; Sainju, Hartwell & Edwards, 2021), organizational citizenship behaviors (Lavi & Littman-Ovadia, 2017; Saxena, Tomar & Tomar, 2019), physical and mental health of the worker (burnout, self-esteem, anxiety, depression) (Faragher, Cass & Cooper, 2005; Kim, Ra, Park & Kwon, 2017), organizational commitment (Eliyana, Ma'arif & Muzakki, 2019; Meyer & Allen, 1991; Meyer, Allen & Smith, 1993), organizational culture (Bellou, 2010; Mesfin, Woldie, Adamu & Bekele, 2020), or workaholism (McMillan, Brady, O'Driscoll & Marsh, 2002).

The study of job satisfaction has presented several definitions over time, considering the theoretical framework in which it falls. According to Judge and Klinger (2000), the most popular definition of job satisfaction, and the one that offers a greater comprehensiveness and theoretical coherence, was promoted by Locke (1976), who defined it as a pleasurable or positive emotional state resulting from the appraisal of one's job or job experiences and professional dissatisfaction is the unpleasant emotional state resulting from the evaluation of the work itself as frustrating or blocking the realization of the values of the individual, related to the work. In this way, job satisfaction and dissatisfaction are a function of the perceived relationship between what the individual wants from his work and his perception of what he is receiving.

Locke (1976) points out three phases of the evolution of the concept of satisfaction: the psycho-economic one, in which the professional satisfaction is translated by the salary obtained in function of the work performed; the psycho-sociological, where the professional satisfaction is analyzed taking into account both the personality of the individual and his relation to work as the external variables; the developmentalist, values the work situation taking into account the effort expended by the individual at work to obtain the desired rewards. Thus, job satisfaction is a variable of an affective nature and is a mental process for the evaluation of work experiences, which results in a pleasant or unpleasant state. It can be influenced by the individual's various mental contents, such as beliefs, values, morality, and the possibility of development at work (Levin & Stocks, 1989; Staw & Ross, 1985; Staw, Bell & Clausen, 1986).

Herzberg and colleagues proposed the Two Factor Theory (Herzberg, Mausner & Snyderman, 1959) which is considered one of the first theories proposed to explain job satisfaction. It proposed the existence of two groups of factors, responsible, separately, for job satisfaction - the motivators - such as work itself, performance, promotion and recognition for the work performed: and by job dissatisfaction - hygienic factors - such as supervision, interpersonal relationships at work, and organizational policies. One of the major contributions of Herzberg and colleagues (1959) was to identify the importance of psychological growth as a necessary condition for job satisfaction and to demonstrate that this growth comes from work itself. For Hackman and Oldham (1974), satisfaction would be determined by critical psychological states, caused by five central characteristics of the work: variety, identity, meaning, task autonomy and feedback received. This ratio would, according to them, be moderated by the need for employee growth. However, for Staw and Ross (1985) the concepts of social information processing, job characteristics can be subjective and perceptually flexible, and attitudes at work can be determined by both situational factors and relatively stable dispositional aspects of the individual through the time.

Job satisfaction is one of the most used constructs in industrial and organizational psychology (Heritage, Pollock & Roberts, 2015), for their relationships with individual and organizational dimensions in the workplace, as well as with working conditions (Dormann & Zapf, 2001). It can be considered as the passive acceptance of a satisfactory situation (Warr & Inceoglu, 2012), and can be based on intrinsic (e.g., affective bond) or extrinsic (e.g., payment) aspects (Warr, Cook & Wall, 1979). A meta-analysis developed by Faragher

et al. (2005) concluded that depression, anxiety, burnout, and self-esteem were all associated with job satisfaction. Thus, the conceptual and operational integrity of this concept has relevance both to the academic context and to applied valence.

Job Satisfaction Scale

Warr et al. (1979) developed a scale of 15 items indicative of intrinsic/extrinsic factors related to job satisfaction. The scale also includes a 16th item, which evaluates the overall satisfaction of the participant. In 2015, Heritage and colleagues reported that, up to that time, the study by Warr et al. had more than 650 citations, according to the records of the web of knowledge, and that this scale continued to be frequently used in the academic context, but also applied. The original two- and three-factor structures provided by Warr et al. (1979) were based on cluster analysis: an exploratory technique. These results are still quoted in contemporary literature (e.g., Falkum & Vaglum, 2005; Mohd Dahlan, Mearns & Flin, 2010; Ose et al., 2010; Solberg et al., 2012; Turner, Ross & Ibbetson, 2011). More often, reference is made to the two-factor solution (intrinsic/extrinsic). Magnavita, Fileni and Bergamaschi (2009) demonstrated a two-factor solution in an Italian translation scale study, although the solution did not accurately reflect the same items as those presented by Warr et al. (1979).

In addition to the studies of two and three factors of previous research, there are still studies that choose an one-dimensional solution. As an illustration, with the exploratory factorial analysis, Morrison (2004) found a solution of a single factor that best represented the structure of the underlying factor of the scale. Hills, Joyce and Humphreys (2012) observed a single factor solution through exploratory techniques.

It should be noted that all the investigations studied used exploratory and non-factorial confirmatory analyzes. There is limited consensus in the research literature on the factor structure of this scale. By 2015, Heritage and colleagues pointed out that to date the use of confirmatory factor analysis (CFA) has not been used to validate the factorial structure of the measure. These authors present a hierarchical model of three factors, in which there is a latent factor that represents the general professional satisfaction and that has high correlations with the three factors.

In view of the importance of the construct and due to the happy-productive worker hypothesis has most often been examined in organizational research by correlating job satisfaction to performance (Wright & Cropanzano, 2000) is our main goal of this study to contribute to the adaptation of a widely used measure to the Portuguese language. Despite the widely used of this measure, only with few studies of analysis to its psychometric characteristics, namely using structural equation models, the present study seeks to respond to this need. Thus, it is our objective to contribute to the adaptation of the scale of Warr and colleagues (1979) to the Portuguese population, as well as to explore its factorial structure.

METHOD

Sample

A total of 632 individuals participated in this study, 252 men and 379 women, aged between 18 and 68 years. Most of the participants are married (n = 369, 58.4%), or single (n = 258, 40.8%). The educational qualifications range from the 1st cycle of basic education (n = 7, 1.1%), the 2nd and 3rd cycles of basic education (n = 60, 9.5%); (n = 201, 31.8%) and higher (n = 361, 57.15%). The employment situation is mainly employed (n = 550, 87.0%).

Questionnaire

Professional satisfaction was evaluated using the *Job Satisfaction Scale* of Warr et al. (1979). This is constituted by fifteen items, all of them in the positive sense, where they refer to various aspects of the work, where the participants indicate how satisfied or dissatisfied, they are in relation to each of the presented work characteristics. There is still a sixteenth item that refers to overall job satisfaction. The response scale is 7 points (1 = extremely unsatisfied to 7 = extremely satisfied). The internal consistency of the scale in the original study is .89.

The scale translation process, to maintain the equivalence between the original measure in English and the version translated into Portuguese, followed the method proposed by Brislin (1970): 1) back-translation method; 2) bilingual technique; 3) committee approach; and 4) pre-test procedure. For the pre-test, the translated version into Portuguese was applied to 15 participants, and the value of internal

consistency was acceptable. These participants were not included in the final sample.

A questionnaire was also applied with sociodemographic questions regarding the age, gender, marital status, literacy, employment status and type of employment contract.

Data collection procedure

Participants completed the questionnaire, which took about 10 minutes to complete. The paper and pencil questionnaires were completed in the presence of the investigator, in companies, universities, and public places. Participants were guaranteed the anonymity of data, as well as voluntary and free participation.

Data analysis procedure

To analyze the data, the SPSS (version 20.0) and SPSS Amos (version 20.0) programs were used. To carry out the adaptation and validation of the scale for the Portuguese population, descriptive statistics of the items, confirmatory factor analysis and internal consistency analysis were performed.

The following indicators were calculated based on the recommendations of Byrne (2010): The χ^2 (chi-square); χ^2/df (which must range between 2 and 5); CFI (which can vary between 0 and 1); RMSEA (values between .05 and .08 indicate a good fit); and SRMR (an appropriate adjustment of the model is indicated by values lower than .05; Hu & Bentler, 1999). To analyze the internal consistency, Cronbach's alpha (α) and coefficient omega (ω) were performed.

The maximum likelihood estimation method was used, which assumes the multivariate normal distribution and is robust when this assumption is not attended (Schermelleh-Engel, Moosbrugger & Müller, 2003), which is the case for the data in this study (Kolmogorov-Smirnov $\leq$.01).

RESULTS

Analysis of items

As can be seen in Table 1, the data have an asymmetric distribution and do not meet the assumption of normality according to the results of the Kolmogorov-Smirnov test.

Confirmatory factor analysis

In the accomplishment of the confirmatory factorial analysis the several solutions found in previous studies were tested: (i) one-factor (16 items); (ii) one-factor (15 items); (iii) bifactorial (extrinsic satisfaction and intrinsic satisfaction) (15 items); (iv) trifactorial (extrinsic satisfaction, intrinsic satisfaction, and satisfaction with professional relations - subordinated to a general factor: general professional satisfaction) (see Table 2).

The model of the three factors, ensured a better adjustment (see Figure 1).

The observed χ^2/df of 6.127 (p = .000) approached the desired values. The CFI of .914 is a benchmark for good adjustment (Byrne, 2010; Joreskog, 1966). Regarding the SRMR and RMSEA should have values lower than .05 to be considered a good fit, although values close to .08 are considered a reasonable adjustment (MacCallum, Browne & Sugawara, 1996). Thus, the SRMR value of .047 and the RMSEA of .089 are indicative of a reasonable adjustment.

Analysis of internal consistency

The scale reliability analysis was performed using the Cronbach's alpha and omega coefficient. The value of reliability with all items was very good (α = .933; ω = .934). The elimination of item 15 maintain the value.

The following internal consistency values were observed: intrinsic satisfaction - α = .866, ω = .868; extrinsic satisfaction - α = .707, ω = .710; satisfaction with professional relationships - α = .877, ω = .878. In any of the dimensions no item should be eliminated, otherwise the internal consistency will be reduced.

DISCUSSION, IMPLICATIONS AND FUTURE DIRECTIONS

The present study had as main objective to analyze the factorial structure of the scale of Warr et al. (1979), in a sample of Portuguese population. Based on the theoretical description of Warr et al. (1979), it would be expected to observe a factorial structure that expresses two factors (intrinsic satisfaction and extrinsic satisfaction) or a unifatorial solution. A three-factor hierarchical model presented the best model adjustment indices.

Table 1 – Location, normality, and asymmetry of the data

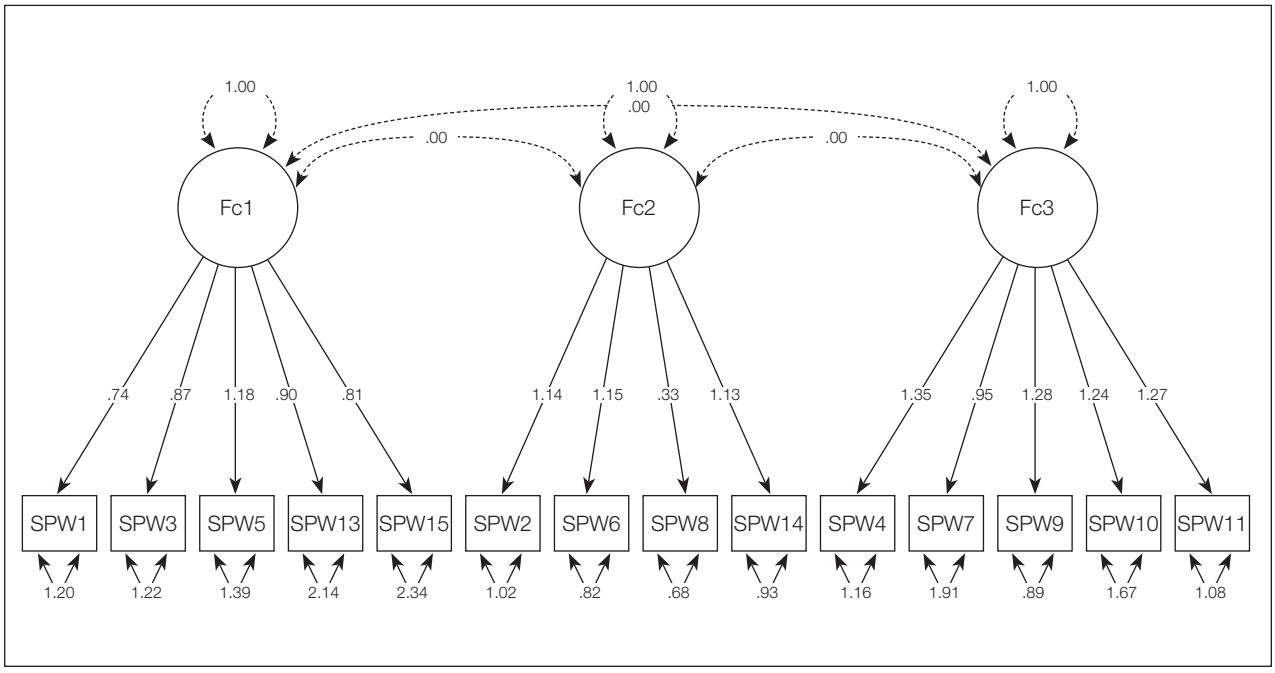
	Percentiles							Central Tendency		Normality	Asymmetry
	5	10	25	50	75	90	95	*Mean*	*SD*	*Kolmogorov-Smirnov*	
Item 1	2.00	3.00	4.00	5.00	6.00	6.00	7.00	4.74	1.323	KS = .191, *p* = .000	−.453
Item 2	2.00	3.00	4.00	5.00	6.00	7.00	7.00	4.88	1.523	KS = .175, *p* = .000	−.531
Item 3	3.00	3.00	4.00	5.00	6.00	7.00	7.00	5.22	1.408	KS = .189, *p* = .000	−.595
Item 4	1.00	2.00	3.00	4.00	6.00	6.00	7.00	4.24	1.728	KS = .141, *p* = .000	−.236
Item 5	1.65	2.00	4.00	5.00	6.00	7.00	7.00	4.81	1.668	KS = .157, *p* = .000	−.538
Item 6	2.00	3.00	4.00	5.00	6.00	7.00	7.00	4.96	1.465	KS = .158, *p* = .000	−.547
Item 7	1.00	1.00	2.00	3.00	5.00	5.00	6.00	3.29	1.679	KS = .135, *p* = .000	.260
Item 8	2.00	2.00	4.00	5.00	6.00	6.00	7.00	4.50	1.564	KS = .155, *p* = .000	−.361
Item 9	1.00	2.00	4.00	5.00	6.00	7.00	7.00	4.56	1.589	KS = .171, *p* = .000	−.427
Item 10	1.00	1.00	2.00	3.00	4.00	6.00	7.00	3.20	1.790	KS = .145, *p* = .000	.388
Item 11	1.00	2.00	3.00	4.00	5.00	6.00	7.00	4.04	1.645	KS = .149, *p* = .000	−.159
Item 12	1.00	2.00	3.00	4.00	5.00	6.00	7.00	4.29	1.557	KS = .154, *p* = .000	−.282
Item 13	1.00	2.00	3.00	5.00	6.00	7.00	7.00	4.52	1.720	KS = .162, *p* = .000	−.408
Item 14	2.00	3.00	4.00	5.00	6.00	7.00	7.00	4.69	1.489	KS = .159, *p* = .000	−.421
Item 15	1.00	2.00	4.00	5.00	6.00	7.00	7.00	4.58	1.733	KS = .150, *p* = .000	−.502
Item 16	2.00	3.00	4.00	5.00	6.00	6.00	7.00	4.68	1.421	KS = .174, *p* = .000	−.489

Table 2 – Confirmatory factorial structures tested

	χ^2/df	CFI	TLI	RMSEA	SRMR
1. One-factor (16 items)	7.323	.889	.872	.100	.052
2. One-factor (15 items)	7.655	.884	.865	.103	.054
3. Bifactorial(15 items)	7.692	.885	.864	.103	.054
4. Trifactorial (15 items)	**6.127**	**.914**	**.896**	**.089**	**.047**

Legenda. df = degree of freedom; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Squared Residual.

Figure 1 – Factorial confirmatory structure



The original article by Warr et al. (1979) have correlations between the subscale scores (r = .58 - .72), but not between latent factors. When strongly correlated latent factors of the scale are considered, then one-dimensionality can be guaranteed for practical issues. This further validates some exploratory results from previous studies (e.g., Hills et al., 2012; Morrison, 2004) reiterating a one-dimensional solution was the best representative of the scale. Our results are consistent to those observed by Heritage et al. (2015), in a study developed with an Australian active population sample. Thus, by the similarity of results, we believe that the factorial solution of three factors, subordinated to a general factor of job satisfaction, allows its use, global or by dimensions, in the Portuguese population.

In terms of internal consistency, it should be noted that the observed results are very satisfactory, above those of the original study and some adaptations (e.g., Heritage et al., 2015). In the future we consider that the sample can be expanded achieving people who have managerial jobs and people who don´t. Miao, Humphrey and Qian (2017) find that job satisfaction is higher in non-managerial jobs than in managerial jobs.

This study presents some limitations. For future research could be interesting to analyze the temporal stability. As previously discussed by Molina and colleagues (Molina, Moliner, Martínez-Tur, Cropanzano & Peiró, 2016) for the assessment of justice climate, we believed that in case of job satisfaction, a longitudinal assessment could have allowed to test possible variability over the time. So, although this study contributes to a better understanding of the *Job Satisfaction Scale*, namely through the performance of factorial validity and assessment of different structures, one of the limitations was to apply only one instrument, which did not allow us to observe, for example, the temporal stability of the measure.

This study has also some methodological limitations, for example sample doesn't guarantee a normal distribution and the invariance wasn't tested.

To have more evidence of psychometric properties of this questionnaire it would be interesting to develop studies considering other validities (e.g., convergent; divergent; content), as also suggested by van Beveren et al. (2017) in a study of psychometric properties of *Global Transformational Leadership Scale*.

Funding. This work was funded by national funds through FCT - Fundação para a Ciência e a Tecnologia - as part the project CIP/UAL - Refª UID/PSI/04345/2020

References

BAKOTIC, D. (2016). Relationship between job satisfaction and organisational performance. *Economic Research-Ekonomska Istraživanja*, *29* (1), 118-130. doi.org/10.1080/1331677X.2016.1163946

BELLOU, V. (2010). Organizational culture as a predictor of job satisfaction: The role of gender and age. *Career Development International*, *15*, 4-19. doi.org/10.1108/13620431011020862

BOTA, O.A. (2013). Job satisfaction of teachers. *Procedia - Social and Behavioral Sciences*, *83*, 634-638. doi.org/10.1016/j.sbspro.2013.06.120

BRISLIN, R.W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, *1* (3), 185-216. doi.org/10.1177/135910457000100301

BYRNE, B.M. (2010). *Structural equation modeling with AMOS. Basic concepts, applications, and programming (2nd ed.)*. Routledge Taylor & Francis Group.

DORMANN, C. & ZAPF, D. (2001). Job satisfaction: A meta-analysis of stabilities. *Journal of Organizational Behavior*, *22*, 483-504. doi.org/10.1002/job.98

ELIYANA, A., MA'ARIF, S. & MUZAKKI, M. (2019). Job satisfaction and organizational commitment effect in the transformational leadership towards employee performance. *European Research on Management and Business Economics*, *25* (3), 144-150. doi.org/10.1016/j.iedeen.2019.05.001

FALKUM, E. & VAGLUM, P. (2005). The relationship between interpersonal problems and occupational stress in physicians. *General Hospital Psychiatry*, *27*, 285-291. doi.org/10.1016/j.genhosppsych.2005.03.006

FARAGHER, E.B., CASS, M. & COOPER, C.L. (2005). The relationship between job satisfaction and health: A meta-analysis. *Occupational and Environmental Medicine*, *62*, 105-112. doi.org/10.1136/oem.2002.006734

GU, Z. & SIU, R.C. (2009). Drivers of job satisfaction as related to work performance in Macao casino hotels: An investigation based on employee survey. *International Journal of Contemporary Hospitality Management*, *21* (5), 561-578. doi.org/10.1108/09596110910967809

HACKMAN, J.R. & OLDHAM, G.R. (1974). *Motivation through the design of work: Test of a theory*. Yale University.

HERITAGE, B., POLLOCK, C. & ROBERTS, L. (2015). Confirmatory analysis of Warr, Cook, and Wall's (1979) Job Satisfaction Scale. *Australian Psychologist*, *50*, 122-129. doi.org/10.1111/ap.12103

HERZBERG, F., MAUSNER, B. & SNYDERMAN, B. (1959). *The motivation to work*. Wiley.

HILLS, D., JOYCE, C. & HUMPHREYS, J. (2012). Validation of a job satisfaction scale in the Australian clinical medical workforce. *Evaluations & the Health Professions*, *35*, 47-76. doi.org/10.1177/0163278710397339

HU, L. & BENTLER, P. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1-55. doi.org/10.1080/10705519909540118

JORESKOG, K.G. (1966). Testing a simple structure hypothesis in factor analysis. *Psychometrika*, *31*, 165-178.

JUDGE, T.A. & KLINGER, R. (2000). Promote job satisfaction through mental challenge. *Handbook of principles of organizational behavior*, *3* (1), 75-89.

KIM, W., RA, Y.-A., PARK, J. & KWON, B. (2017). Role of burnout on job level, job satisfaction, and task performance. *Leadership & Organization Development Journal 38* (5), 630-645. doi.org/10.1108/LODJ-11-2015-0249

LAVY, S. & LITTMAN-OVADIA, H. (2017). My better self: Using strengths at work and work productivity, organizational citizenship behavior, and satisfaction. *Journal of Career Development*, *44*, 95-109. doi.org/10.1177/0894845316634056

LEVIN, I. & STOCKS, J.P. (1989). Dispositional approach to the job satisfaction: Role of negative affectivity. *Journal of Applied Psychology*, *74*, 752-758. doi.org/10.1037/0021-9010.74.5.752

LOCKE, E.A. (1976). The nature and causes of job satisfaction. In M.D. Dunnette (Ed.), *Handbook of industrial and organizational psychology*. Rand McNally College Publishing Company.

LUSINE, P., JIANFANG, L., JINGJING, S. & THOMAS, D. (2017). Practice environments and job satisfaction and turnover intentions of nurse practitioners: Implications for primary care workforce capacity. *Health Care Management Review*, *42*, 162-171. doi.org/10.1097/HMR.0000000000000094

MacCALLUM, R.C., BROWNE, M.W. & SUGAWARA, H.M. (1996). Power analysis and determination of size sample for covariance structure modeling. *Psychology Methods*, *1*, 130-149.

MAGNAVITA, N., FILENI, A. & BERGAMASCHI, A. (2009). Satisfaction at work among radiologists. *La Radiologia Medica*, *114*, 1330-1344. doi.org/10.1007/s11547-009-0461-z

McMILLAN, L.H.W., BRADY, E.C., O'DRISCOLL, M.P. & MARSH, N.V. (2002). A multifaceted validation study of Spence and Robbins' (1992) workaholism battery. *Journal of Occupational and Organizational Psychology*, *75*, 357-368. doi.org/10.1348/096317902320369758

MESFIN, D., WOLDIE, M., ADAMU, A. & BEKELE, F. (2020). Perceived organizational culture and its relationship with job satisfaction in primary hospitals of Jimma zone and Jimma town administration, correlational study. *BMC Health Services Research*, *20* (438), 1-9. doi.org/10.1186/s12913-020-05319-x

MEYER, J.P. & ALLEN, N.J. (1991). A three-component conceptualization of organizational commitment. *Human Resource Management Review*, *1*, 61-98. doi.org/10.1016/1053-4822(91)90011-z

MEYER, J.P., ALLEN, N.J. & SMITH, C.A. (1993). Commitment to organizations and occupations: Extension and test of a three-component conceptualization. *Journal of Applied Psychology*, *78*, 538-551. doi.org/10.1037/0021-9010.78.4.538

MIAO, C., HUMPHREY, R. & QIAN, S. (2017). A meta-analysis of emotional intelligence effects on job satisfaction mediated by job resources, and test of moderators. *Personality and Individual Differences*, *116*, 281-288. doi.org/10.1016/j.paid.2017.04.031

MOHD DAHLAN, A.M., MEARNS, K. & FLIN, R. (2010). Stress and psychological well-being in UK and Malaysian fire fighters. *Cross Cultural Management*, *17*, 50-61. doi.org/10.1108/13527601011016907

MOHR, A.T. & PUCK, J.F. (2007). Role conflict, general manager job satisfaction and stress and the performance IJVs. *European Management Journal*, *25* (1), 25-35. doi.org/10.1016/j.emj.2006.11.003

MOLINA, A., MOLINER, C., MARTÍNEZ-TUR, V., CROPANZANO, R. & PEIRÓ (2016). Validating justice climate and peer justice in a real work setting. *Journal of Work and Organizational Psychology*, *32*, 191-205. doi.org/10.1016/j.rpto.2016.09.002

MORRISON, R. (2004). Informal relationships in the workplace: Associations with job satisfaction, organisational commitment and turnover intentions. *New Zealand Journal of Psychology*, *33*, 114-128.

NUNNALLY, J. (1978). *Psychometric theory (2nd ed.)*. McGraw-Hill.

OSE, D., FREUND, T., KUNZ, C.U., SZECSENYI, J., NATANZON, I., TRIESCHMANN, J. & MIKSCH, A. (2010). Measuring organizational attributes in primary care: A validation study in Germany. *Journal of Evaluation in Clinical Practice*, *16*, 1289-1294. doi.org/10.1111/j.1365-2753.2009.01330.x

OSTROFF, C. (1992). The relationship between satisfaction, attitudes, and performance: An organizational level analysis. *Journal of Applied Psychology*, *77* (6), 963-974. doi.org/10.1037/0021-9010.77.6.963

ROMEO, M., YEPES-BALDÓ, M. & LINS, C. (2020). Job satisfaction and turnover intention among people with disabilities working in special employment centers: The moderation effect of organizational commitment. *Frontiers in Psychology*, *11*, 1035. doi.org/10.3389/fpsyg.2020.01035

SAINJU, B., HARTWELL, C. & EDWARDS, J. (2021). Job satisfaction and employee turnover determinants in Fortune 50 companies: Insights from employee reviews from Indeed.com. *Decision Support Systems*, 113582. doi.org/10.1016/j.dss.2021.113582

SAXENA, S., TOMAR, K. & TOMAR, S. (2019). Impact of job satisfaction on organizational citizenship behavior. *Proceedings of 10th International Conference on Digital Strategies for Organizational Success*. doi.org/10.2139/ssrn.3323753

SCHAUMBERG, R.L. & FLYNN, F.J. (2017). Clarifying the link

between job satisfaction and absenteeism: The role of guilt proneness. *Journal of Applied Psychology*, *102*, 982-992. doi.org/10.1037/apl0000208

SCHERMELLEH-ENGEL, K., MOOSBRUGGER, H. & MÜLLER, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research - Online*, *8*, 23-74.

SOLBERG, I., AASLAND, O., GUDE, T., MOUM, T., VAGLUM, P. & TYSSEN, R. (2012). The impact of change in a doctor's job position: A five-year cohort study of job satisfaction among Norwegian doctors. *BMC Health Services Research*, *12*, 1-7. doi.org/10.1186/1472-6963-12-41

STAW, B.M., BELL, N.E. & CLAUSEN, J.A. (1986). The dispositional approach to job attitudes: A Lifetime Longitudinal Test. *Administrative Science Quarterly*, *31*, 56-77. doi: 10.2307/2392766

STAW, B.M. & ROSS, J. (1985). Stability in the midst of change: A dispositional approach to job attitudes. *Journal of Applied Psychology*, *70*, 469-480. doi.org/10.1037/0021-9010.70.3.469

TURNER, S., ROSS, M.K. & IBBETSON, R.J. (2011). Job satisfaction among dually qualified hygienist-therapists in UK primary care: a structural model. *British Dental Journal*, *210*, 166-167. doi.org/10.1038/sj.bdj.2011.50

VAN BEVEREN, P., DIMAS, I., LOURENÇO, P. & REBELO, T. (2017). Psychometric properties of the Portuguese version of the Global Transformational Leadership (GTL) scale. *Journal of Work and Organizational Psychology*, *33*, 109-114. doi.org/10.1016/j.rpto.2017.02.004

WARR, P., COOK, J. & WALL, T. (1979). Scales for measurement of some work attitudes and aspects of psychological well-being. *Journal of Occupational Psychology*, *52*, 129-148. doi.org/10.1111/j.2044-8325.1979.tb00448.x

WARR, P. & INCEOGLU, I. (2012). Job engagement, job satisfaction, and contrasting associations with person-job fit. *Journal of Occupational Health Psychology*, *17*, 129-138. doi.org/10.1037/a0026859

WENING, N. & CHOERUDIN, A. (2015). The influence of religiosity towards organizational commitmen, job satisfaction, and personal performance. *Polish Journal of Management Studies*, *11* (2), 181-191.

WRIGHT, T.A. & CROPANZANO, R. (2000). Psychological well-being and job satisfaction as predictors of job performance. *Journal of Occupational Health Psychology*, *5* (1), 84-94. doi.org/10.1037/1076-8998.5.1.84