vol. LXXXI January-April 2023
N° 296

Four-monthly Journal
ISSN 0006-6761

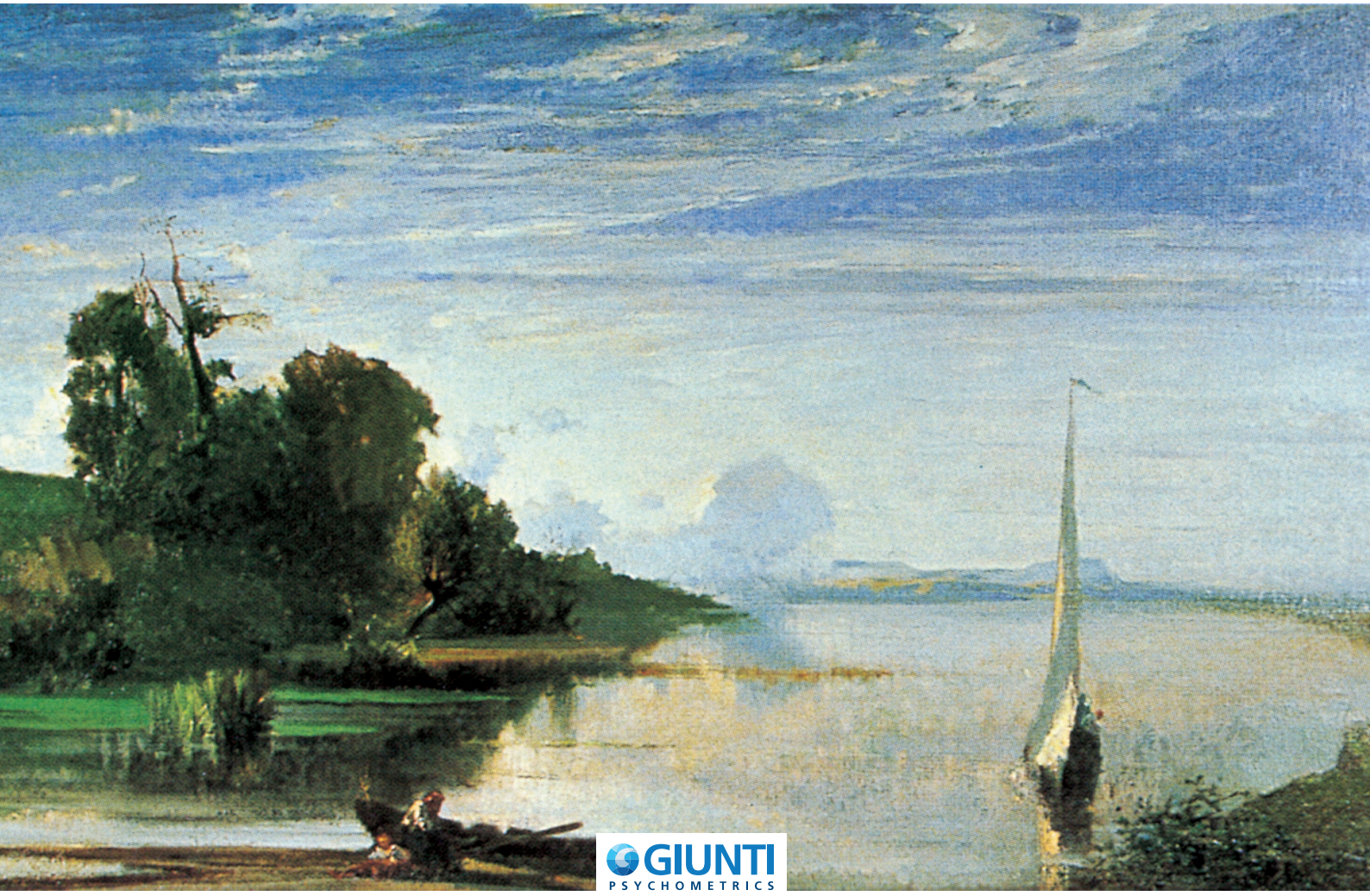
BPA

1/3

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

Research

Experiences & Tools



GIUNTI PSYCHOMETRICS

Scientific Director Alessandro Zennaro

CONTENTS

⬢ Research

▲ Experiences & Tools

DOI: 10.26387/bpa.2023.00004

Exploratory Structural Equation Modeling (ESEM): Methodological considerations and empirical results using the Orientation for Teaching Survey (OTS)

Palmira Faraci, Rossella Bottaro

Faculty of Human and Social Sciences - University of Enna "Kore", Italy

rossella.bottaro@unikorestudent.it; rossella.bottaro1@gmail.com

ABSTRACT. Il presente studio si è posto l'obiettivo di fornire un quadro di riflessione metodologica sull'utilizzo della tecnica Exploratory Structural Equation Modeling (ESEM), presentando un esempio applicativo per mezzo dell'Orientation for Teaching Survey (OTS). 338 insegnanti italiani hanno partecipato ad una survey online per la valutazione della motivazione all'insegnamento e del benessere psicologico. Modelli confermativi ed ESEM sono stati comparati secondo un approccio olistico, che ha mostrato il modello bifattoriale ESEM quale scelta appropriata (χ^2 = 178.02, *df* = 102, CFI = .94, TLI = .92, RMSEA = .05, SRMR = .03; ω = .88) per la versione breve dell'OTS. I risultati della network analysis hanno, altresì, messo in luce le differenze tra insegnanti precari e di ruolo. Gli ESEM sembrano offrire soluzioni analitiche promettenti per le misure multidimensionali complesse in generale e potrebbero fornire una prospettiva innovativa per la valutazione della motivazione all'insegnamento in particolare.

SUMMARY. Exploratory Structural Equation Modeling (ESEM) has emerged as an intriguing approach for investigating the dimensionality of psychometric tools. The present study aimed to reflect on the advantages and drawbacks of the ESEM technique using the Orientation for Teaching Survey (OTS) as an application example. 338 Italian teachers, mostly women (77.5%; M_{age} = 46.84, SD = 10.65) completed an online survey that evaluated the motivation for teaching and indicators of psychological well-being. Confirmatory and ESEM models were compared to test the OTS factor solution. A series of network analysis were performed to investigate the relationship between motivation for teaching and depression, anxiety, stress, optimism, intolerance of uncertainty, and self-esteem in the total sample as well as the precarious and in-role groups. Our study revealed the superiority of the bi-factor ESEM model for the OTS short version (χ^2 = 178.021, df = 102, CFI = .949, TLI = .923, RMSEA = .047, SRMR = .034; ω = .88). The motivation for teaching was affected by psychological well-being indicators, particularly in the precarious group. The current study supported the use of the ESEM technique, especially for a complex and culturally oriented tool, as a simultaneously theory-driven and data-driven approach. Furthermore, the network analysis showed preliminary findings helpful to overcome a knowledge gap in the motivation of Italian teachers. Cross-cultural studies that compare different school systems using ESEM might provide an innovative perspective on assessing motivation for teaching.

Keywords: *Exploratory Structural Equation Model, Motivation for teaching, Psychometrics*

INTRODUCTION

Various methods were traditionally used to investigate the factor structure of the psychometric instruments. The factor analysis was developed to represent psychological constructs through latent factors that serve as the rationale for human behaviors (Morin, Myers & Lee, 2020). In other words, the observed and measured variables co-occur to identify a latent factor. Exploratory factor analysis (EFA) and Confirmatory factor analysis (CFA) were the most commonly used multivariate factor analysis techniques in psychological research within the family of structural equation modeling techniques. However, despite their widespread adoption, both EFA and CFA have some limitations.

In recent years, Exploratory Structural Equation Modeling (ESEM) has emerged as a challenge for psychometric evaluations and in the context of broader psychometric modeling (Alamer, 2022; Asparouhov & Muthén, 2009; Marsh, Liem, Martin, Morin & Nagengast, 2011; Perry, Nicholls, Clough & Crust, 2015). Previous research (Marsh, Morin, Parker & Kaur, 2014) suggested that ESEM is primarily a confirmatory technique, but it can be used exploratorily through a target rotation and cross-loading between items un-forced to zero. The ESEM approach, in particular, combines the advantages of traditional techniques. With a data-driven perspective, the EFA suggests the lowest number of dimensions that can adequately explain the covariation observed among a set of observed variables (Brown, 2015). However, it can make interpreting the factors extracted through labels difficult at times. Furthermore, researchers are frequently required to make multiple decisions, which, if not guided by strong theoretical knowledge, can lead to poor results (Morin et al., 2020). Based on these considerations, the ESEM advantages reflect the ability to freely estimate previously unknown parameters (i.e., cross-loadings). Indeed, the cross-loadings are not constrained to zero, allowing us to estimate the association between each item and all possible factors. The CFA, on the other hand, suggests the best factor solution from a theory-driven standpoint. In a parsimonious model, each item is a priori assigned to a specific latent factor based on previous theories. However, it is frequently more simplistic and restrictive for most psychological constructs because it assumes that an item loads to a single pure factor (i.e., cross-loadings are constrained to be zero), which represents an operationalized simplification and artificially idealistic representation of a portion of human complexity (Mai, Zhang & Wen, 2018; Marsh et al., 2009). The ESEM approach, like the CFA, allows the researcher to test an a priori theory-driven hypothesis and compare the goodness-of-fit indices provided by various models. Instead, ESEM may provide a more accurate representation of complex psychological constructs: freely esteemed cross-loadings may compensate for wording effects and cross-cultural shades in the interpretation of previously ignored variance. As expected, when CFA and ESEM models were compared, the latter fit the data better than the former because the estimated factor correlation is less biased than in traditional CFA (Marsh et al., 2014). Indeed, the ESEM could solve common multicollinearity problems between latent factors by incorporating not previously estimated variability parts (i.e., cross-loadings) into the model and reducing the correlation between factors. Consequently, the relationships with other variables (e.g., convergent and discriminant validity) will also be more accurate. Moreover, the exploratory approach could be seen in parameter estimation (Boffo, Mannarini & Munari, 2012): using maximum likelihood (ML) estimation or estimation methods robust to non-normality and EFA familiar loading matrix rotation methods that allow the identification of latent factors. Regarding the choice of rotation, the geomin rotation is preferable with not very complex models; the target rotation is used in a confirmatory way, that is when we make assumptions a priori compared to cross-loadings; the orthogonal rotation is used when one general factor (G) and specific factors (S) should be estimated as totally independent as well as in bifactor-ESEM models (Asparouhov & Muthén, 2009; Morin, Arens & Marsh, 2016; Xiao, Liu & Hau, 2019). Despite the fact that ESEM is mostly used to investigate the factorial structure of psychometric tools, its advantages can also be seen as part of more complex Structural Equation Models (SEM). Indeed, the presence of ESEM parts in articulated SEM models may improve their fit indices. Again, inserting new portions of variance into the model will result in a more accurate representation of reality.

However, several weakness and areas of application that have yet to be fully implemented should be discussed: i) at this time, it is not possible to estimate the bootstrap confidence intervals required to estimate the indirect effects within the mediation reports; ii) ESEM are not appropriate for mixed or multilevel models because the second-order latent factor should be based on estimates of first-order latency

correlations; iii) it is not possible to estimate partial invariance between factors and ESEM placed within larger models may present convergence problems; iv) ESEM models may be over-fitted; this evidence requires constant attention also at item analysis levels. To address some of these limitations, two types of ESEM have been proposed: set-ESEM and ESEM-within-CFA (Marsh, Guo, Dicke, Parker & Craven, 2020). The set-ESEM allows for more complex parameter estimation in accordance with both data-driven and theory-driven set determination: the parameters are freely estimated into the set (i.e., specific part of the model) and constrained between sets (e.g., in a complex SEM model, the set could be represented by items of specific psychometric tools. In a confirmatory way, we could constrain the parameters into known relationships between variables and freely estimate the factor structures between the administered instruments). The ESEM-within-CFA model is not an alternative to the full-ESEM model (i.e., they show the same results as full-ESEM). Rather, they outline simpler computational aspects that effectively address ESEM's convergence problems as part of more complex SEM models. In conclusion, these two types can broaden the technique's applicability within more complex analysis frameworks.

Finally, several studies have shown the advantages of using ESEM for the study of the dimensionality of psychometric instruments (e.g., Marsh, Nagengast, Morin, Parada, Craven & Hamilton, 2011; Perera, 2015; Tóth-Király, Morin, Bőthe, Orosz & Rigó, 2018) also in the Italian context (Boffo et al., 2012). Our goal was to discuss a recent useful technique for the validation of psychometric tools and show an empirical application.

The Orientation for Teaching Survey: A culturally oriented psychometric tool

The organization of work in Italian public schools is a widespread issue. In 2022, there were 914.839 teachers working in schools, with 703.169 on permanent contracts and 211.670 (23.1%) on fixed-term contracts (Anief, 2022). These data are worrying because they favour precarious work. Recent findings (Marzano, De Angelis & Vegliante, 2015) highlighted the effects of precarious work on teaching motivation in a group of Italian workers. The condition of instability affected most aspects of life, such as limited future planning due to economic uncertainty, difficulty recognizing oneself within a defined professional role, and slow rejection of vocational ideals (e.g., desire to work with young people, desire to teach and transfer their discipline's content, desire to contribute to society's improvement). Furthermore, teachers with a permanent employment contract face challenges in the Italian school (Marzano et al., 2015): some people enter this field by chance or in search of economic stability; however, the general lack of resources (i.e., financial and materials), unsatisfactory social prestige for the role of teacher, disruptive student behavior, and salaries that are frequently considered too low may reduce their motivation to teach (Kelly & Northrop, 2015; You & Conley, 2015).

The Orientation for Teaching Survey (OTS) was a psychometric instrument developed in America by Ferrell and Daniel (1993) to evaluate motivation for teaching. It was a 58-item tool with a 5-points Likert scale ranging from 1 (strongly agree) to 5 (strongly disagree), based on eight themes from Lortie's (1975) and Joseph and Green's (1986) theories. However, after preliminary research, the authors concluded that the 6-factor solution fits the data better than the 8-factor solution (i.e., Job security, Worthwhile service to society, Interpersonal relationships, Intellectual stimulation, Material benefits, and Continuation of work in a familiar setting). Unfortunately, various cross-cultural validation studies for OTS (e.g., Anghel, 2013; Sinclair, Dowson & Mcinerney, 2006) revealed inconsistent factor solutions (for a summary, see Simić, Purić & Stančić, 2018). Indeed, various motivation theories have been developed that seem to fit better in different cultural contexts. Additionally, in the study of the factorial structure of the OTS, the most common theories on teaching motivation based on two (i.e., intrinsic and extrinsic motivation; Ryan & Deci, 2000) or three factors (i.e., intrinsic, extrinsic, and altruistic motivation; Morgan, Kilpatrick, Abbott, Dallat & McClune, 2001) produced significant results. Despite the fact that the original paper (Ferrell & Daniel, 1993) did not include information about reliability properties, the Australian version (Sinclair et al., 2006) revealed Cronbach's alpha ranged from .58 to .78 for subscales, and the Rumen (Anghel, 2013) and Serbian (Jovanović, Bogdanović & Simić, 2013) version revealed Cronbach's alpha of .86 to .94 for the total score, respectively. Finally, Sinclair et al. (2006) highlighted high factor loadings and relatively low cross-loadings at the item analysis level. In conclusion, the OTS has been a useful tool for the purpose of this work due to its cultural specificities and psychometric features.

The present study

The current study sought to consider the strengths and drawbacks of the ESEM technique for investigating the dimensionality of psychometric instruments. In particular, we investigated and compared various CFA, full ESEM, and bi-factor ESEM models based on previously suggested factor solutions (i.e., 6, 3, and 2-factor models, as well as a short version) and exploratory structures of OTS, a culturally oriented tool. Furthermore, a series of network analyses were carried out in order to fill a knowledge gap regarding the unexplored relationship between motivation for teaching and indicators of psychological well-being (i.e., intolerance of uncertainty, optimism, anxiety, depression, and stress) in Italy, and the network was compared between precarious and in-role groups to test the model's invariance.

METHOD

Participants and procedure

Participants were 338 Italian teachers who take part in an online survey in May 2022. They were mostly women (women: n = 262, 77.5%; men: n = 72, 21.3%) with four (1.2%) preferring not to specify their gender. Their mean age was 46.84 (SD = 10.65) and their mean seniority was 15.30 (SD = 11.53). At the time of administration, 229 (67.8%) participants had a permanent employment contract in school (generally referred to as 'in-role' in Italy) and 109 (32.2%) had a fixed-term contract (generally referred to as 'precarious' in Italy). Participants spent an average of 15 minutes responding to the survey via a Google form with a mandatory response format to avoid missing data. According to our eligibility criteria, Italians aged 18 and up who worked in Italian schools voluntarily participated in the research. Following an explanation of the study's objectives, all participants provided their informed consent.

The research project proposal was carried out in accordance with the Declaration of Helsinki, and it was approved by the internal review board for psychological research of the University of Enna "Kore". The measures were administered in accordance with the privacy guarantee regulations outlined in legislative decree no. 196/2003 and the GDPR (EU Regulation no. 2016/679).

Measures

Our online survey consisted of ad hoc items – which were created to detect the sample's sociodemographic characteristics – and self-report psychometric instruments to assess our selected variables. They will be presented below.

- *Motivation for teaching*. The English version of OTS was independently translated into Italian by two Italian native speakers. The two translations were then compared, and no significant differences were found. The first final version was back-translated into English by one bilingual speaker who was familiar with the psychological topic. After comparing the back-translation to the original version, a minor revision was required (see Appendix for items). In addition to the OTS, ad hoc items were developed to detect the sample's sociodemographic characteristics and other constructs were evaluated to assess relationships with motivation for teaching.
- *Self-esteem*. We administered the *Rosenberg Self-Esteem Scale* (*RSES*; Rosemberg, 1989) to evaluate self-esteem levels. It is a 10-items self-report scale (e.g., "I think I have a number of qualities", "I guess I don't have much to be proud") with a 4-points Likert scale ranging from 1 (strongly agree) to 4 (strongly disagree). We used the Italian version (Prezza, Trombaccia & Armento, 1997), which showed good internal consistency (Cronbach's α =.84). Cronbach's α for the present sample was .92.
- *Anxiety, depression, and stress*. We used the *Depression Anxiety Stress Scales – 21* (*DASS-21*; Lovibond & Lovibond, 1995) in its Italian version (Bottesi et al., 2015) to assess stress, anxiety, and depression in a unique psychometric instrument. It is a 21-items self-report scale (e.g., "I felt a lot of tension and I had difficulty recovering a state of calm", "I just couldn't feel any positive emotions", "I felt stressed out") using a 4-point Likert scale ranging from 0 (never happened to me) to 3 (it happened to me almost always) with good internal consistency and temporal stability both in the original version (Lovibond & Lovibond, 1995) (anxiety Cronbach's α = .74; depression Cronbach's α = .82; stress Cronbach's α = .85; total Cronbach's α = .90) and in the present sample (anxiety Cronbach's α = .86; depression Cronbach's α = .90; stress Cronbach's α =.89; total Cronbach's α = .95).
- *Intolerance of uncertainly*. The short version of the *Intolerance of Uncertainty Scale* was used (*IUS12*; Carleton, Norton & Asmundson, 2007). It is a 12-item scale with a

5-point Likert response format ranging from 1 (not at all characteristic of me) to 5 (entirely characteristic of me), with higher scores indicating higher levels of intolerance of uncertainty. In addition to a total score, two dimensions of uncertainty intolerance can be evaluated: perspective intolerance of uncertainty (e.g. "When things suddenly happen, I get very nervous"), and inhibitory uncertainty intolerance (e.g. "Feeling uncertain blocks me in making the most of things"). In the present study, we used the Italian version of IUS-12 (Bottesi et al., 2015) which demonstrated good psychometric properties. Cronbach's α for the present sample was .92.

– *Optimism*. The *Life Orientation Test-Revised* (*LOT-R*; Scheier, Carver & Bridges, 1994) in its Italian version (Giannini, Schuldberg, Di Fabio & Gargaro, 2008) was used to explore the teachers' optimism. The LOT-R is a 10-items scale. It employs a 5-point scale ranging from 0 (strongly disagree) to 4 (strongly agree) for 10 items (e.g., "I'm always optimistic about my future", "I hardly ever expect things to go right"). The total score ranges from 6 to 30, with higher scores indicating higher optimism levels. Cronbach's α values ranged from .74 to .78. Cronbach's α for the present sample was .70.

Data analysis

Following preliminary analysis to ensure that the data was normal, we ran and compared a series of CFA and ESEM models to the previously suggested factorial solutions for the OTS. The Maximum Likelihood Robust (MLR) estimation method was used because it did not assume the data's normality distribution. We reported the goodness-of-fit indices and measurement quality indicators and selected only the models that met the following cut-off criterion: Root Mean Square Error of Approximation (RMSEA) .06-.08 marginally acceptable and .01-.05 excellent, non-significant (p>.01), 90% confidence interval range should not include zero; Comparative Fit Index (CFI) .90-.95 marginally acceptable and .96-.99 excellent; Tucker-Lewis Index (TLI) .90-.95 marginally acceptable and .96-.99 excellent; Standardized Root Mean Residual (SRMR) .06-.08 marginally acceptable and .01-.05 excellent (Marsh, Hau & Wen, 2004; Marsh, Hau & Grayson, 2005). We used Mplus Version 7 (Muthén & Muthén, 2012) for the current analysis and supplemented the selected model's information with indices estimated by Dueber bifactor indices calculator (Dueber, 2017). Finally, we used the R-packages psychonetrics (Epskamp, 2020) and qgraph (Epskamp, Cramer, Waldorp, Schmittmann & Borsboom, 2012) to conduct a series of network analyses in RStudio v.2022.12.0+353 (RStudioTeam, 2020) to investigate the relationships between motivation for teaching and other related variables and compared the evidence between groups (i.e., in-role and precarious teachers). The psychometrics network model for Gaussian graphical models (GGMs) in particular creates an undirected network of partial correlations and it is useful for cross-sectional data (Epskamp, 2020).

RESULTS

Preliminary data processing

First, the minimum sample size required for ESEM estimation was a priori calculated (Cohen, 2013; Soper, 2022; Westland, 2010). We estimated it for a medium effect size (i.e., .3), a desired statistical power level of 80%, a confidence interval of 95%, and the most expensive model among those provided (i.e., 58 observed variables and 3 latent factors). The minimum sample size required to detect the effect was 119. As a result, our sample meets the requirements to a large extent.

Following data collection, the dataset was screened for potential issues. The multivariate normality of data was checked by the computation of Mahalanobis' distance, which revealed that the data were not distributed normally (i.e., Mardia's multivariate Kurtosis coefficient = 3799.94; df = 59~60, chi-square critical value = 99.608 when p<.001). Furthermore, no missing data were discovered as a result of our Google form's mandatory response format.

Comparison between CFA and ESEM models

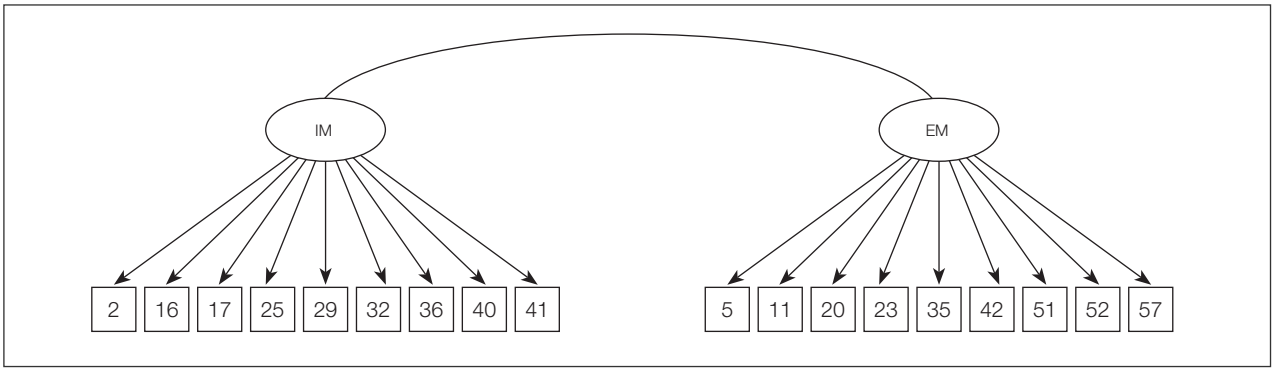
To identify the best fitting model, we performed a series of CFA, full ESEM, and bi-factor ESEM models based on both previously suggested and new models, according to the exploratory nature of the ESEM approach (i.e., 6, 3, 2, factors models and short version models; see Table 1 and Figure 1a, 1b, 1c). In line with previous suggestions (Morin et al., 2016),

Table 1 – Estimated models

Model	Type	χ^2	*p*	*df*	CFI	TLI	RMSEA [90% C.I.]	SRMR	AIC	BIC	aBIC	Meets criteria
6-factors												
Model 1	CFA	4937.419	.00	1580	.605	.587	.079 [.077-.082]	.109	54222.760	54945.316	54345.778	No
Model 2	ESEM	2647.520	.00	1320	.838	.797	.055 [.052-.058]	.041	51477.558	53187.414	51763.156	No
Model 3	ESEM bifactor	2608.684	.00	1268	.820	.765	.056 [.053-.059]	.037	51328.810	53236.689	51647.484	No
3-factors												
Model 4	CFA	4553.428	.00	1592	.652	.639	.074 [.072-.077]	.096	53814.769	54491.448	53929.977	No
Model 5	ESEM	3405.011	.00	1482	.741	.711	.062 [.060-.065]	.053	52031.810	53124.747	52214.364	No
Model 6	ESEM bifactor	3097.389	.00	1427	.775	.739	.059 [.056-.062]	.048	51781.682	53084.066	51999.220	No
2-factors												
Model 7	CFA	4640.517	.00	1594	.642	.629	.075 [.073-.078]		53897.857	54566.890	54011.764	No
Model 8	ESEM	3901.168	.00	1538	.711	.690	.068 [.065-.071]	.058	52295.206	53174.887	52442.140	No
Model 9	ESEM bifactor	3525.772	.00	1482	.750	.722	.064 [.062-.067]	.053	52031.810	53124.747	52214.364	No
Short version 2-factors												
Model 10	CFA	742.915	.00	134	.656	.607	.116 [.108-.124]	.113	17457.608	17667.876	17493.407	No
Model 11	ESEM	274.782	.00	118	.907	.880	.063 [.053-.073]	.042	16665.785	16936.164	16710.947	Yes
Model 12	ESEM bi-factor	178.021	.00	102	.949	.923	.047 [.036-.059]	.034	16613.545	16944.854	16668.884	Yes

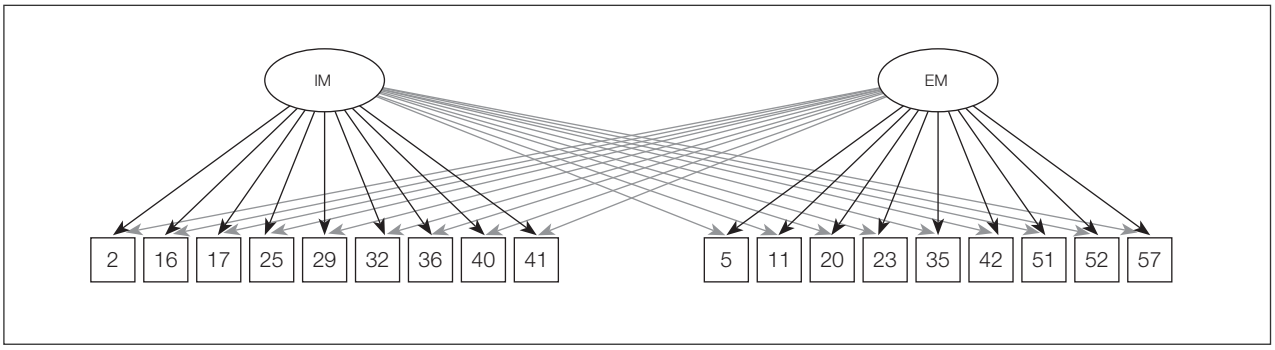
Legenda. CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; aBIC = Akaike's Bayesian Information Criterion.

Figure 1a – CFA two factor model for OTS short version (Model 10)



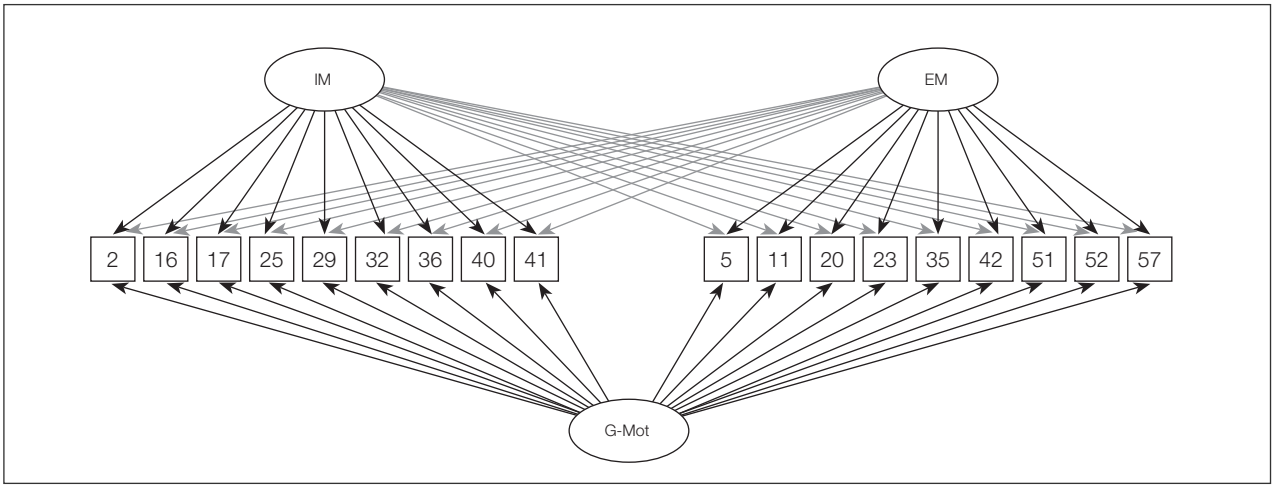
Legenda. IM = Implicit motivation for teaching; EM = Explicit motivation for teaching.

Figure 1b – Full-ESEM model for OTS short version (Model 11)



Legenda. IM = Implicit motivation for teaching; EM = Explicit motivation for teaching.

Figure 1c – ESEM Bi-factor model for OTS short version (Model 12)



Legenda. IM = Implicit motivation for teaching; EM = Explicit motivation for teaching; G-Mot = General factor-motivation for teaching.

Table 2 – Item level descriptive statistics, standardized loadings, reliability estimates and explained common variance of ESEM and ESEM Bi–factor model of OTS short version

Factors						Model 11 ESEM						Model 12 ESEM–bifactor								
	Item	Mean	*SD*	Skewness	Kurtosis	OTS–IM		OTS–EM				G–Mot		OTS–IM		OTS–EM				I–ECV
						λ	S.E.	λ	S.E.	δ	R^2	λ	S.E.	λ	S.E.	λ	S.E.	δ	R^2	
OTS–IM																				
	2	4.64	.65	−2.00	4.45	**.442**	.053	−.029	.054	.811	.189	**.168**	.051	**.407**	.066	−.004	.062	.806	.194	.146
	16	1.76	1.10	1.52	1.56	−.082	.057	**.646**	.041	.604	.396	**.549**	.063	**−.138**	.061	**.365**	.063	.547	.453	.664
	17	4.48	.82	−1.89	4.08	**.449**	.053	.017	.054	.794	.206	**.223**	.053	**.396**	.067	−.019	.064	.793	.207	.240
	25	2.68	1.34	.31	−.97	**.193**	.055	**.558**	.041	.595	.405	**.647**	.041	.030	.057	−.023	.067	.579	.421	.997
	29	4.14	1.12	−1.20	.68	**.779**	.036	−.062	.043	.420	.580	**.299**	.063	**.710**	.064	−.045	.057	.405	.595	.150
	32	3.62	1.39	−.59	−.90	**.584**	.046	.045	.049	.643	.357	**.302**	.058	**.535**	.068	.037	.076	.621	.379	.241
	36	2.86	1.27	.11	−.90	**.220**	.052	**.630**	.041	.482	.518	**.781**	.041	−.009	.070	**−.177**	.054	.359	.641	.951
	40	3.82	1.17	−.76	−.20	**.669**	.040	.018	.045	.546	.454	**.313**	.059	**.614**	.054	.025	.068	.524	.476	.206
	41	4.55	.73	−1.91	4.47	**.677**	.042	**−.093**	.046	.566	.434	**.249**	.065	**.581**	.059	**−.191**	.078	.564	.436	.142
OTS–EM																				
	5	2.38	1.19	.43	−.73	−.047	.048	**.647**	.041	.595	.405	**.612**	.044	**−.187**	.056	.042	.095	.589	.411	.911
	11	2.14	1.16	.76	−.24	.087	.047	**.623**	.041	.567	.433	**.628**	.053	−.030	.046	.142	.105	.584	.416	.949
	20	4.21	.94	−1.34	1.86	**.478**	.052	−.002	.060	.772	.228	**.241**	.067	**.390**	.070	−.153	.091	.766	.234	.249
	23	2.89	1.40	.18	−1.17	**.116**	.054	**.477**	.050	.730	.270	**.539**	.051	−.031	.061	−.072	.095	.703	.297	.979
	35	3.72	1.16	−.63	−.38	**.213**	.054	**.391**	.053	.758	.242	**.518**	.048	.052	.061	−.144	.099	.708	.292	.920
	42	2.92	1.49	.05	−1.34	**.326**	.053	**.310**	.055	.754	.255	**.472**	.052	.193	.057	−.093	.085	.731	.269	.829
	51	2.58	1.29	.33	−.93	**.162**	.049	**.554**	.045	.619	.381	**.606**	.048	.031	.046	.058	.108	.629	.371	.974
	52	2.01	1.23	1.15	.34	.001	.048	**.628**	.042	.605	.395	**.560**	.068	−.039	.046	**.456**	.097	.477	.523	.600
	57	1.59	.97	1.93	3.48	−.083	.049	**.628**	.043	.627	.373	**.521**	.073	−.114	.042	**.487**	.103	.478	.522	.520

	Omega	ECV	Omega	Omega_{hs}
G–Mot	—	.603	.880	—
OTS–IM	.753	.293	.880	.127
OTS–EM	.782	.104	.880	.003

Legenda. G–Mot = General factor–motivation for teaching; OTS–IM = Implicit motivation for teaching; OTS–EM = Explicit motivation for teaching; ECV = Explained Common Variance; I–ECV = Item–Explained Common Variance.

Bold items = significant target loadings (p<.05); underlined items indicate cross–loading items; S.E. = Standard Error; λ = standardized factor loadings; δ = item uniqueness.

we did not investigate hierarchical models because they simply involve replacing CFA factor correlations into a high-order factor, thus resulting in an empirically equivalent model in terms of degrees of freedom and fit indices. Furthermore, we did not consider any constraints because they were not supported by reference theory or previous validation studies.

Specifically, our findings showed that all CFA models failed to meet the criteria. Indeed, among the ESEM estimated models - with a target rotation - only the short versions (i.e., 2-factors and bi-factor with 2 first-order factors models) met the required criteria (Model 11: χ^2 = 274.782, *df* = 118, CFI = .907, TLI = .880, RMSEA = .063, SRMR = .042, AIC = 16665.785, BIC = 16936.164 and Model 12: χ^2 = 178.021, *df* = 102, CFI = .949, TLI = .923, RMSEA = .047, SRMR = .034, AIC = 16613.545, BIC = 16944.854, respectively). After model comparison, Model 12 (i.e., ESEM bi-factor model with two specific first-order factors) demonstrated the best data fitting (i.e., model with highest CFI, TLI, ΔRMSEA≤.015, ΔSRMR≤.015, and lowest χ^2, AIC, BIC, aBIC). Despite the Model 12 had a slightly high BIC than the Model 11, recent research (Cao & Liang, 2022) has revealed that this information criterion was biased in the ESEM technique, favouring the more parsimonious model, as well as in our results. According to previous research (Ryan & Deci, 2000), the general factor was labelled 'General motivation for teaching' and the two specific first-order factors were labelled 'Implicit and explicit motivation for teaching', respectively (see Figure 1c).

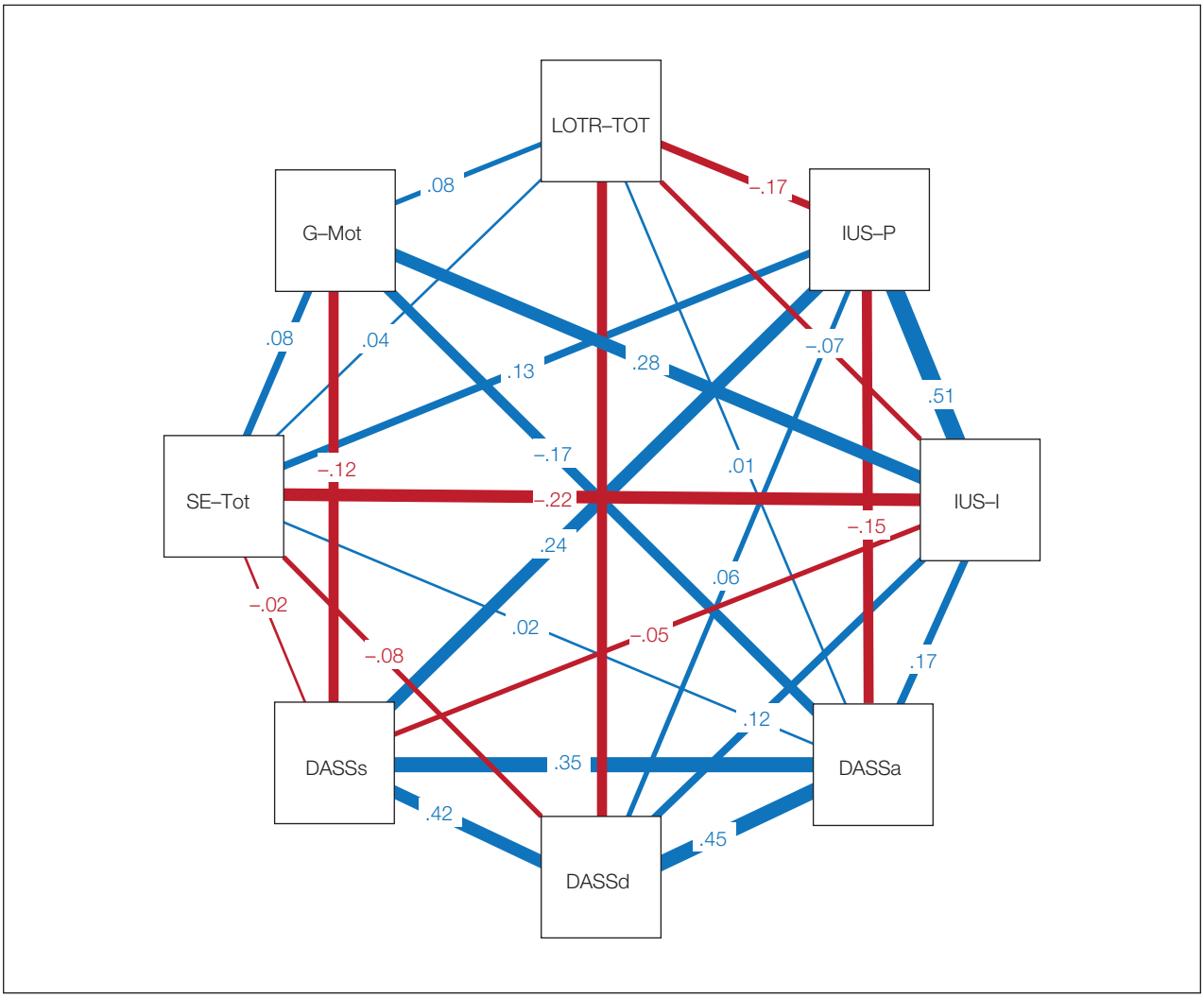
Beyond the fit indices, the comparison between Model 11 and 12 revealed that the latter had better defined factors (see Table 2). Notably, at the item level parameters, Model 11 displayed some issues in both the explicit and implicit dimensions: the items 16, 25, 36 did not exceeded the threshold (loadings λ>.30, *p*<.05) for the expected factor (i.e., implicit motivation), but did exceed it to the explicit motivation. Likewise, item 20 exceeded the threshold only for the unexpected factor (i.e., implicit motivation). Moreover, the items 20, 23, 35, 42, and 51 showed statistically significant loadings (*p*<.05) also for the implicit motivation. These findings suggested an unclear factor definition. On the other hand, the Model 12 generated a well-defined general factor with all items' *p*<.001 that explains most of the variance (loadings λ>.30, except for items 2, 17, 29, 41, and 20). The first-order specific factors were also well-defined for the same criteria, with the exception of items 16, 25, and 36, which refer to implicit motivation, and items 5, 11, 20, 23, 35, 42, and 51, which refer to explicit motivation. However, these results are consistent with the bi-factor model assumptions that the general factor explained the most part of the items' variance and the specific first-order factors - which showed the most minus cross-loadings - reflect the portion of the items' residual variance that the general factor does not explain (Morin et al., 2016). Additionally, although McDonalds's omega showed adequate reliability for Model 11 (Omega = .75 and .78, respectively), this value was increased in Model 12 (Omega = .88). Finally, as an additional support for the bifactorial structure, the proportion of common variance was mostly explained by General motivation for teaching (i.e., ECV for general factor = .603 and specific factors = .293, and .104, respectively). Indeed, the I-ECV (Stucky, Thissen & Edelen, 2013) - which refers to the proportion of common variance explained by the general factor at the item level (i.e., I-ECV>.80) - suggested that items 25, 36, 5, 11, 23, 35, 42, and 51 are essentially unidimensional. In summary, the Model 12 showed better psychometric properties than the Model 11 in terms of fit indices, item level parameters, and measurement quality criteria.

Network analysis for OTS

Based on the previous findings, we performed a series of network analyses for the OTS short version using the bi-factor ESEM solution. We investigated the relationship between teaching motivation and various indicators of psychological well-being (i.e., intolerance of uncertainty, optimism, anxiety, depression, stress, and self-esteem). On a graphical level, each variable is represented by a 'node', and the partial correlations are represented by 'edges' (Epskamp, 2020). First, using OTS's bi-factor structure, we investigated the relationship between general motivation for teaching and our indicators (*df* = 0, χ^2 = 0, *p* = 1, CFI = 1, RMSEA = 0; see Figure 2 and Table 3). The partial correlation matrix showed significant positive partial correlations between general motivation for teaching and inhibitory intolerance of uncertainty and anxiety (β = .028, *p*<.001; β = .17, *p*<.01, respectively) and significant negative correlations between general motivation for teaching and stress (β = −.012, *p*<.05), regardless of the presence of the other selected indicators in the network.

Secondly, we explored a network for in-depth knowledge that included first-order factors (i.e., extrinsic and intrinsic

Figure 2 – Network analysis between motivation for teaching and psychological well-being indicators



Legenda. Red lines: negative correlations; Blu lines: positive correlations; line thickness: association's intensity.
G-Mot = General factor-motivation for teaching; LOTR-tot = Revised Life Orientation Test; SE-tot = Rosenberg Self-esteem Scale; DASSs = Depression, Anxiety, and Stress Scale subscale Stress; DASSd = Depression, Anxiety, and Stress Scale subscale Depression; DASSa = Depression, Anxiety, and Stress Scale subscale Anxiety; IUS-I = Intolerance of Uncertainly Scale subscale Inhibitory; IUS-P = Intolerance of Uncertainly Scale subscale Perspective.

Table 3 – Partial correlation matrix for network analysis which include general motivation for teaching and psychological well–being indicators

Variable 1	Variable 2	β	Standard Error	*p*
IUS–P	LOTR–tot	–.17	.05	.002
IUS–I	LOTR–tot	–.07	.05	.20
DASSa	LOTR–tot	.01	.05	.80
DASSd	LOTR–tot	–.11	.05	.04
DASSs	LOTR–tot	–.00	.05	.96
SE–tot	LOTR–tot	.04	.05	.44
G–Mot	LOTR–tot	.075	.05	.17
IUS–I	IUS–P	.51	.04	.00
DASSa	IUS–P	–.15	.05	.00
DASSd	IUS–P	.06	.05	.25
DASSs	IUS–P	.24	.00	.25
SE–tot	IUS–P	.13	.05	.02
G–Mot	IUS–P	–.01	.05	.89
DASSa	IUS–I	.17	.05	.00
DASSd	IUS–I	.12	.05	.02
DASSs	IUS–I	–.05	.05	.33
SE–tot	IUS–I	–.22	.05	.00
G–Mot	IUS–I	.28	.05	.00
DASSd	DASSa	.45	.04	.00
DASSs	DASSa	.35	.05	.00
SE–tot	DASSa	.02	.05	.72
G–Mot	DASSa	.17	.05	.00
DASSs	DASSd	.42	.04	.00
SE–tot	DASSd	–.08	.05	.14
G–Mot	DASSd	–.01	.05	.89
SE–tot	DASSs	–.015	.05	.78
G–Mot	DASSs	–.12	.05	.03
G–Mot	SE–tot	.08	.05	.13

Legenda. G–Mot = General factor–motivation for teaching; LOTR–tot = Revised Life Orientation Test; SE–TOT = Rosenberg Self–esteem Scale; DASSs = Depression, Anxiety, and Stress Scale subscale Stress; DASSd = Depression, Anxiety, and Stress Scale subscale Depression; DASSa = Depression, Anxiety, and Stress Scale subscale Anxiety; IUS–I = Intolerance of Uncertainly Scale subscale Inhibitory; IUS–P = Intolerance of Uncertainly Scale subscale Perspective.

motivation for teaching) (df = 0, χ^2 = 0, p = 0, CFI = 1, RMSEA = 0; see Figure 3 and Table 4). The results revealed significant differences between the models. Specifically, the partial correlation matrix showed a significant positive association between i) extrinsic and intrinsic motivation (β = .028, p<.001), ii) anxiety, depression, and stress, and iii) inhibitory and perspective intolerance of uncertainty, which reflected the psychometric instruments' factor structure. Furthermore, focused on intrinsic motivation, the findings revealed positive partial correlations with self-esteem and anxiety (β = .015, p<.01; β = .014, p<.01, respectively), as well as negative partial correlation with stres (β = −.012, p<.05). On the other hand, focused on extrinsic motivation, we found a significant positive partial correlation with inhibitory intolerance of uncertainly (β = .018, p<.001).

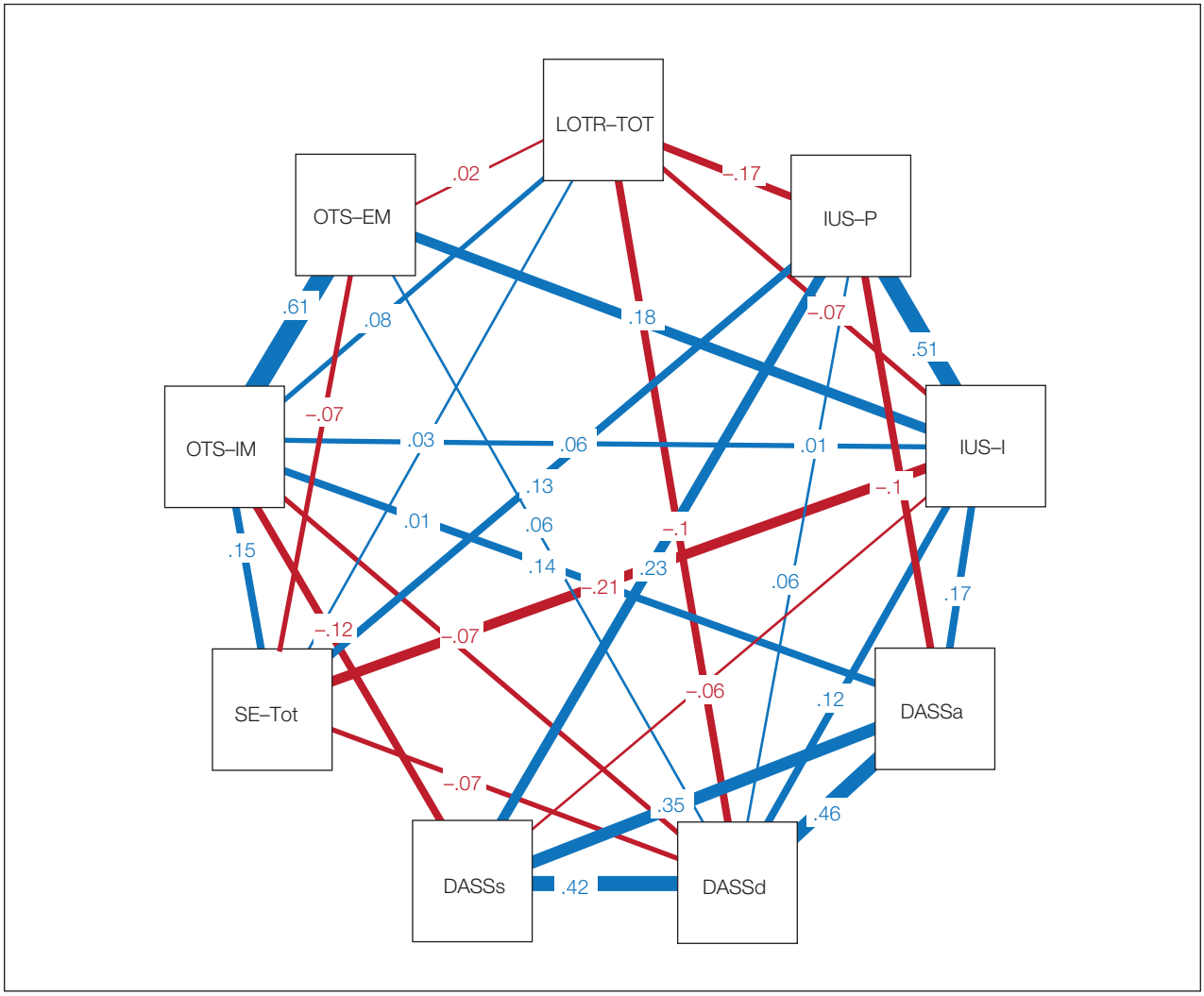
Finally, we tested the invariance between precarious and in-role groups using the general motivation for teaching in order to highlight the eventual differences. We compared two models: one model with free parameters (i.e., model-free) and the second model with parameters forced to be equal (i.e., model equal), using the 'compare' function of the psychometrics R-package, and found significant differences (model free: df = 0; RMSEA = not applicable; χ^2 = 0; model equal: df = 28; RMSEA = 0; χ^2 = −9.33). The findings revealed no significant differences in the direction of the associations between the two groups (see Figure 4 and 5; see Table 5). However, the most significant differences were found in the strength of these associations: high levels of motivation to teach in the precarious group are more positively associated with inhibitory intolerance of uncertainty (β = .37, p<.001), optimism (β = .026, p<.01) and negatively associated with stress (β = −.019, p<.05) compared to the in-role group (inhibitory intolerance of uncertainly: β = .24, p<.001; optimism: β = .01, p = .82; stress: β = −10, p = .13, respectively).

DISCUSSION

The present study aimed to reflect on the potential of the ESEM technique for the psychometric tools' factor structure. The OTS, a culturally oriented instrument, was studied as an application example. Indeed, the previous validation suggested various factor solutions that reflected cultural differences in various educational systems. First of all, the comparison between CFA, full-ESEM, and bi-factor ESEM models based on both previously suggested models and exploratory approach, according to the bivalent nature of the ESEM, highlighted the inadequacy of CFA models for the OTS in the Italian context. However, the ESEM models met the criteria for the short version in both full-ESEM (Model 11) and bi-factor ESEM (Model 12), with the latter performing better. This result may suggest that the ESEM technique explains the complexities of psychological construct better, especially when they are culturally determined. Indeed, cross-loadings un-forced to zero could introduce parts of unexplained variance into the model (Asparouhov & Muthén, 2009). This would allow us to estimate latent factors more adherent to reality and not pure. The Model 12 outperformed the Model 11 at various levels. In addition to improved fit indices, the factors are better defined at the item level. Although some factor loadings did not exceed the threshold criteria (λ>.30), these results could be considered acceptable due to their statistical significance. Furthermore, these results are consistent with the bi-factor model assumptions that the general factor explained the most part of items' variance (Morin et al., 2016). As an additional support, the bi-factor structure showed its potential: as suggested by omega - which showed higher values in Model 12 than in Model 11 - ECV, and I-ECV indices, the most common part of variance was explained by the general motivation for teaching. However, as a practical implication, maintaining a complex factorial structure rather than a one-dimensional structure reflects strong reference theories (Ryan & Deci, 2000) about motivation (i.e., intrinsic and extrinsic motivation) and allows for a more finely tuned assessment of motivation for teaching in a parsimonious short form.

Additionally, the network analysis also explored the relationship between motivation for teaching and psychological well-being indicators in the previously unexplored Italian context. Simultaneously, it may enable us to bridge the knowledge gap and suggest associations between variables. First, we explored the association between general motivation for teaching and the selected indicators. The results showed that high levels of general motivation for teaching are related to high levels of inhibitory intolerance of uncertainty and anxiety, as well as low levels of stress. Based on these findings and according with the traditional Yerkes-Dodson anxiety curve which considers anxiety also as a positive performance boost (Broadhurst, 1957), teachers who are highly motivated to teach may benefit from good levels of anxiety, which improves their performance

Figure 3 – Network analysis between extrinsic and intrinsic motivation and psychological well-being indicators



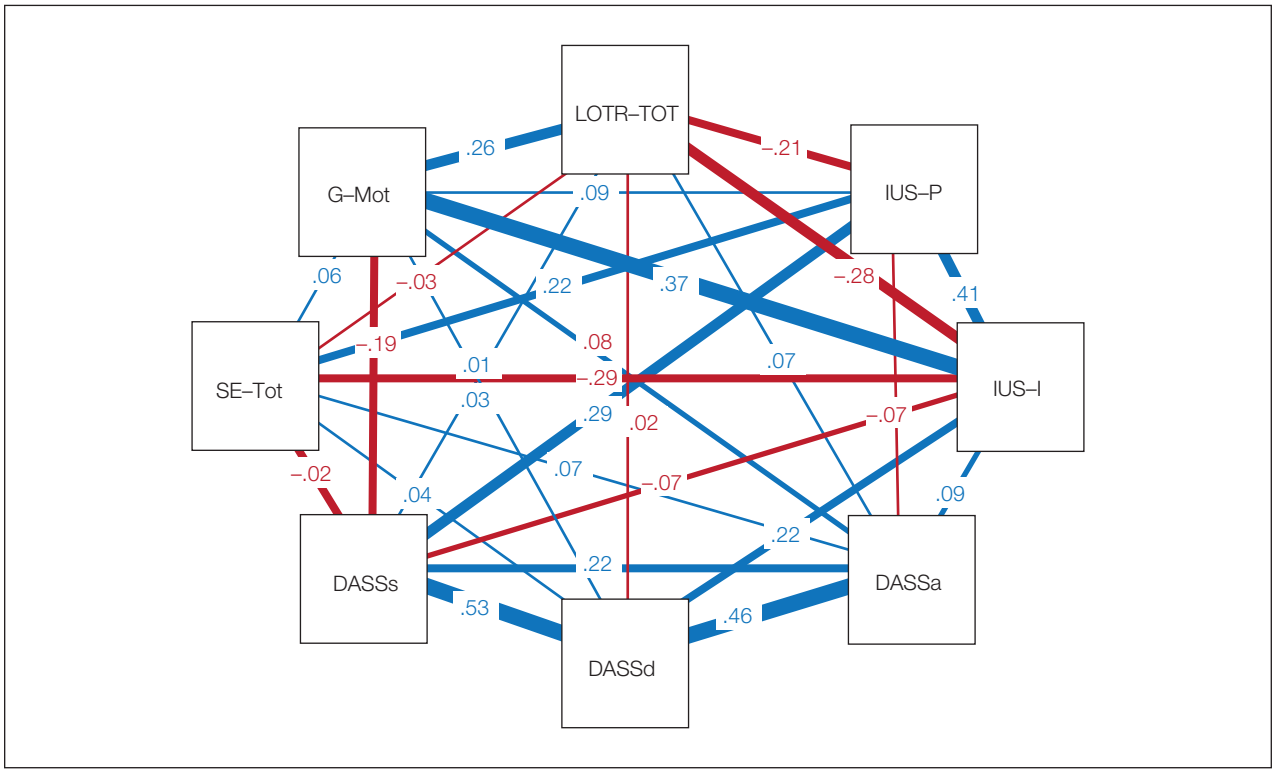
Legenda. Red lines: negative correlations; Blu lines: positive correlations; line thickness: association's intensity.
OTS-IM = Implicit motivation for teaching; OTS-EM = Explicit motivation for teaching; LOTR-tot = Revised Life Orientation Test; SE-tot = Rosenberg Self-esteem Scale; DASSs = Depression, Anxiety, and Stress Scale subscale Stress; DASSd = Depression, Anxiety, and Stress Scale subscale Depression; DASSa = Depression, Anxiety, and Stress Scale subscale Anxiety; IUS-I = Intolerance of Uncertainly Scale subscale Inhibitory; IUS-P = Intolerance of Uncertainly Scale subscale Perspective.

Table 4 – Partial correlation matrix for network analysis which include extrinsic and intrinsic motivation for teaching and psychological well–being indicators

Variable 1	Variable 2	β	Standard Error	*p*
IUS-P	LOTR-tot	−.17	.05	.002
IUS-I	LOTR-tot	−.07	.05	.22
DASSa	LOTR-tot	.01	.05	.87
DASSd	LOTR-tot	−.10	.05	.05
DASSs	LOTR-tot	.00	.05	.98
SE-tot	LOTR-tot	.035	.05	.53
OTS-IM	LOTR-tot	.08	.05	.12
OTS-EM	LOTR-tot	−.015	.05	.78
IUS-I	IUS-P	.51	.04	.00
DASSa	IUS-P	−.15	.05	.00
DASSd	IUS-P	.06	.05	.26
DASSs	IUS-P	.23	.05	.00
SE-tot	IUS-P	.13	.05	.02
OTS-IM	IUS-P	−.001	.05	.02
OTS-EM	IUS-P	−.00	.05	.99
DASSa	IUS-I	.17	.05	.00
DASSd	IUS-I	.12	.05	.03
DASSs	IUS-I	−.06	.05	.30
SE-tot	IUS-I	−.21	.05	.00
OTS-IM	IUS-I	.06	.05	.27
OTS-EM	IUS-I	.18	.05	.00
DASSd	DASSa	.46	.04	.00
DASSs	DASSa	.35	.05	.00
SE-tot	DASSa	.01	.05	.87
OTS-IM	DASSa	.14	.05	.01
OTS-EM	DASSa	.01	.05	.88
DASSs	DASSd	.42	.045	.00
SE-tot	DASSd	−.07	.05	.20
OTS-IM	DASSd	−.07	.05	.20
OTS-EM	DASSd	.06	.02	.27
SE-tot	DASSs	−.00	.05	.91
OTS-IM	DASSs	−.12	.05	.03
OTS-EM	DASSs	.01	.05	.91
OTS-IM	SE-tot	.15	.05	.01
OTS-EM	SE-tot	−.07	.05	.20
OTS-EM	SE-tot	.61	.03	.00

Legenda. OTS-IM = Implicit motivation for teaching; OTS-EM = Explicit motivation for teaching; LOTR-tot = Revised Life Orientation Test; SE-tot = Rosenberg Self-esteem Scale; DASSs = Depression, Anxiety, and Stress Scale subscale Stress; DASSd = Depression, Anxiety, and Stress Scale subscale Depression; DASSa = Depression, Anxiety, and Stress Scale subscale Anxiety; IUS-I = Intolerance of Uncertainly Scale subscale Inhibitory; IUS-P = Intolerance of Uncertainly Scale subscale Perspective.

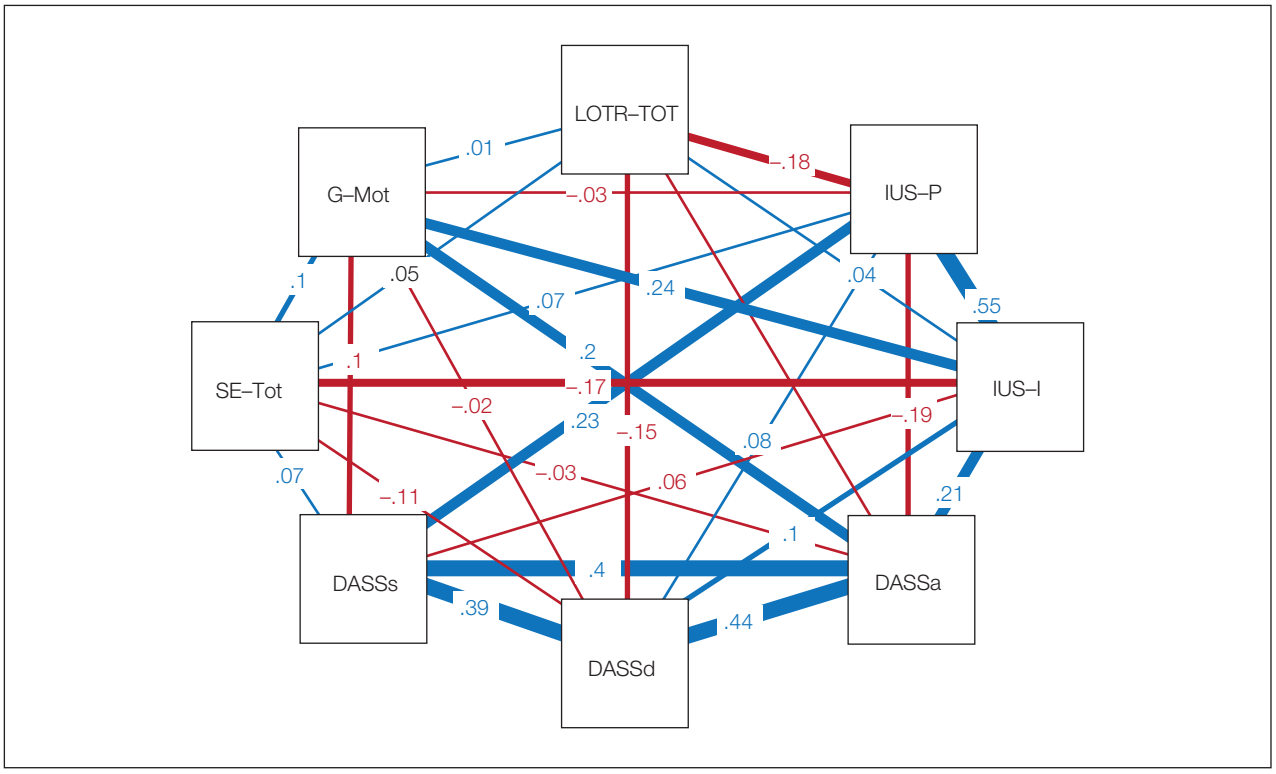
Figure 4 – Network analysis between motivation for teaching and psychological well-being indicators in precarious teachers group



Legenda. Red lines: negative correlations; Blu lines: positive correlations; line thickness: association's intensity.
G-Mot = General factor–motivation for teaching; LOTR-tot = Revised Life Orientation Test; SE-tot = Rosenberg Self-esteem Scale; DASSs = Depression, Anxiety, and Stress Scale subscale Stress; DASSd = Depression, Anxiety, and Stress Scale subscale Depression; DASSa = Depression, Anxiety, and Stress Scale subscale Anxiety; IUS-I = Intolerance of Uncertainly Scale subscale Inhibitory; IUS-P = Intolerance of Uncertainly Scale subscale Perspective.

and allows them to try a push to action. Working in a stressful environment, such as that found in some Italian schools (e.g., bad relationships with colleagues, old school structures, low pay), on the other hand, may have an impact on their motivation (Marzano et al., 2015). As a practical implication, it is crucial to pay close attention to the whole working environment and invest considerable resources in the teacher's well-being to promote quality education. Secondly, we estimated a network that includes the first-order factors (i.e., implicit and explicit motivation). High levels of intrinsic motivation are linked to high levels of self-esteem and anxiety and low levels of stress. As one possible interpretation, believing in oneself boosts self-confidence in one's abilities and good anxiety levels, allowing one to be more productive. On the other hand, high levels of extrinsic motivation are associated with high levels of inhibitory intolerance of uncertainty. As a viable explanation, being in a state of paralyzing uncertainty about the future can push one to pursue a teaching career for the financial stability and work-life balance that this work provides. Finally, we examined the network's invariance in the precarious and in-role groups. The results showed that there are no substantial

Figure 5 – Network analysis between motivation for teaching and psychological well-being indicators in in-role teachers group



Legenda. Red lines: negative correlations; Blu lines: positive correlations; line thickness: association's intensity.
G-Mot = General factor motivation for teaching; LOTR-tot = Revised Life Orientation Test; SE-tot = Rosenberg Self-esteem Scale; DASSs = Depression, Anxiety, and Stress Scale subscale Stress; DASSd = Depression, Anxiety, and Stress Scale subscale Depression; DASSa = Depression, Anxiety, and Stress Scale subscale Anxiety; IUS-I = Intolerance of Uncertainly Scale subscale Inhibitory; IUS-P = Intolerance of Uncertainly Scale subscale Perspective.

changes in the direction of association between variables between the two groups. Previous research (Marzano et al., 2015) discovered significant differences in the intensity of these associations: when compared to in-role teachers, high levels of motivation to teach in the precarious group are more associated with inhibitory and prospective intolerance of uncertainty and optimism (positive associations) and stress (negative association). Based on these findings, the motivation for teaching precarious teachers may be more oriented by uncertainty about the future and the search for stable employment than it is for in-role teachers who do not experience these dynamics. Also, optimism and stress were essential for the precarious: as a plausible explanation, spending every year in various schools, having to reinvent and re-environment their job in different situations, and frequently changing pupils and colleagues (situations not experienced by in-role teachers) were all very stressful experiences. The optimism for a better future remains the spring that allows precarious teachers to continue to work with motivation.

Table 5 – Partial correlation matrix for network analysis which general motivation for teaching and psychological well–being indicators between precarious and in rule groups

		Model Free						Model Equal					
		Precarious			In Rule			Precarious			In Rule		
Variable 1	Variable 2	β	Standard Error	*p*	β	Standard Error	p	β	Standard Error	*p*	β	Standard Error	*p*
IUS−P	LOTR−tot	−.21	.09	.02	−.18	.06	.01	−.48	<.0001	<.0001	−.48	<.0001	<.0001
IUS−I	LOTR−tot	−.28	.09	.00	.04	.07	.55	−.37	<.0001	<.0001	−.37	<.0001	<.0001
DASSa	LOTR−tot	.06	.10	.49	−.03	.07	.69	−.43	<.0001	<.0001	−.43	<.0001	<.0001
DASSd	LOTR−tot	−.02	.10	.84	−.15	.06	.02	−.58	<.0001	<.0001	−.58	<.0001	<.0001
DASSs	LOTR−tot	.01	.10	.90	.00	.07	.98	−.53	<.0001	<.0001	−.53	<.0001	<.0001
SE−tot	LOTR−tot	−.03	.10	.73	.05	.07	.47	.17	<.0001	<.0001	.17	<.0001	<.0001
G−Mot	LOTR−tot	.26	.09	.73	.01	.07	.82	−.09	<.0001	<.0001	−.09	<.0001	<.0001
IUS−I	IUS−P	.41	.08	.00	.55	.05	.00	1	<.0001	<.0001	1	<.0001	<.0001
DASSa	IUS−P	−.07	.10	.48	−.19	.06	.00	1	<.0001	<.0001	1	<.0001	<.0001
DASSd	IUS−P	−.01	.10	.89	.08	.07	.23	1	<.0001	<.0001	1	<.0001	<.0001
DASSs	IUS−P	.29	.09	.00	.23	.06	.00	1	<.0001	<.0001	1	<.0001	<.0001
SE−tot	IUS−P	.22	.09	.02	.07	.07	.30	−.27	<.0001	<.0001	−.27	<.0001	<.0001
G−Mot	IUS−P	.09	.09	.35	−.02	.07	.71	.60	<.0001	<.0001	.60	<.0001	<.0001
DASSa	IUS−I	.09	.09	.32	.21	.06	.00	1	<.0001	<.0001	1	<.0001	<.0001
DASSd	IUS−I	.22	.09	.02	.01	.07	.13	1	<.0001	<.0001	1	<.0001	<.0001
DASSs	IUS−I	−.07	.10	.44	−.06	.07	.37	1	<.0001	<.0001	1	<.0001	<.0001
SE−tot	IUS−I	−.29	.09	.00	−.17	.06	.00	−.48	<.0001	<.0001	−.48	<.0001	<.0001
G−Mot	IUS−I	.37	.08	.00	.24	.06	.00	.82	<.0001	<.0001	.82	<.0001	<.0001
DASSd	DASSa	.46	.07	.00	.44	.05	.00	1	<.0001	<.0001	1	<.0001	<.0001
DASSs	DASSa	.22	.09	.01	.40	.06	.00	1	<.0001	<.0001	1	<.0001	<.0001
SE−tot	DASSa	.07	.09	.46	−.03	.07	.63	−.52	<.0001	<.0001	−.52	<.0001	<.0001
G−Mot	DASSa	.08	.10	.39	.20	.06	.63	.88	<.0001	<.0001	.88	<.0001	<.0001
DASSs	DASSd	.53	.07	.00	.39	.06	.00	1	<.0001	<.0001	1	<.0001	<.0001
SE−tot	DASSd	.04	.10	.65	−.11	.07	.09	−.66	<.0001	<.0001	−.66	<.0001	<.0001
G−Mot	DASSd	.03	.10	.76	−.02	.07	.71	.79	<.0001	<.0001	.79	<.0001	<.0001
SE−tot	DASSs	−.20	.09	.03	.07	.07	.31	−.58	<.0001	<.0001	−.58	<.0001	<.0001
G−Mot	DASSs	−.19	.10	.04	−.10	.07	.13	.57	<.0001	<.0001	.57	<.0001	<.0001
G−Mot	SE−tot	.06	.10	.56	.10	.07	.12	−.11	<.0001	<.0001	−.11	<.0001	<.0001

Legenda. G−Mot = General factor−motivation for teaching; LOTR−tot = Revised Life Orientation Test; SE−TOT = Rosenberg Self−esteem Scale; DASSs = Depression, Anxiety, and Stress Scale subscale Stress; DASSd = Depression, Anxiety, and Stress Scale subscale Depression; DASSa = Depression, Anxiety, and Stress Scale subscale Anxiety; IUS−I = Intolerance of Uncertainly Scale subscale Inhibitory; IUS−P = Intolerance of Uncertainly Scale subscale Perspective.

Limitations and implications for further research

The results of the present study should be interpreted in light of some limitations. First, a convenient online sample and cross-sectional design were used. Further research in an Italian-speaking sample will be required to improve the generalizability of the findings. Secondly, we were unable to test the OTS's validity because there were no previous studies relevant to the Italian context that could be used to formulate hypotheses. Based on these considerations, the present paper provides preliminary exploratory knowledge that can be used in future research. In addition, based on the collected socio-demographic information we divided our sample into two groups (i.e., precarious and in-role teachers); however, further research could be divided into other groups (e.g., degree of school, years of experience, teaching subject) to gain more specific knowledge. Finally, cross-cultural studies that compare different school systems using the ESEM approach could provide an innovative perspective on assessing motivation for teaching.

CONCLUSION

The ESEM may provide an intriguing technique for testing the dimensionality of the psychometric tools, especially for complex and culturally oriented psychological constructs. The ESEM's combination of confirmatory and exploratory nature allows to combine theory-driven and data-driven advantages. Specifically, the OTS application of the ESEM technique for the Italian teachers offered the opportunity to reflect on the issues of cross-cultural validation while maintaining a complex structure and a strong reference to the theory of motivation in a parsimonious short version. Furthermore, preliminary findings from the network analysis might help Italian teachers overcoming a motivational gap. Indicators of psychological well-being had an impact on it. More specifically, intrinsic motivation seems to be influenced by intra-subject indicators (e.g., anxiety), whereas extrinsic motivation appears to be influenced by extra-subject indicators (e.g., uncertainty). Finally, focusing on teachers' contextual and personal well-being, especially in precarious situations, may improve their motivation. This can lead to better teaching quality with far-reaching implications for students and society as a whole.

In conclusion, we will not claim that this psychometric technique is necessarily useful for all instruments commonly used in psychological assessment. More broadly, we argue that this framework has the potential to become a standard tool in the study of multifaceted psychometric tools, especially those that present problems in cross-cultural validation. Ultimately, we encouraged the adoption of the ESEM technique, emphasizing that the common guidelines (Asparouhov & Muthén, 2009; Marsh, Liem et al., 2011a; Marsh et al., 2014; Morin et al., 2016) supported the comparison between CFA and ESEM models in order to select the more parsimonious model, while systematically demonstrating the advantages of the latter both theoretically (i.e., accurate representation of latent factors in reference to theory without "pure" factors) and technically (i.e., well-defined latent factors and better estimation of relationships with other variables) (Marsh et al., 2014). Indeed, an ESEM-based approach is more likely to produce better factorial solutions than the overly restrictive and widely used CFA strategy.

Author note: we have no conflicts of interest to disclose.

References

ALAMER, A. (2022). Exploratory structural equation modeling (ESEM) and bifactor ESEM for construct validation purposes: Guidelines and applied example. *Research Methods in Applied Linguistics*, *1* (1), 100005. doi.org/https://doi.org/10.1016/j.rmal.2022.100005

ANGHEL, I. (2013). Motivating aspects to become teachers in T-VET system. *Scientific Bulletin – Education Sciences Series*, *1*, 112-123.

ANIEF (2022). *Dossier Anief: "Nella scuola italiana, aumenta l'esercito di precari, il 22,4% del personale in servizio"*. Retrieved 20/12/2022 from https://anief.org/stampa/news/39736-dossier-anief-%E2%80%9Cnella-scuola-italiana,-aumenta-l%E2%80%99esercito-di-precari,-il-22,4-del-personale-in-servizio%E2%80%9D

ASPAROUHOV, T. & MUTHÉN, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16* (3), 397-438. doi.org/10.1080/10705510903008204

BOFFO, M., MANNARINI, S. & MUNARI, C. (2012). Exploratory structure equation modeling of the UCLA loneliness scale: A contribution to the Italian adaptation. *TPM: Testing, Psychometrics, Methodology in Applied Psychology*, *19* (4), 345-363.

BOTTESI, G., GHISI, M., ALTOÈ, G., CONFORTI, E., MELLI, G. & SICA, C. (2015). The Italian version of the Depression Anxiety Stress Scales – 21: Factor structure and psychometric properties on community and clinical samples. *Comprensive Psychiatry*, *60*, 170-181. doi.org/10.1016/j.comppsych.2015.04.005

BOTTESI, G., GHISI, M., NOVARA, C., BERTOCCHI, J., BOIDO, M., DE DOMINICIS, I. & FREESTON, M.H. (2015). Intolerance of Uncertainty Scale (IUS-27 e IUS-12): Due studi preliminari. [Intolerance of Uncertainty Scale (IUS-27 and IUS-12): Two preliminary studies.]. *Psicoterapia Cognitiva e Comportamentale*, *21*, 345-365. Retrieved (17/03/2023) from https://www.researchgate.net/publication/281899656_Intolerance_of_Uncertainty_Scale_IUS-27_and_IUS-12_Two_preliminary_studies

BROADHURST, P.L. (1957). Emotionality and the Yerkes-Dodson Law. *Journal of Experimental Psychology*, *54*, 345-352. doi.org/10.1037/h0049114

BROWN, T.A. (2015). *Confirmatory factor analysis for applied research*. Guilford publications.

CAO, C. & LIANG, X. (2022). Sensitivity of fit measures to lack of measurement invariance in exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *29* (2), 248-258. doi.org/10.1080/10705511.2021.1975287

CARLETON, R.N., NORTON, M.A.P.J. & ASMUNDSON, G.J.G. (2007). Fearing the unknown: A short version of the intolerance of uncertainty scale. *Journal of Anxiety Disorders*, *21* (1), 105-117. doi.org/https://doi.org/10.1016/j.janxdis.2006.03.014

CHRISTOPHER WESTLAND, J. (2010). Lower bounds on sample size in structural equation modeling. *Electronic Commerce Research and Applications*, *9* (6), 476-487. doi.org/https://doi.org/10.1016/j.elerap.2010.07.003

COHEN, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.

DUEBER, D.M. (2017). *Bifactor indices calculator: A Microsoft Excel-based tool to calculate various indices relevant to bifactor CFA models*. doi.org/https://doi.org/10.13023/edp.tool.01

EPSKAMP, S. (2020). Psychometric network models from time-series and panel data. *Psychometrika*, *85* (1), 206-231. doi.org/10.1007/s11336-020-09697-3

EPSKAMP, S., CRAMER, A.O.J., WALDORP, L.J., SCHMITTMANN, V.D. & BORSBOOM, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, *48* (4), 1-18. doi.org/10.18637/jss.v048.i04

FERRELL, C.M. & DANIEL, L.G. (1993). *Construct validation of an instrument measuring teacher career motivations*. Annual Meeting of the Mid-South Educational Research Association, New Orleans, LA.

GIANNINI, M., SCHULDBERG, D., DI FABIO, A. & GARGARO, D. (2008). Misurare l'ottimismo: Proprietà psicometriche della versione italiana del Life Orientation Test – Revised (LOT-R). [Measuring optimism: Psychometric properties of the Italian version of the Life Orientation Test – Revised (LOT-R)]. *Counseling: Giornale Italiano di Ricerca e Applicazioni*, *1*, 73-84.

JOSEPH, P.B. & GREEN, N.S. (1986). Perspectives on reasons for becoming teachers. *Journal of Education*, *37* (6), 28-33.

JOVANOVIĆ, O., BOGDANOVIĆ, S. & SIMIĆ, N. (2013). Povezanost bazičnih crta ličnosti i motivacije studenata „nastavničkih" fakulteta za posao nastavnika [Relationship between the basic personality traits and motivation of the students from the teacher faculties for the future profession]. In *Zbornik radova sa XIX naučnog skupa Empirijska istraživanja u psihologiji*. Filozofski fakultet.

KELLY, S. & NORTHROP, L. (2015). Early career outcomes for the "best and the brightest": Selectivity, satisfaction, and attrition in the beginning teacher longitudinal survey. *American Educational Research Journal*, *52 (4)*, 624-656. doi.

org/10.3102/0002831215587352

LOVIBOND, S. & LOVIBOND, P. (1995). *Manual for the depression anxiety stress scales*. Psychology Foundation of Australia.

MAI, Y., ZHANG, Z. & WEN, Z. (2018). Comparing exploratory structural equation modeling and existing approaches for multiple regression with latent variables. *Structural Equation Modeling: A Multidisciplinary Journal*, *25* (5), 737-749. doi.org/10.1080/10705511.2018.1444993

MARSH, H.W., GUO, J., DICKE, T., PARKER, P.D. & CRAVEN, R.G. (2020). Confirmatory factor analysis (CFA), exploratory structural equation modeling (ESEM), and set-ESEM: Optimal balance between goodness of fit and parsimony. *Multivariate Behavioral Research*, *55* (1), 102-119.

MARSH, H.W., HAU, K.T. & GRAYSON, D. (2005). Goodness of fit evaluation in structural equation modeling. In A. Maydeu-Olivares & J. McArdle (Eds.), *Contemporary psychometrics. A festschrift to Roderick P. McDonald*. Hillsdale, NJ: Erlbaum.

MARSH, H.W., HAU, K.T. & WEN, X. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indices and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, *11*, 320-341. doi:10.1207/s15328007sem1103_2

MARSH, H.W., LIEM, G.A.D., MARTIN, A.J., MORIN, A.J.S. & NAGENGAST, B. (2011). Methodological measurement fruitfulness of exploratory structural equation modeling (ESEM): New approaches to key substantive issues in motivation and engagement. *Journal of Psychoeducational Assessment*, *29* (4), 322-346. doi.org/10.1177/0734282911406657

MARSH, H.W., MORIN, A.J.S., PARKER, P.D. & KAUR, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, *10* (1), 85-110. doi.org/10.1146/annurev-clinpsy-032813-153700

MARSH, H.W., MUTHÉN, B., ASPAROUHOV, T., LÜDTKE, O., ROBITZSCH, A., MORIN, A.J. & TRAUTWEIN, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling: A Multidisciplinary Journal*, *16* (3), 439-476. doi.org/10.1080/10705510903008220

MARSH, H.W., NAGENGAST, B., MORIN, A.J., PARADA, R.H., CRAVEN, R.G. & HAMILTON, L.R. (2011). Construct validity of the multidimensional structure of bullying and victimization: An application of exploratory structural equation modeling. *Journal of Educational Psychology*, *103* (3), 701. doi.org/10.1037/a0024122

MARZANO, A., DE ANGELIS, M. & VEGLIANTE, R. (2015). The precarious teacher in Italy: An exploratory research. *Proceedings of EDULEARN15 Conference*.

MORAN, A., KILPATRICK, R., ABBOTT, L., DALLAT, J. & McCLUNE, B. (2001). Training to teach: Motivating factors and implications for recruitment. *Evaluation and Research in Education*, *5* (1), 17-32. doi.org/https://doi.org/10.2298/PSI170327012S

MORIN, A.J.S., ARENS, A.K. & MARSH, H.W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal*, *23* (1), 116-139. doi.org/10.1080/10705511.2014.961800

MORIN, A.J.S., MYERS, N.D. & LEE, S. (2020). Modern factor analytic techniques. In *Handbook of Sport Psychology*. doi.org/10.1002/9781119568124.ch51

MUTHÉN, L.K. & MUTHÉN, B.O. (2012). *Mplus User's Guide (7th ed.)*. CA: Muthén & Muthén.

PERERA, H.N. (2015). The internal structure of responses to the Trait Emotional Intelligence Questionnaire – Short form: An exploratory structural equation modeling approach. *Journal of Personality Assessment*, *97* (4), 411-423. doi.org/10.1080/00223891.2015.1014042

PERRY, J.L., NICHOLLS, A.R., CLOUGH, P.J. & CRUST, L. (2015). Assessing model fit: Caveats and recommendations for confirmatory factor analysis and exploratory structural equation modeling. *Measurement in Physical Education and Exercise Science*, *19* (1), 12-21. doi.org/10.1080/1091367x.2014.952370

PREZZA, M., TROMBACCIA, F.R. & ARMENTO, L. (1997). La scala dell'autostima di Rosenberg: Traduzione e validazione italiana. [The Rosenberg Self-Esteem Scale: Italian translation and validation]. *BPA Bollettino di Psicologia Applicata*, *223*, 35-44.

ROSENBERG, M. (1989). *Society and the adolescent self-image – Revised version*. Wesleyan University Press.

RSTUDIO TEAM (2020). *RStudio: Integrated development environment for R.* RStudio, PBC. www.rstudio.com/

RYAN, R.M. & DECI, E.L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25* (1), 54-67. doi.org/https://doi.org/10.1006/ceps.1999.1020

SCHEIER, M.F., CARVER, C.S. & BRIDGES, M.W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology*, *67* (6), 1063-1078. doi.org/10.1037//0022-3514.67.6.1063

SCRIVEN, E.G. & PARRIOTT, K. (1975). Schoolteacher: A

sociological study. *Journal of Teacher Education*, *26* (4), 363-364. doi.org/10.1177/002248717502600425

SIMIĆ, N., PURIĆ, D. & STANČIĆ, M. (2018). Motivation for the teaching profession: Assessing psychometric properties and factorial validity of the orientation for teaching survey on in-service teachers. *Psihologija*, *51* (3), 309-331. doi.org/10.2298/PSI170327012S

SINCLAIR, C., DOWSON, M. & MCINERNEY, D.M. (2006). Motivations to teach: Psychometric perspectives across the first semester of teacher education. *Teachers College Record, 108* (6), 1132-1154. doi.org/10.1111/j.1467-9620.2006.00688.x

SOPER, D. (2022). *A-priori sample size calculator for structural equation models*. Retrieved (17/03/2023) from https://www.danielsoper.com/statcalc

STUCKY, B.D., THISSEN, D. & EDELEN, M.O. (2013). Using logistic approximations of marginal trace lines to develop short assessments. *Applied Psychological Measurement*, 37, 41-57. doi.org/10.1177/0146621612462759

TÓTH-KIRÁLY, I., MORIN, A.J., BŐTHE, B., OROSZ, G. & RIGÓ, A. (2018). Investigating the multidimensionality of need fulfillment: A bifactor exploratory structural equation modeling representation. *Structural Equation Modeling: A Multidisciplinary Journal*, *25* (2), 267-286. doi.org/10.1080/10705511.2017.1374867

XIAO, Y., LIU, H. & HAU, K.-T. (2019). A comparison of CFA, ESEM, and BSEM in test structure analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *26* (5), 665-677. doi.org/10.1080/10705511.2018.1562928

YOU, S. & CONLEY, S. (2015). Workplace predictors of secondary school teachers' intention to leave: An exploration of career stages. *Educational Management Administration & Leadership*, *43* (4), 561-581. doi.org/10.1177/1741143214535741

APPENDIX

Orientation for Teaching Survey (OTS) Italian version

Ho deciso di insegnare perché…

1. vorrei lavorare con i ragazzi
2. l'insegnamento mi permette di prestare un prezioso servizio di valore morale
3. mi piace stare nell'ambiente scolastico
4. ho/avrò la possibilità di percepire un buon stipendio
5. gli insegnanti godono di buoni benefit associati al loro lavoro
6. mi piacciono gli orari di lavoro e le vacanze scolastiche
7. l'insegnamento mi dà/darà la possibilità di aiutare i meno fortunati
8. l'insegnamento mi dà/darà l'opportunità di aiutare gli studenti ad acquisire un senso di realizzazione e autostima
9. l'insegnamento mi dà/darà la possibilità di "ripagare" i buoni insegnanti che ho avuto
10. i miei genitori ritenevano che l'insegnamento sarebbe stato una buona carriera per me
11. l'insegnamento mi dà/darà l'opportunità di avere autorità
12. l'insegnamento mi permette/permetterà di vivere l'amore e il rispetto dei ragazzi
13. l'insegnamento è un'occupazione relativamente non competitiva
14. ho una passione per una particolare materia
15. non ero soddisfatto/a del lavoro che avevo svolto in altri campi
16. è meno costoso prepararsi per insegnare di quanto non lo sia prepararsi per molti altri campi
17. è un'occupazione intellettualmente stimolante
18. l'insegnamento è un'occupazione appagante e stimolante
19. mi sento più a mio agio con i ragazzi che con gli adulti
20. mi piacerebbe risolvere alcuni dei problemi del sistema educativo
21. mi piace l'idea di essere al centro dell'attenzione in una stanza piena di persone
22. c'è così tanto bisogno di bravi insegnanti
23. l'insegnamento era il lavoro migliore tra quelli più prontamente disponibili per me
24. l'insegnamento è un'occupazione prestigiosa
25. l'insegnamento mi dà/darà la possibilità di essere il capo di me stesso/a
26. amo i ragazzi
27. mi è piaciuto lavorare con i ragazzi in altri contesti e ho pensato che l'insegnamento sarebbe altrettanto piacevole
28. l'insegnamento era il miglior lavoro tra quelli per cui sono più tagliato/a
29. sento una "vocazione" personale all'insegnamento
30. ho il desiderio di impartire conoscenze ad altre persone
31. l'insegnamento mi dà/darà la possibilità di avere un impatto sulla società
32. ho sempre voluto insegnare
33. l'insegnamento è una professione creativa

continued on next page

continued

34. da insegnante, posso avere l'opportunità di svolgere attività extracurriculari che mi piacciono
35. l'orario di lavoro è compatibile con la mia situazione domestica
36. l'insegnamento mi dà/darà la possibilità di migliorare la mia posizione sociale
37. l'insegnamento mi dà/darà la possibilità di fungere da modello positivo per i ragazzi
38. l'insegnamento si adatta bene alla mia personalità
39. insegnare è una tradizione di famiglia
40. le persone spesso mi considerano un insegnante nato
41. l'insegnamento mi dà/darà l'opportunità di promuovere il rispetto per la conoscenza e l'apprendimento
42. alcuni dei miei amici si sono laureati in ambito pedagogico
43. mi sono formato/a in un altro campo ma non sono riuscito/a a trovare un lavoro
44. mi sono formato/a in un altro campo ma non mi sentivo a mio agio in quell'ambito
45. qualcuno che stimo molto mi ha detto che sarei stato/a un buon insegnante
46. mi hanno parlato di una borsa di studio o di un programma di rimborsi delle tasse universitarie disponibile per coloro che intraprendono i percorsi formativi per diventare insegnanti
47. l'insegnamento mi offre/offrirà una buona opportunità per l'avanzamento di carriera
48. l'insegnamento può facilmente portarmi ad altre carriere
49. insegnare può aiutarmi a sviluppare il carattere
50. gli insegnanti vivono un ambiente di lavoro piacevole
51. l'insegnamento mi dà/darà opportunità di leadership
52. è facile formarsi per lavorare come insegnante
53. insegnare mi dà/darà l'opportunità di imparare per tutta la vita
54. l'insegnamento mi dà/darà l'opportunità di interagire con colleghi interessanti
55. l'insegnamento mi dà/darà l'opportunità di incontrare molte persone
56. l'insegnamento mi offre/offrirà un lavoro sicuro
57. l'insegnamento è un lavoro molto facile
58. ho sentito un discorso motivazionale sull'insegnamento oppure sono stato influenzato da materiale mediatico focalizzato sui benefici dell'insegnamento

Note. The items for the short version are: a) implicit motivation: 2, 16, 17, 25, 29, 32, 36, 40, 41; b) explicit motivation: 5, 11, 20, 23, 35, 42, 51, 52, 57.

DOI: 10.26387/bpa.2023.00005

Rapid Visual Information Processing neuropsychological test for remote cognitive assessment in multiple sclerosis: A preliminary study

Daniele Lozzi

Dep. of Information Engineering, Computer Science, and Mathematics, University of L'Aquila, L'Aquila, Italy

daniele.lozzi@graduate.univaq.it

ABSTRACT. L'obiettivo di questo lavoro preliminare è di valutare l'applicabilità del Rapid Visual Information Processing (RVIP) test in full-remote, insieme ai test DASS-21 per depressione ansia e stress, CFQ e MSNQ per i fallimenti cognitivi, al fine di misurare la velocità dell'elaborazione delle informazioni (Information Speed Processing - IPS) che risulta essere deficitaria nei pazienti con SM. I risultati suggeriscono come l'RVIP veicoli molte informazioni utili e che possa essere utilizzato in una valutazione dell'IPS in modalità full-remote. Infine, il modello di regressione logistica, costruito usando il True Positive Rate dell'RVIP, mostra una buona specificità di classificazione dei soggetti nei due gruppi. Sviluppi futuri dovranno considerare un campione più ampio e valutare preliminarmente le competenze digitali.

SUMMARY. Multiple sclerosis (MS) is a neurodegenerative disease that causes many neuro-cognitive deficits. The test batteries used for clinical evaluation are not always able to detect the slight variations since they do not always consider the subject's baseline but only the reference value computed on the population. Furthermore, the low frequency of assessment, also due to the sparse distribution of MS centres makes the test unable to detect slight fluctuations that could indicate the onset of new deficits. Remote neuropsychological evaluation performed with a high frequency, can provide useful information on the subject's baseline, going beyond the limits of population normative values, and become enough sensitive to measure significant variations on a single subject. In line with the European Union (EU) directives on telemedicine, this work aims to evaluate the usefulness of the remote Rapid Visual Information Processing (RVIP) test to assess the Information Speed Processing (IPS) which is impaired in 40% of MS patients. The other already validated tests used in this work were the Depression Anxiety Stress Short Scale (DASS-21) tests for depression, anxiety and stress, the Cognitive Failure Questionnaire (CFQ) and the Multiple Sclerosis Neuropsychological Questionnaire (MSNQ) both for cognitive failures. Data from 44 subjects (12 with MS and 32 healthy) were acquired. The results show that the True Positive Rate (TPR) of the RVIP is lower in the MS group compared to the Healthy Group. The logistic regression model can classify the subjects into two groups with good specificity. This quickly and non-invasive neuropsychological test performed in remotely, allow us to estimate of the most reliable true parameter of IPS and increase the signal-to-noise ratio performing repeated measures with a high frequency. The present work shows that a remote evaluation of some specific neuropsychological domains is possible and together with a traditional assessment, it can support the clinician in adjusting the therapy, reducing costs and the impact on the individual, and improving the outcome.

Keywords: *Cognitive Processing Speed, Telemedicine, Multiple sclerosis, Rapid Visual Information Processing*

INTRODUCTION

Multiple sclerosis (MS) is a neurodegenerative disease that leads to a progressive loss of myelin, affecting central nervous system (CNS) functions, including the cognitive domain (Lassmann, 2018). MS is more common in female people and on the North of the Earth. Actually, is not clear which is the most important risk factor associated with this distribution between solar radiation, environmental factors, pathogens or genetics (Compston & Coles, 2008). The diagnostic process is complex and includes several biomedical, neurological, radiological and biochemical tests. The course of the disease depends on the MS type. It can be remitting-relapsing or progressive, with several subgroups. The remitting-relapsing MS is characterized by periods with symptoms interspersed with periods without symptoms: this trend has a discrete ladder shape. The progressive MS type is generally slower but without any remission and with increasing severity of symptoms: this trend has a more linear shape without negative derivative (Brownlee, Hardy, Fazekas & Miller, 2017).

It is well known that in MS, cognitive impairment could be present 40-70% of total cases (Chiaravalloti & DeLuca, 2008) and for this, neuropsychological assessment is necessary to address the clinician to personalized cognitive, physical and pharmacological therapy (Rao, Leo, Bernardin & Unverzagt, 1991). Furthermore, cognitive impairment in MS people, which leads to a quality of life (QoL) deterioration, is presented with a high variance throughout the population. Principal cognitive functions impaired in people with MS are:

- long-term memory;
- sustained and selective attention;
- problem-solving;
- attention switching;
- information processing speed (IPS).

In addition, it is necessary to assess psychological conditions due to depressive or anxiety disorders onset (Macias & Ciampi, 2019). Many psychological and neurological batteries for cognitive assessment were used in the literature. Mini-Mental State Examination (MMSE) (Tombaugh & McIntyre, 1992) is useful but not sufficiently sensitive, due to the proof effect, because it is devised for elderly people. More common are the *Brief Repeatable Battery of Neuropsychological Test* (*BRB-NT*) (Rao et al., 1991), the *Minimal Assessment of Cognitive Function in MS* (*MACFIMS*) (Benedict et al., 2006) and the *Brief International Cognitive Assessment for Multiple Sclerosis* (*BICAMS*) (Langdon et al., 2012). All of them assess IPS, specifically using *Symbol Digit Modalities Test* (*SDMT*) (Smith, 1973), but its tests not use the response time (RT) with very high precision, useful to catch minimal difference performance between two or more sessions. In general, neuropsychological tests are validated on populations, and for this, they cannot take into account the baseline of the subject (Eckert, Keren, Roberts, Calhoun & Harris, 2010). Despite this, it is crucial to use the derivative of the performance trend acquired over time to detect the slightest difference from the previous session or the prior healthy state before the onset of MS (Schoenberg et al., 2011). The derivative could be useful to identify some prodromal symptoms of cognitive decline onset or worsening symptoms. Additionally, the bias in the data might be a consequence of the hospital environment (similar to the white coat hypertension effect). Finally, the MS centre also does not have a wider distribution throughout the region since it is expensive, difficult to get to the hospital, and difficult for patients to travel throughout the pathology wards, making evaluations even less regular. Through telemedicine, a neuropsychological assessment of important pathology-related aspects might assist the doctor in quickly changing the course of treatment for a better outcome (Ziemssen et al., 2015).

Related works

Neuropsychological assessment traditionally requires an in-person clinical expert to evaluate cognitive functions, but there are some advantages to using a computerized-remote assessment system.

In (Settle, Robinson, Kane, Maloni & Wallin, 2015), the authors used neuropsychological instruments in two modalities: live-in-office (LIO) and remote-in-office (RIO). To assess the IPS, they used *Symbol Digit Modalities Test* (*SDMT*) (Smith, 1982): in this task, the participants have to match the correct number corresponding to a symbol in a correct key looking at the legend. The time to perform this task is measured. The authors also used the *Automated Neuropsychological Assessment Metrics* (*ANAM-MS*) (Reeves, Winter, Bleiberg & Kane, 2007) to assess several cognitive functions: attention, reaction time, information processing speed, memory, and decision-making. Results show no significant difference in ANAM-MS and little difference in

SDMT between the two modalities (LIO and RIO). Moreover, they find that the SDMT score is highly predictive of ANAM-MS scores and this suggests SDMT could be used for a screening assessment. Moreover, based on the SDMT score, clinicians can choose to perform a more detailed assessment.

In (Barcellos et al., 2018), the authors used the *California Verbal Learning Test – second version* (*CVLT-II*) (Delis, Kramer, Kaplan & Ober, 1987), a part of *Brief International Cognitive Assessment for Multiple Sclerosis* battery (*BICAMS*) (Benedict et al., 2012). The CVLT-II assesses verbal memory: clinicians read a list of words five times and participant have to recall as many words as possible. They collect data from 180 people with MS and 90 without MS diagnosis (Healthy group). The difference in the score was statistically significant between the Healthy and MS group, both in-person and in remote modality assessment.

No differences were found in the two modalities between MS assessed in-person and remote.

In (Barcellos et al., 2021), the authors collected data from 100 people with MS, divided into two groups: 50 people to in-person assessment group and 50 people to remote assessment group. All subjects were assessed by CVLT-II and SDMT. The authors observe the strong correlation between the final score of two modalities (r = .85, $p<10^{-28}$ for SDMT and r = .71, $p<10^{-15}$ for CVLT-II). Moreover, shuffling the order of the two modalities does not affect the results.

In (Rogers et al., 2022), authors used SDMT, CVLT-II, *Brief Visuospatial Memory Test-Revised* (*BVMT-R*) (Benedict, 1997), where subjects have to draw figures previously seen, *Trail-Making Test* (*TMT*) (Reitan & Wolfson, 1992) for visual attention and task switching assessment, and *Multiple Sclerosis Neuropsychological Questionnaire* (*MSNQ*) (Benedict et al., 2003) for a self-screening on cognitive impairments. The research was conducted considering two groups both composed of 34 people, one with in-person and remote assessment and one only with remote assessment. The results show no significant differences between the two groups for CVLT-II and SDMT final scores.

In a recent review (Wojcik et al., 2019), the literature about computerized neuropsychological assessment devices (CNADs) in MS was analyzed. The results show how some computerized instrument to assess cognitive impairment in multiple sclerosis could be useful to clinicians due to psychometric qualities although several computerized tests for cognition does not yet demonstrate adequate reliability and validity. Nevertheless, computerized tests could be useful for identifying the prodromal symptoms capturing the very tiny difference in some parameters as RT that is impossible with traditional tests.

Our work

This preliminary work aims to perform a web-based assessment method for cognitive functions to evaluate the utility of a remote assessment of cognitive function in MS people, with a focus on IPS and its relation with some psychological aspects. Our hypothesis is based on the assumption confirmed by literature (Chiaravalloti & DeLuca, 2008; Rao et al., 1991) that MS people have the worst performance in IPS tasks. This work tries to evaluate this cognitive domain in remote modalities during the Covid-19 pandemic period. Due to Covid-19 Italian law, it was not possible to compare in-remote with in-person assessments and also was not possible to explain the usage of the platform used for that study. The *Rapid Visual Information Processing* (*RVIP*) test (Talland, 1966) was used to assess the speed of information processing and the number of correct, wrong or miss answers to an information flow based on specific rules. The RVIP test is already present in *Cambridge Neuropsychological Test Automated Battery* (CANTAB, 2016), a battery used in the evaluation of functions cognitive composed of several tests with automatic scoring. Our result shows how RVIP is a useful neuropsychological instrument to assess IPS in MS patients in a web-based context, and it conveys the same information as the Go/No-Go test (Donders, 1969) with more details on sustained attention and working memory. Moreover, the score of MSNQ (Migliore et al., 2021) was in line with the result of the *Cognitive Failure Questionnaire* (*CFQ*) (Broadbent, Cooper, Fitzgerald & Parkes, 1982). Although the weak sample and the different distribution of age in the two groups that not allow a generalization of results, the logit model showed the relation between the TPR (True Positive Rate) of the RVIP and the MS diagnosis.

METHOD

The IPS is a part of the cognitive domain of complex attention, referring to the storing and manipulation of information for a short time. The RVIP test evaluates the continuous performance on the detection or avoidance of

specific targets. Based on how many targets or non-target are correct detecting (HIT), wrong detecting (FA), missed (MISS) or correct rejecting (CR). RVIP complementary assesses impulsivity using False Positive Rate (*FPR*) index:

$$FPR = \frac{nFA}{nNonTarget}$$

where *nFA* is the total of the wrong answers to non-target sequences/stimulus, and *nNonTarget* is the total of the non-target sequences/stimulus presented.

Before starting, participants have to learn the target sequences. After a very short training, participants observe a stream of single digits (from 0 to 9) appear on the screen and they have to press a button (space bar) as soon as they identify the last digit of the target sequence, consisting of three digits. As shown in Table 1, the indexes used to assess the results are:

- HIT: correct answers to a target sequence over total target stimuli;
- False alarm (FA): wrong answer to a non-target sequence over the number of total non-target stimuli;
- MISS: not-answer to a target sequence over total target stimuli;
- Correct rejection (CR): correct rejected answers to a non-target sequences/stimulus.

This paradigm as a cognitive task was used as a measure of sustained attention began in the 1960s to estimate the effect of alcohol intake (Talland, 1966).

Rapid Visual Information Processing

The test used in the present work is a simplified version of the RVIP (Jones, Sahakian, Levy, Warburton & Gray, 1992). This version solves some ambiguities due to the possible succession of two consequential target sequences. The original target sequences proposed were the following: 2-4-6, 3-5-7, 4-6-8, 5-7-9.

This test was used to evaluate sustained attention in visual sensory modality, following the subcutaneous administration of nicotine to subjects with Alzheimer's dementia and is also present in the CANTAB battery. The main rule for this task is to detect the target sequence by pressing the button only when the last (third) digit appears: the first two digits can alert the participant to be ready for pressing the button. The version used in this work uses only two target sequences: 2-4-6, 3-5-7.

This avoids the improbable but possible cases of two target sequences appearing one after another. Instead, if two concatenated sequences should appear, the participant could not understand which triplet is the target and lose the last target sequence and for this reason, two target sequences were excluded. Each stimulus lasts 400 milliseconds (ms) followed by an interval between stimuli (ITI - inter-trial interval) of variable ms, between 300 and 600 ms, to prevent the subject from habituation to the distance between two stimuli. Thus, it can be possible to stimulate sustained attention and max-

Table 1 – Possibile answers

	Target	**Non-Target**
Answer	*HIT*	*FA*
Non-Answer	*MISS*	*CR*

Legenda. HIT = correct detecting; FA = wrong detecting; MISS = missed; CR = correct rejecting.

imise the probability of FA occurrences. Through the RVIP is possible to analyze many scores: in addition to the four parameters that can also be calculated in the Go /No-Go to verify if the sustained attention (Hit, Miss, False alarm, Correct rejection) it is possible to assess the speed information processing through the analysis of these parameters over time because is necessary to store the target sequences.

Self-assessment questionnaires

Data from questionnaires were collected. The questionnaires used were the following.

- *Depression Anxiety Stress Scales short version* (Lovibond & Lovibond, 1995): the DASS-21 is a tool, already validated on the Italian population (Bottesi et al., 2015), useful for discriminating against depressive from anxious symptoms, and whose principal component analysis showed that 1/3 of the items refer to a third factor that has been identified as a measure of stress (characterized by irritability, nervous tension, difficulty relaxing and agitation). The original article, using the extended version with 42 items, analyzed the psychometric properties and identified a strong correlation between the depression component and the *Beck Depression Inventory* (*BDI*; Beck, Steer & Brown, 1996) score. Moreover, it was shown the correlation between the anxiety component and the *Beck Anxiety Inventory* (*BAI*; Beck, Epstein, Brown & Steer, 1993) score, a very common tool used in clinical practice. The scale, in addition to measuring the three components (depression, anxiety and stress), can discriminate very well between depression and anxiety, maximizing the internal consistency between items of the same construct. The short version (Henry & Crawford, 2005), the DASS-21, was created to have a smaller instrument than the original 42-item version, keeping the psychometric components within adequate limits to be used in a clinical context. Therefore, DASS-21 has good construct validity, and factor analysis identified that although the items refer to a general factor of psychological discomfort, a part of the variance of each of them is explained by the three factors (Depression – Anxiety – Stress), and shown high reliability. Compared to the extended version of 42 items, the DASS-21 has the advantage of being able to be administered to people with poor attention sustained. The advantage of using DASS-21 compared to DASS-42 is to improve the discrimination power between a score due to psychological disorder instead of neurological damage.
- *Cognitive Failures Questionnaire*: the CFQ is a tool used in the self-assessment context of cognitive functions which, through 25 questions, investigates the errors of memory, perception or distractions that can occur in daily life. Operationally, it investigates errors by asking how often some common cognitive failures occur (e.g. "Do you fail to notice signposts on the road?"). This tool was used for the validation of the MSNQ, showing a high correlation between the two scores. Furthermore, it has been included in the present work since it has already been validated on the Italian population (Stratta, Rinaldi, Daneluzzo & Rossi, 2006), where the multi-factorial nature of the measured constructs was confirmed to measure: memory, concentration, inattention/distractibility, interpersonal intelligence, memory for names.

  The validation study of CFQ showed a high Cronbach's alpha, as well as no relationship between test score and age, while there was a slight but significant correlation between total test score and years of education. However, the validation study has a group of university students as a sample, and if this eliminates disturbing factors of a too-heterogeneous sample with many variables, it does not guarantee an extension of the results to the whole Italian population, which is noted by the authors themselves who recommend carrying out more accurate analyzes on more numerous and heterogeneous samples. Despite this, they assume that years of education could be interpreted as an operational measure of intelligence and therefore the results could be a "reflection of the cognitive variables related to intellectual performance".
- *Multiple Sclerosis Neuropsychological Questionnaire* (*Patient and Informant* sections): the *Multiple Sclerosis Neuropsychological Questionnaire* (*MSNQ*) is a short questionnaire composed of 30 self-administered questions: 15 are for the person with MS (patient, MSNQ-P) and 15 are for the informant (MSNQ-I, i.e. a person who knows the patient and his daily routine), designed precisely to be administered quickly by non-specialist staff. The participants have to rate each question on a Likert scale, from "Problem not encountered" (0) to "Occurs very frequently" (4). A higher score indicates more cognitive deficits. It was built based on an analysis of the scientific literature on the existing tools and on the clinical experience of the authors to capture all the

variations in cognitive functions that occur in people with MS. The items referring to depressive symptoms were excluded, although the initial intention was to combine some items referring to depressive disorders (such as those of the BDI) with questions on cognitive failures. Originally, self-administered questionnaires were found to predict *Expanded Disability Status Scale* (*EDSS*) (Kurtzke, 1983) scores sufficiently, but for assessment of cognition, they may not be reliable because memory impairment is strongly associated with depressive symptoms. People with dementia tend to overestimate their cognitive abilities, and for this, an association of information collected by the patient together with the information from the informant could increase the predictivity of the results. From an analysis of the collected data on MSNQ (O'Brien et al., 2007), it was possible to observe that MSNQ-P is very useful for estimating depressive symptoms while MSNQ-I is useful for effectively estimating cognitive functioning in daily activity as a screening measure, due to difficulty in self-estimate cognitive impairments. Finally, the dissociation between the responses of the patient and those of the informant may be an indication of dementia, with implications of anosognosia and reduced self-awareness. This instrument is recently validated on the Italian population (Migliore et al., 2021).

Data acquisition and descriptive statistics

Data collection was done using the Psytoolkit platform (Stoet, 2010, 2017), and analyzed using Phyton programming language (Van Rossum & Drake Jr, 1995) and Jamovi software (Jamovi Project, 2021). In Figure 1 and in Figure 2 are shown some screenshots of the website built for the data acquisition.

Data from 44 subjects were collected, 12 and 32 from the MS and Healthy group, respectively, during the period between November and December 2020, during the second lockdown in Italy, through the support of associations and online communities due to the difficulty to recruit people within a hospital or clinical centre, both for Healthy and MS groups. In particular, through the Association "Sclerosi Multipla Albero di Kos" it was possible to recruit patients among their subscribers for the MS group, and the Healthy group it was published online a public notice specifying the exclusion criteria. Also due to Italian law, it was not possible to perform study design with one group in presence and one remote to make the comparison. The inclusion criteria were to have had an MS diagnosis and for Healthy groups was not to have had an MS diagnosis. Moreover, the exclusion criteria for both groups were related to not having had other diagnoses of psychiatric and/or neurological diseases or previous neurological damage. Figure 3 shows the age distribution of the sample dividend into gender groups.

In particular, the TPR was used as a metric of performance:

$$TPR = \frac{nHIT}{nTarget}$$

where *nHIT* is the total number of the correct answer (HIT) to a target sequence/stimulus, and *nTarget* is the total of the target sequence/stimulus.

RESULTS

Due to the low sample size, it was impossible to perform separate analyses based on different types of MS.

Cognitive performance

The results showed in Table 2 show a significant difference of *TPR RVIP* between the MS and Healthy group, where the first group had better performance. The *t*-test was used to verify if the difference was statistically significant (on the last line of Table 2) and the same result was achieved for the Go/No-Go test (TPR Go/No-Go, see Table 3). The *t*-test was used to verify if the difference was statistically significant as shown in Table 3 on the last line.

Despite this, RT does not differ between the two groups. The results in Table 4 of the questionnaires show a strong correlation between the total scores of the CFQ and the MSNQ-P, which confirms the construct validity of the two diagnostic tools.

Unfortunately, the number of MSNQ-I answers does not allow any statistical analysis. An interesting finding shown in Table 5 is a negative correlation between the TPR of the Go/No-Go test of the Healthy group and the score of the CFQ inattention subscale and this shows the high sensitivity for Go/No-Go to measure the inattention domain. The correlation between the TRP of Go/No-Go and the CFQ inattention score in the MS group was not statistically

Figure 1 – Screenshot of RVIP test

SECONDA PARTE

Vedrete comparire in rapida successione una serie di numeri singoli tra zero e nove (0-9). Prestando la massima attenzione, se dovesse comparire, in questo esatto ordine, la sequenza di numeri

2 – 4 – 6

oppure

3 – 5 – 7

va premuta la BARRA SPAZIATRICE al momento della comparsa dell'ultima cifra della sequenza (quindi alla comparsa del 6 o del 7)

Figure 2 – Screenshot of DASS-21

Item	Molto	Abbastanza	Occasionalmente	Raramente	Mai
Si distrae facilmente?	◯	◯	◯	◯	◯
Perde il filo del discorso mentre ascolta qualcuno che parla?	◯	◯	◯	◯	◯
È rallentato quando prova a risolvere problemi?	◯	◯	◯	◯	◯
Dimentica gli appuntamenti?	◯	◯	◯	◯	◯
Dimentica cosa ha letto?	◯	◯	◯	◯	◯
Ha difficoltà a descrivere programmi o varietà televisivi che ha recentemente visto?	◯	◯	◯	◯	◯
Le istruzioni le devono essere ripetute?	◯	◯	◯	◯	◯
Le devono essere ricordate le cose da fare?	◯	◯	◯	◯	◯
Dimentica gli appuntamenti che ha pianificato?	◯	◯	◯	◯	◯
Ha difficoltà a rispondere alle domande?	◯	◯	◯	◯	◯
Ha difficoltà a fare due cose nello stesso momento?	◯	◯	◯	◯	◯
Dimentica cosa una persona le sta dicendo?	◯	◯	◯	◯	◯
Ha difficoltà a controllare gli impulsi?	◯	◯	◯	◯	◯
Ride o piange per futili motivi?	◯	◯	◯	◯	◯
Parla eccessivamente o si focalizza molto su un suo interesse personale?	◯	◯	◯	◯	◯

Figure 3 – Age average by group divided by gender

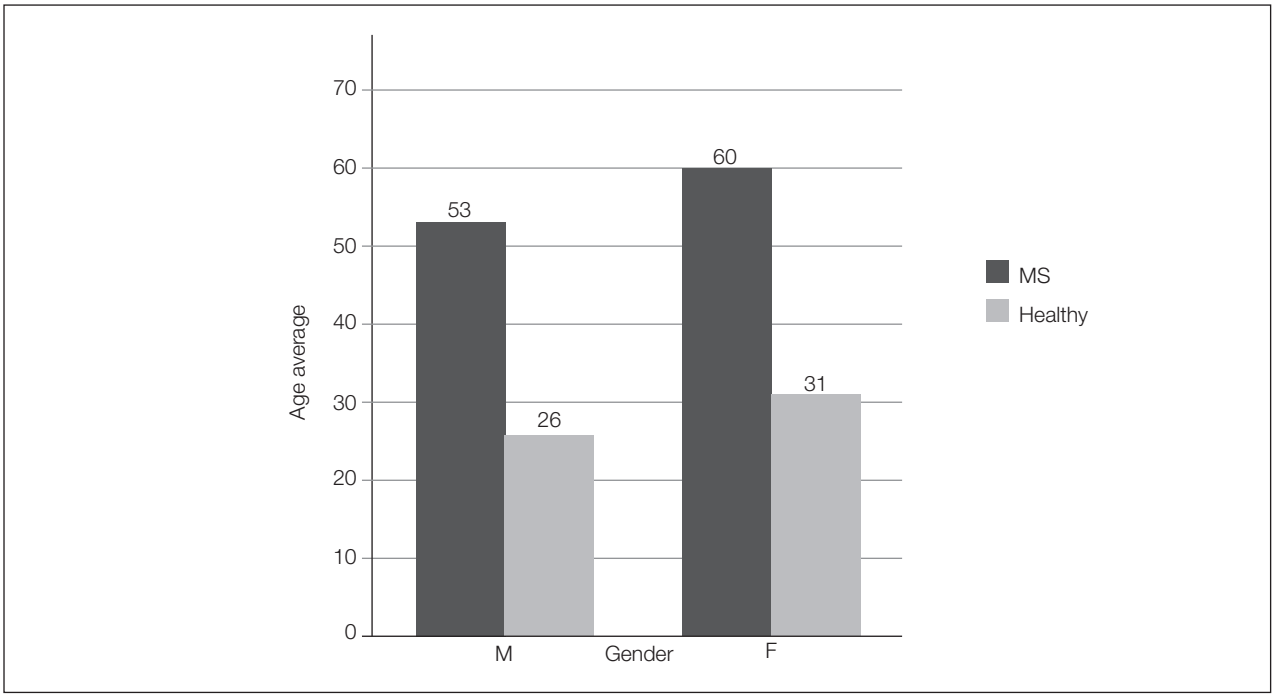


Table 2 – TPR RVIP over group

	MS group±*SD*	**Healthy group±*SD***
TPR RVIP	.36±.29	.67±.30
n	12	32
***t*-test**	***p*-value**	***df***
3.03	$p<.005$	42

Legenda. TPR = True Positive Rate; RVIP = *Rapid Visual Information Processing; df* = degree of freedom.

Table 3 – Go/No-Go RVIP over group

	MS group±*SD*	**Healthy group±*SD***
TPR Go/No-Go	.34±.25	.8±.21
n	12	32
***t*-test**	***p*-value**	***df***
6.218	*p*<.001	42

Legenda. df = degree of freedom.

Table 4 – Correlation coefficients between MSNQ-patients and CFQ

	MS group	**Healthy group**
Corr. (MSNQ-P – CFQ)	.86	.77
p-value	*p*<.001	*p*<.001

Legenda. MSNQ = *Multiple Sclerosis Neuropsychological Questionnaire;* CFQ = *Cognitive Failure Questionnaire.*

Table 5 – Correlation coefficients between TRP Go/No-Go and CFQ inattention

	Healthy group	***p*-value**
Corr. (TPR Go/No-Go – CFQ inatt.)	−.52	*p*<.005
Corr. RT avg HIT (RVIP – Go/No-Go)	.326	*p*<.05

Legenda. TPR = True Positive Rate; CFQ = *Cognitive Failure Questionnaire*; RT = response time; HIT = correct detecting; RVIP = *Rapid Visual Information Processing*.

significant, probably due to the low number of samples.

Furthermore, a positive correlation shown in Table 5 was observed between the results of the RT HIT mean between the Go/No-Go and the RVIP test, both in the MS and Healthy groups. This suggests that the use of the RVIP maintains the information acquired by the Go/No-Go and adds the information relating to the number of sequences detected which shows correct retention of the information in memory and a sufficient IPS. This result shows how Go/No-Go can be replaced by RVIP to avoid information redundancy by acquiring more detailed data.

Depression – Anxiety – Stress

Results on DASS-21 (Table 6 for mean and Table 7 for *t*-test results) show that scores do not show any statistically significant difference between the two groups (MS, Healthy). The result may support the hypothesis that IPS is lower in the MS group due to the disease (Hauser & Oksenberg, 2006).

Interaction between cognitive performance, depression and MSNQ(P-I) scores

As said before, self-perception of cognitive decline could be masked by cognitive decline itself and accepted as depression. Unless the low sample size, the result shows that MSNQ-P score is more correlated with depression compared to MSNQ-I score. The results are merely indicative (see Table 8) considering the high *p*-value probably due to the low sample size.

Table 6 – Average scores on DASS-21 dimensions

	Depression±*SD*	**Anxiety±*SD***	**Stress±*SD***
Healthy	14.9±10.5	9.06±7.38	20.3±9.87
MS	14.2±12.7	7.67±9.14	14.3±12.1

Table 7 – *t*-test on DASS-21 dimensions

***t*-test Depression**	***p*-value**	***df***
.161	*p*>.05	42
***t*-test Anxiety**	***p*-value**	***df***
.523	*p*>.05	42
***t*-test Stress**	***p*-value**	***df***
1.681	*p*>.05	42

Legenda. df = degree of freedom.

Table 8 – Correlation between MSNQ scores and Depression score

	Depression (DASS-21)	*p*-value
MSNQ-P score	.598	*p*>.05
MSNQ-I score	.362	*p*>.05

Legenda. MSNQ = *Multiple Sclerosis Neuropsychological Questionnaire.*

Conversely, a higher correlation between performance at cognitive tests (RT and HIT, FA, CR, MISS for Go/No-Go and RVIP) with MSNQ-I compared to MSNQ-P scores are expected, but no differences were found in this data.

Logistic regression

The logit model was built using as an independent variable the MS diagnosis (0 = Healthy Group, 1 = MS Group) and as a dependent variable the TPR RVIP. The result shows that the model can detect a Healthy person from MS person with a .844 specificity score and a .705 precision score, as shown in Table 10, Table 11 and the Area Under the Curve (AUC) of the model in Figure 4. Figure 5 it is shown the probability of an MS diagnosis given the empirical value (round black dot), and the theoretical curve (black line). Parameters shown in Table 9 and the general equation of the logit model are the following:

$$P(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

After training the model, the parameters estimate were shown in Table 9 (in this model Healthy group is the reference). As a consequence, the real equation after replacement of the β_0 and β_1 terms becomes:

$$P(x) = \frac{1}{1 + e^{-(0{,}6 - 3{,}03x)}}$$

CONCLUSION

This preliminary work shows how a web-based assessment of some cognitive domains is possible, which could be useful if carried out constantly and with a higher frequency. The combination of remote assessments with in-person assessments can help the clinician to understand the progression of the disease and adapt the therapy. Moreover, continuous data acquisition can be useful to link data of visible injuries acquired with medical imaging techniques with cognitive performance (Sperling et al., 2001). Furthermore, a remote evaluation can be useful to reduce any bias due to the clinical environment and reduce the stress due to travel in people with motor disabilities. Nevertheless, the domestic environment should be checked before the assessment to guarantee the absence of distracting stimuli, as well as internet connection stability and the quality of the

Table 9 – The estimated parameters of logit model

Predictor	Estimate	*p*-value
Intercept (β_0)	.6	.326
Beta (β_1)	−3.03	.009

Table 10 – Confusion matrix of logit model

Classification Table				
		Predict		
		Healthy	**MS**	
Observ	**Healthy**	27	5	84.4%
	MS	8	4	33.3%

Table 11 – Metrics of the logit model

Precision	Specificity	Sensitivity	AUC
.705	.844	.364	.795

Legenda. AUC = Area Under the Curve.

Figure 4 – AUC curve plot

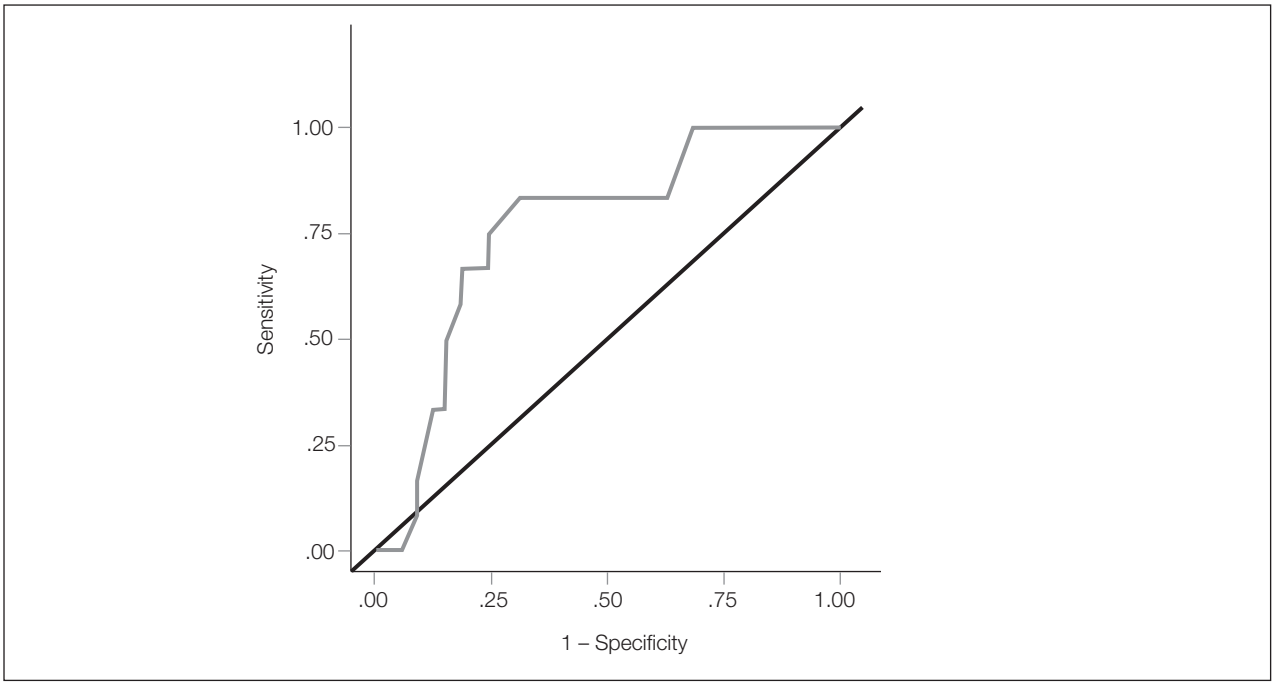
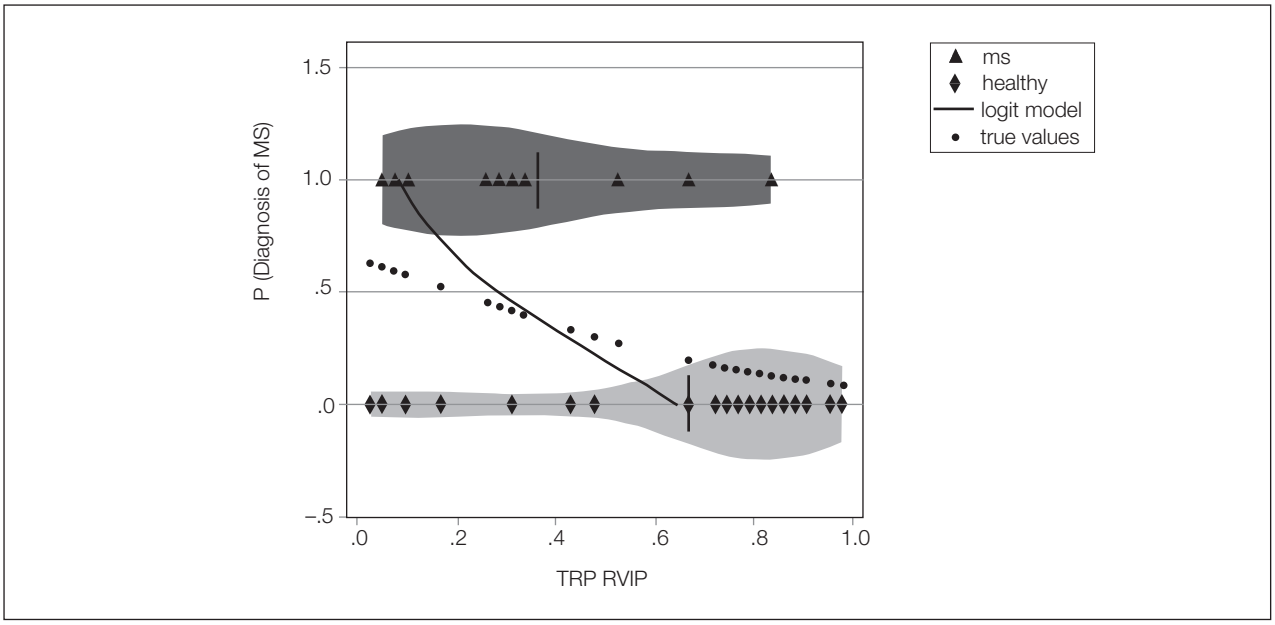


Figure 5 – Logit model on violin distribution



Legenda. Grey and light grey dots represent the empirical value of TRP RVIP with probability equal to 1 or 0; round black dots represent the probability of empirical value after computing the probability with Equation 4; the black line represents the model based on Equation 4.

device, as shown in (Rogers et al., 2022). The data shown are promising even if affected by the low sample size and the high age difference that can affect the result due to physiological cognitive deterioration (Eckert et al., 2010), as well as how the gender could influence the results (Roivainen, 2011). It could be interesting to build a mediation/moderation model to better understand the interaction between age, gender, cognitive performance and illness presence/absence. Another limitation of this work was the absence of a measurable difference between the in-person and remote assessment of the same group, due to Italian law on Covid-19. Furthermore, future studies will include the assessment of digital skills, also including questionnaires that investigate the propensity of patients to be part of telemedicine pathway (Toscano et al., 2022). Moreover, this type of assessment can be framed in telemonitoring interaction, integrating traditional health-care services (Gallo et al., 2022). This is needed to exclude some interactions and confounding factors between age, gender, digital skills and cognitive impairments due to MS. In addition, a future remote assessment will include an evaluation of the usability of the platform, to exclude some difficulties due to technical problems. Moreover, future research should acquire several measures on the same subject performing a within study to analyze the trend of performance in function of time going beyond of normative population and compute better the IPS parameters increasing the signal-to-noise ratio. Future studies should consider larger samples to validate these conclusions and, using a computer vision algorithm for image segmentation (Placidi, Cinque, Polsinelli, Splendiani & Tommasino, 2019), link the performance of cognitive measurement to the newest plaques and their morphological and topographical proprieties, to improve knowledge of cognitive functions and the cognitive correlates of MS. Moreover, it could be useful to reduce the invasive exams needs to understand the course of the disease. Finally, it could be useful to perform both between and within study design, to find personal variation patterns (over her/his baseline) on some specific cognitive domain. The within part of the research could be useful to estimate the self-perception of the cognitive impairments, with caregiver information.

Acknowledgments. This preliminary work was not possible without the support of the Association "Sclerosi Multipla Albero di Kos" (https: //ccsvi-sm.org/) and many online communities.

References

BARCELLOS, L.F., BELLESIS, K.H., SHEN, L., SHAO, X., CHINN, T., FRNDAK, S., DRAKE, A., BAKSHI, N., MARCUS, J., SCHAEFER, C. et al. (2018). Remote assessment of verbal memory in ms patients using the california verbal learning test. *Multiple Sclerosis Journal*, *24* (3), 354–357.

BARCELLOS, L.F., HORTON, M., SHAO, X., BELLESIS, K.H., CHINN, T., WAUBANT, E., BAKSHI, N., MARCUS, J., BENEDICT, R.H. & SCHAEFER, C. (2021). A validation study for remote testing of cognitive function in multiple sclerosis. *Multiple Sclerosis Journal*, *27* (5), 795–798.

BECK, A.T., EPSTEIN, N., BROWN, G. & STEER, R. (1993). Beck anxiety inventory. *Journal of Consulting and Clinical Psychology*.

BECK, A.T., STEER, R.A. & BROWN, G. (1996). Beck Depression Inventory – II. *Psychological assessment*.

BENEDICT, R.H. (1997). *Brief visuospatial memory test – revised*.

BENEDICT, R.H., AMATO, M.P., BORINGA, J., BROCHET, B., FOLEY, F., FREDRIKSON, S., HAMALAINEN, P., HARTUNG, H., KRUPP, L., PENNER, I. et al. (2012). Brief international cognitive assessment for ms (bicams): International standards for validation. *BMC Neurology*, *12* (1), 1–7.

BENEDICT, R.H., COOKFAIR, D., GAVETT, R., GUNTHER, M., MUNSCHAUER, F., GARG, N. & WEINSTOCK-GUTTMAN, B. (2006). Validity of the minimal assessment of cognitive function in multiple sclerosis (macfims). *Journal of the International Neuropsychological Society*, *12* (4), 549–558.

BENEDICT, R.H., MUNSCHAUER, F., LINN, R., MILLER, C., MURPHY, E., FOLEY, F. & JACOBS, L. (2003). Screening for multiple sclerosis cognitive impairment using a self- administered 15-item questionnaire. *Multiple Sclerosis Journal*, *9* (1), 95–101.

BOTTESI, G., GHISI, G., MARTA, CONFORTI, E., MELLI, G. & SICA, C. (2015). The Italian version of the Depression Anxiety Stress Scales – 21: Factor structure and psychometric properties on community and clinical samples. *Comprehensive Psychiatry*, *60*, 170–181.

BROADBENT, D.E., COOPER, P.F., FITZGERALD, P. & PARKES, K.R. (1982). The Cognitive Failures Questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology*, *21* (1), 1–16.

BROWNLEE, W.J., HARDY, T A., FAZEKAS, F. & MILLER, D.H. (2017). Diagnosis of multiple sclerosis: Progress and challenges. *The Lancet*, *389* (10076), 1336–1346.

CANTAB, C.C. (2016). *Cognitive assessment software*. Cambridge Cognition: Cambridge, UK.

CHIARAVALLOTI, N.D. & DeLUCA, J. (2008). Cognitive impairment in multiple sclerosis. *The Lancet Neurology*, *7* (12), 1139–1151.

COMPSTON, A. & COLES, A. (2008). Multiple sclerosis. *The Lancet*, (372), 1502–1517.

DELIS, D.C., KRAMER, J.H., KAPLAN, E. & OBER, B.A. (1987). California Verbal Learning Test. *Assessment*.

DONDERS, F.C. (1969). On the speed of mental processes. *Acta Psychologica*, *30*, 412–431.

ECKERT, M.A., KEREN, N.I., ROBERTS, D.R., CALHOUN, V.D. & HARRIS, K.C. (2010). Age-related changes in processing speed: Unique contributions of cerebellar and prefrontal cortex. *Frontiers in Human Neuroscience*, *10*.

GALLO, P., BATTAGLIA, M., AMATO, M., BERTOLOTTO, A., FILIPPI, M., GASPERINI, C., GHIRARDI, A., PATTI, F., SERASCHI, C., TEDESCHI, G., TOZZI, V., TROJANO, M., URSILLO, P., ABBRUZZESE, M., COSENTINO, M. & MARVULLI, M. (a cura di) (2022). *Agenas promuove PDTA per la sclerosi multipla: Indicazioni per la creazione delle reti di assistenza*. Agenzia Nazionale per i Servizi Sanitari Regionali.

HAUSER, S.L. & OKSENBERG, J.R. (2006). The neurobiology of multiple sclerosis: Genes, inflammation, and neurodegeneration. *Neuron*, *52* (1), 61–76.

HENRY, J.D. & CRAWFORD, J.R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non- clinical sample. *British Journal of Clinical Psychology*, *44* (2), 227–239.

JAMOVI PROJECT, T. (2021). *Jamovi (version 1.6)*. Computer Software.

JONES, G., SAHAKIAN, B., LEVY, R., WARBURTON, D.M. & GRAY, J. (1992). Effects of acute subcutaneous nicotine on attention, information processing and short-term memory in Alzheimer's disease. *Psychopharmacology*, *108* (4), 485–494.

KURTZKE, J.F. (1983). Rating neurologic impairment in multiple sclerosis: An Expanded Disability Status Scale (EDSS). *Neurology*, *33* (11), 1444–1444.

LANGDON, D., AMATO, M., BORINGA, J., BROCHET, B., FOLEY, F., FREDRIKSON, S., HÄMÄLÄINEN, P., HARTUNG, H., KRUPP, L., PENNER, I. et al. (2012). Recommendations for a brief international cognitive assessment for multiple sclerosis (bicams). *Multiple Sclerosis Journal*, *18* (6), 891–898.

LASSMANN, H. (2018). Multiple sclerosis pathology. *Cold Spring Harbor Perspectives in Medicine*, *8* (3), a028936.

LOVIBOND, P.F. & LOVIBOND, S.H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, *33* (3), 335–343.

MACIAS ISLAS, M.A. & CIAMPI, E. (2019). Assessment and impact of cognitive impairment in multiple sclerosis: An overview. *Biomedicines*, *7* (1), 22.

MIGLIORE, S., LANDI, D., PROIETTI, F., D'AURIZIO, G., SQUITIERI, F., MATALUNI, G., NICOLETTI, C.G., CURCIO, G. & MARFIA, G.A. (2021). Validity of the Italian multiple sclerosis neuropsychological screening questionnaire. *Neurological Sciences*, *42* (11), 4583–4589.

O'BRIEN, A., GAUDINO-GOERING, E., SHAWARYN, M., KOMAROFF, E., MOORE, N.B., & DeLUCA, J. (2007). Relationship of the Multiple Sclerosis Neuropsychological Questionnaire (MSNQ) to functional, emotional, and neuropsychological outcomes. *Archives of Clinical Neuropsychology*, *22* (8), 933–948.

PLACIDI, G., CINQUE, L., POLSINELLI, M., SPLENDIANI, A. & TOMMASINO, E. (2019). Automatic framework for multiple sclerosis follow-up by magnetic resonance imaging for reducing contrast agents. *International Conference on Image Analysis and Processing*, 367–378.

RAO, S.M., LEO, G.J., BERNARDIN, L. & UNVERZAGT, F. (1991). Cognitive dysfunction in multiple sclerosis.: I. frequency, patterns, and prediction. *Neurology*, *41*(5), 685–691.

REEVES, D.L., WINTER, K.P., BLEIBERG, J. & KANE, R.L. (2007). Anam® genogram: Historical perspectives, description, and current endeavors. *Archives of Clinical Neuropsychology*, *22* (Suppl 1), S15–S37.

REITAN, R.M. & WOLFSON, D. (1992). *Neuropsychological evaluation of older children*. Neuropsychology Press.

ROGERS, F., BANE, E., DWYER, C.P., ALVAREZ-IGLESIAS, A., JOYCE, R.A. & HYNES, S.M. (2022). Remote administration of bicams measures and the trail-making test to assess cognitive impairment in multiple sclerosis. *Neuropsychological Rehabilitation*, 1–24.

ROIVAINEN, E. (2011). Gender differences in processing speed: A review of recent research. *Learning and Individual differences*, *21* (2), 145–149.

SCHOENBERG, M., LANGE, R., MARSH, P., SAKLOFSKE, D., KREUTZER, J., DELUCA, J. & CAPLAN, B. (2011). Premorbid intelligence. *Encyclopedia of Clinical Neuropsychology*, 2004–2010.

SETTLE, J.R., ROBINSON, S.A., KANE, R., MALONI, H.W. & WALLIN, M.T. (2015). Remote cognitive assessments for patients with multiple sclerosis: A feasibility study. *Multiple Sclerosis Journal*, *21* (8), 1072–1079.

SMITH, A. (1973). *Symbol Digit Modalities Test*. Los Angeles: Western Psychological Services.

SMITH, A. (1982). *Symbol Digit Modalities Test (SDMT). Manual (revised)*. Los Angeles: Western Psychological Services.

SPERLING, R.A., GUTTMANN, C.R., HOHOL, M.J., WARFIELD, S.K., JAKAB, M., PARENTE, M., DIAMOND, E. L., DAFFNER, K.R., OLEK, M.J., ORAV, E.J. et al. (2001). Regional magnetic resonance imaging lesion burden and cognitive function in multiple sclerosis: A longitudinal study. *Archives of Neurology*, *58* (1), 115–121.

STOET, G. (2010). Psytoolkit: A software package for programming psychological experiments using linux. *Behavior Research Methods*, *42* (4), 1096–1104.

STOET, G. (2017). Psytoolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44* (1), 24–31.

STRATTA, P., RINALDI, O., DANELUZZO, E. & ROSSI, A. (2006). The use of the Italian version of the Cognitive Failures Questionnaire (CFQ) in a sample of undergraduate students: A validation study. *Rivista di Psichiatria*, *41* (4), 260.

TALLAND, G.A. (1966). Effects of alcohol on performance in continuous attention tasks. *Psychosomatic Medicine*, *28* (4), 596–604.

TOMBAUGH, T.N. & McINTYRE, N.J. (1992). The mini-mental state examination: A comprehensive review. *Journal of the American Geriatrics Society*, *40* (9), 922–935.

TOSCANO, S., PATTI, F., CHISARI, C.G., ARENA, S., FINOCCHIARO, C., SCHILLACI, C.E. & ZAPPIA, M. (2022). Reliability of televisits for patients with mild relapsing–remitting multiple sclerosis in the covid-19 era. *Neurological Sciences*, *43* (4), 2253–2261.

VAN ROSSUM, G. & DRAKE Jr, F.L. (1995). *Python reference manual*. Amsterdam: Centrum voor Wiskunde en Informatica.

WOJCIK, C.M., BEIER, M., COSTELLO, K., DELUCA, J., FEINSTEIN, A., GOVEROVER, Y., GUDESBLATT, M., JAWORSKI III, M., KALB, R., KOSTICH, L. et al. (2019). Computerized neuropsychological assessment devices in multiple sclerosis: A systematic review. *Multiple Sclerosis Journal*, *25* (14), 1848–1869.

ZIEMSSEN, T., DE STEFANO, N., SORMANI, M.P., VAN WIJMEERSCH, B., WIENDL, H. & KIESEIER, B.C. (2015). Optimizing therapy early in multiple sclerosis: An evidence-based view. *Multiple sclerosis and related disorders*, *4* (5), 460–469.

DOI: 10.26387/bpa.2023.00006

Assessment by telematic means and artificial agents: A new challenge for psychometrics?

Daniela Conti[1,2], Santo Di Nuovo[3]

[1] Department of Humanities, University of Catania, Catania, Italy
[2] Oasi Research Institute-IRCCS, Troina (EN), Italy
3 Department of Educational Sciences, University of Catania, Catania, Italy

daniela.conti@unict.it

ABSTRACT. L'utilizzo di strumenti psicodiagnostici somministrati elettronicamente e/o con agenti artificiali ha aperto una nuova sfida. A un campione di 122 psicologi abilitati è stato somministrato uno strumento che comprendeva un questionario composto da 21 items su una scala a 5 punti e un Differenziale semantico. L'atteggiamento generale degli psicologi nei confronti delle modalità innovative di valutazione è risultato positivo. Dall'analisi dei dati emerge come la somministrazione di test psicodiagnostici per via elettronica o con agenti artificiali richiede studi di adattamento e in molti casi una riformulazione degli strumenti e una adeguata formazione dei professionisti.

SUMMARY. During the COVID-19 restrictions, the administration of psychodiagnostic tools not through direct interaction with the psychologist, but administered electronically and/or with artificial agents, opened a new challenge. A change of setting and interactive conditions are introduced which can alter the reliability and validity of tools consolidated for diagnostic use in face-to-face assessments. A sample of 122 licensed psychologists participated in the study. They were either attending or teaching post-graduate specialization courses of different theoretical-methodological focus. The participants were given an online survey via Google forms that included a questionnaire composed of 21 items on a 5-point scale, built according to the UTAUT model and adapted for the acceptability and willingness to use online questionnaires for mental well-being, and a semantic differential for evaluating the attitude towards technology in general. The overall attitude of the psychologists toward the innovative modalities of assessment is positive, and the intention to use online testing is very high. The predictors of the positive attitude and intention of use are analyzed. A multidimensional analysis suggested that the attitude towards online testing, and the intention to use it, are located in the crossing dimensions of technical (psychometric) aspects and concrete usability. The administration of tests electronically or through artificial agents requires adaptation studies and in many cases a reformulation of the tools that are offered using these modalities. The need to raise awareness amongst psychologists about the new forms of assessment, and to train those who intend to use them, has be underlined.

Keywords: *Online assessment, Psychometry, Diagnostic process, Artificial agents*

INTRODUCTION

The recent pandemic has prompted the generalization of a practice that was already spreading, albeit amidst many doubts and contrasts: the administration of psychodiagnostic tools not through direct interaction with the psychologist, but administered electronically and/or with artificial agents.

This methodological scenario is very different from the use of technologies in the administration and evaluation of psychological tests.

The well-known *Computerized Adaptive Testing* allows tailoring of the instruments, and automating their adaptation to the subject's skill starting level, thus making tests more appropriate and precise, especially for extreme cases (Wainer, Dorans, Flaugher, Green & Mislevy, 2000).

Artificial intelligence systems facilitate complex diagnoses. As a recent example, Grazioli and colleagues (2022) developed supervised machine-learning algorithms to support the diagnostic process for attention deficit hyperactivity disorder (ADHD) in children. A decision tree classifier and random forest algorithms were used to identify the most relevant features in discriminating between the presence or absence of the ADHD diagnosis; they reliably aid the diagnostic process as an initial step.

However, in these cases the psychologist still administers the tests, while the conditions of adaptation to the user and the methods of scoring or interpretating the results have changed, now being completely automated.

When providing for remote administration, a change of setting and interactive conditions are introduced which can alter the reliability and validity of tools consolidated for diagnostic use in face-to-face assessments. Online versions of tests may not always measure the same constructs as their traditional antecedents, therefore equivalence cannot be assumed (Buchanan, 2002). The psychologists themselves may not adequately realize this difference.

Even the use of artificial agents (such as humanoid robots) for administering tests, which has been growing in recent years thanks to the advances in artificial intelligence systems based on Recursive Neural Networks, poses technical problems. These concern the psychometric properties of reliability and validity, and also problems of usability and acceptability of the technological modalities by both professionals and users. In fact, even when the relationship remains remote, and is not limited to the administration of the questionnaire online or in asynchronous mode, the remote modality cannot fully replace the interactive aspects, both verbal and non-verbal, which are relevant for the diagnosis in face-to-face situations.

We will summarize the results of research aimed at deepening the psychological and psychometric aspects of assessment using artificial agents (i.e., humanoid robots), telematic devices, or both.

The robot as a tool for psychometric diagnosis

Recent literature has shown that robots have the potential to be successful assistants in psychological assessment (Conti, Commodari & Buono, 2017; Di Nuovo et al., 2019; Fiske, Henningsen & Buyx, 2019; Riches et al., 2022; Rossi et al., 2018, 2020; Varrasi, Di Nuovo, Conti & Di Nuovo, 2018a, 2018b). However, when considering this method of administration using robots, perhaps the most innovative findings recently presented in the international literature on testing, show that there are relevant empirical results, even if still preliminary.

The literature on robotics that tries to address psychometric evaluation is quite limited. This is probably due to the skepticism of the practitioners (Conti, Di Nuovo, Buono & Di Nuovo, 2017) and to the perception of users on the reliability shown by the robots (de Graaf, Ben Allouch & van Dijk, 2019). Specifically, robots can be programmed to perform specific, repeatable actions, providing the benefit of attainable standardization. Therefore, the robotic implementation of quick screening tests could be promising, because they are often repetitive and easy to take, but can be time-consuming for human assessors (Di Nuovo et al., 2019). An example where this would work is the observation of developmental history and social skills, where clinicians with different specialisations often do not agree when evaluating the same patients (Scassellati, Admoni & Matarić, 2012).

However, since an essential characteristic of psychodiagnostic tests is that the stimuli and the methods for their administration should be just as rigorously standardized to guarantee the reliability (i.e. its repeatability in different times and places) and the validity of the results, assistive robots can represent a valuable way to meet these requirements and provide a reliable automatic tool for psychometric assessment (Di Nuovo et al., 2019; Rossi et al., 2018; Varrasi et al., 2018b; Varrasi, Lucas, Soranzo, McNamara & Di Nuovo, 2019). A robot-led assessment can provide a series of advantages,

including: assessor neutrality, objective measurement of social behaviour, standardization of the interaction, and better acceptance of the robotic platform than a non-embodied computer (Feingold Polak et al., 2018; Varrasi et al., 2018a).

In a recent study (Di Nuovo et al., 2019) aimed at the preliminary evaluation of the feasibility of entire or partial robotic testing, the neuropsychological test *Montreal Cognitive Assessment* (*MoCA*) was administered to a pilot sample of 16 adults (age range 19-61) using the Pepper robot. The purpose was to compare the standard score obtained traditionally, using paper-pencil assessments, with the automatic score obtained by the software integrated with the robot, and the supervised score, calculated by a psychologist correcting the automatic score through audio-video analysis (thus evaluating the communication and relationship aspects). In this case, the score was more correlated with the standard version, demonstrating to the psychologist the need to supervise automated assessments. The analyzes were repeated for the individual subtests of the MoCA, demonstrating that the different tasks within the test are affected differently by the method of robotic administration.

Another study conducted by Rossi et al. (2020) on a sample of 19 adults (8 males, 11 females, age range 53-82) involved both the traditional paper-pencil assessments of the MoCA test and robotic assessments, in addition to the of NEO-PI-3 personalities and the UTAUT questionnaire to evaluate the acceptance of computerized procedures. The results showed that the personality trait *Openness to experience* influenced the performance obtained during the interaction with the robot, and facilitated the unsupervised application of the robotic tool for cognitive assessment. Furthermore, anxiety or empathy towards robotic assessment also influence the differences from traditional assessment.

For these reasons, an important objective of future research will be to evaluate the reliability and validity of the use of the humanoid robot as a tester by acting remotely on its control. The overall aim will be to use more robots in different contexts with centralized supervision and therefore a considerable expansion of the testing possibilities for large screenings and patients with frailty that need extra support or care.

Moreover, the possibility of using a virtual robot for testing at distance should be considered.

Baig and Kavakli (2019), presenting a review of methods and measurements that are currently used in psychophysiological analysis to measure cognitive or mental states, also report on techniques of virtual reality and telerobotics, concluding that virtual reality simulations can be used to study the relationship between brain responses and stress levels. A telepresence robot was designed to measure the physical and psychological health of elderly people (Cortellessa et al., 2018). The results of recent reviews (Virkus, Leoste, Marmor, Kasuk & Talisainen, 2023; Zhang & Hansen, 2022) suggest that from a psychological perspective, telepresence robots could enable greater social presence in remote communication, but that their acceptance for wider use is still challenged.

Studies on online assessment with telematic tools

Currently there are few empirical types of research aimed at evaluating assessment using telematic tools, despite an increase in tools that are not limited to automatic scoring, but also provide for the possibility of remote administration.

Since the 1980s, e-marking has allowed administering tests to large groups at the same time in the fields of educational and industrial training within e-learning paradigms (e.g., Bukie, 2014; Ivanović & Jain, 2013; Preston & Shackelford, 1999). The diffusion of telematic testing has led to the critical discussion of technical aspects: e.g., copyright of the material, modification of administration procedures, the need for dedicated platforms, *ad hoc* standards and thresholds that compensate for the measurement error components introduced by the telematic medium. Furthermore, by reducing the interaction between the psychologist and the user, the relational and non-verbal aspects which are useful for the global understanding of the examined subject are underestimated. In this way, essential components of the diagnostic work would be taken away from the psychologist, especially when working with patients with frailty that need extra support or care and children, or in diagnoses of forensic relevance, which could hardly be adapted to the different remote settings.

Some experimental research has studied the equivalence between standard and remote administration, for example by comparing the traditional and telematic versions of the *General Health Questionnaire – 28* and of the *Symptoms Check-List – 90 –* R (Vallejo, Jordán, Díaz, Comeche & Ortega, 2007).

A meta-analytic review (Brearly et al., 2017) compared neuropsychological tools administered face-to-face and telematically, without highlighting significant differences for tests that use the verbal channel, such as the *Mini-Mental State Examination* (*MMSE*), the digit span, the Boston Naming Test and the fluency test. Few experimental verifications exist regarding tests where the manipulation of complex perceptual stimuli is necessary.

Furthermore, the opportunity to use remote neuropsychological assessment tools for elderly patients or patients with mobility difficulties has been studied (Cullum, Hynan, Grosch, Parikh & Weiner, 2014; Wadsworth et al., 2018). The concrete problem is choosing tests that are more or less suitable for remote administration, using those whose basic constructs are best designated to the type of user and the objectives of the evaluation. The elderly, children and adolescents, family systems or couples all have very different characteristics for which telematics testing needs to be specifically adapted.

As with all technologies applied to psychology, remote assessment is not a solution for all logistical or organizational problems. It poses the problems of usability and acceptability that should demonstrate which techniques are most appropriate to be translated profitably into the remote format, acceptable to both users and psychologists.

On these issues, previous survey research (Di Nuovo & Narzisi, 2021) studied the acceptance by potential users of telematic interventions (including assessment) by a comparison group of psychologists. The results show that a greater acceptance of being evaluated and treated online is linked - in addition to the possession of adequate tools and technical skills - to the belief in obtaining health benefits. The two personality traits (evaluated using the Big Five Factors model) that most predispose potential users to the use of technology and telematics are Extroversion and Open-mindedness. These results are in line with the international literature on the subject.

However, some important concerns are raised by the answers from professional psychologists. Their positive evaluation of online assessment correlates with the perception of being able to use it with children and adolescents too, with the understanding that tests to be administered remotely require specific standardization and calibration. However, a more positive assessment of psychologists is also associated with a greater belief that specific preparation is not necessary to administer online tests. It seems that the technical efficiency of the tool (either the telematic questionnaire or the robot tester) may be enough to ensure its diagnostic effectiveness.

Aims of the study

Reassuming the results of the literature, the need emerges for a further study aimed at deepening not only the general acceptance of the possibility of testing that does not foresee the direct relationship with the psychologist, but, more specifically, the technical conditions deemed necessary for the acceptance of these online testing methods by the professional psychologist.

Based on the preliminary purpose of the study, an exploratory approach was used. The specific aims are:

1. To verify the degree of acceptance of different aspects of telematic assessment, and the intention to use it, comparing these variables in the subgroups of psychologists based on gender and age.
2. To correlate the acceptance of remote assessment with the frequency of previous use of online questionnaires and with positive evaluation of the technology.
3. To ascertain what aspects of acceptance are more predictive of the intention of use.
4. To find dimensional relationships between the more relevant variables emerging in the study.

MATERIALS AND METHODS

Instruments

– The first instrument was a questionnaire involving the collection of responses to 21 items on a 5-point scale (see Appendix). Some of these items had been built according to the UTAUT model (Unified Theory of Acceptance and Use of Technology; Venkatesh, Morris, Davis & Davis, 2003), and had been adapted for the acceptability and use of online questionnaires for mental well-being (e.g., Sora, Nieto, Del Campo & Armayones, 2021). A specific section covered online psychodiagnostic testing. Here, 13 items were identified from a survey of online psychological assessments, promoted during the pandemic by the European Federation of Psychologists Associations.

  Three additional areas of enquiry looked at socio-

demographic information (gender and age) and the previous use of online instruments (never used; rarely; sometimes; often used).

Preliminarily the questionnaire was verified using a panel of judges, i.e. four psychologists who were experts in psychometrics, who evaluated the comprehensibility of each item and its pertinence for the study. Afterwards, the critical items were reformulated, and only items with an interrater agreement for suitability ≥.80 were maintained in the final version. A specific question was dedicated to the intention to use online testing in own professional practice, to be rated on a 5-point scale, from *not at all* to *surely*.

– Moreover, a s*emantic differential* (SD) was administered, consisting of 12 pairs of polar adjectives (e.g.: easy/difficult; useful/useless) on which to evaluate technology in general (Technology is…). For each pair, 6 intervals of response were foreseen, 3 for each polarity; e.g. for easy/difficult: very easy, quite easy, not very easy, not very difficult, quite difficult, very difficult. This tool allowed for a more connotative than denotative evaluation of the research object, and therefore was more linked to emotionality than to purely cognitive judgment (Osgood, Suci & Tannenbaum, 1957; Takahashi, Ban & Asada, 2016). A preliminary analysis of the instrument in a previous study (Di Nuovo & Narzisi, 2021) has demonstrated its monofactoriality, based mainly on the evaluation factor, and the consequent possibility of obtaining a total score by adding the scores of the individual scales, after having rotated all the items in a positive direction. The analysis of the reliability of the total score showed sufficient internal consistency: Cronbach's alpha index was .68 in the pilot study.

Participants

A sample of 122 licensed psychologists participated in the study. They were either attending or teaching post-graduate specialization courses of different theoretical-methodological focus (i.e., Cognitive-behavioral, Gestalt and Psychodynamic).

The gender of the sample (M = 40: 32.79%; F = 82: 67.21%) reflects the different proportions existing in the population of psychological professionals. In terms of age groups, 52.46% (N = 64; M = 19, F = 45) were under 45; while 47.54% (N = 58; M = 21, F = 37) of the sample was aged 46 or older.

Participants came from various Italian regions: Lombardy, Veneto, Liguria, Lazio, Campania, Apulia, Sardinia, and Sicily.

Regarding the previous use of online tests and questionnaires, 40.98% declared that they had never used them; while 15.57% used them on rare occasions, 32.79% occasionally, and 10.66% frequently.

Procedure

The participants were given an online survey via Google forms that included the questions shown in Appendix (Italian version was administered). The sample was recruited from lists of Italian psychologists from different psychotherapy schools and their teachers/supervisors.

The participants were invited to take part in the research via email and subsequently gave online written informed consent to the aggregated use of their data and answered the questionnaire anonymously. They were informed that the questionnaire would take ~10-15 minutes to complete, and the instructions indicated that they could withdraw at any time.

The study was approved by the Ethics Committee of the Department of Educational Sciences of the University of Catania (Italy).

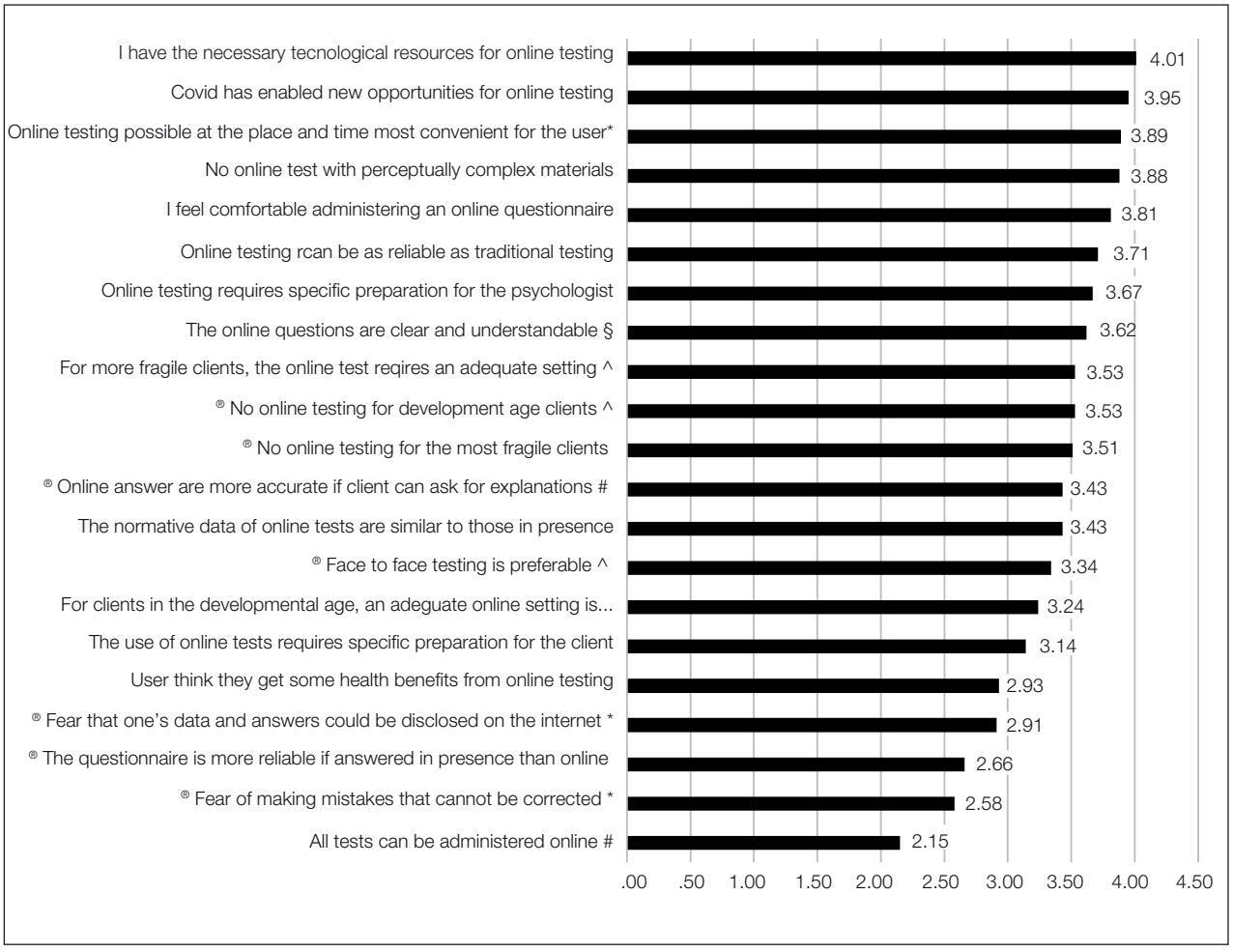
RESULTS AND DISCUSSION

Descripive statistics and subgroup comparison

The first analysis calculated the level of agreement for each item on the questionnaire across the total sample (see Figure 1). Figure 1, also shows the differences between gender and age, where significant on the t-test (p<.05).

In the psychologists' sample the items most agreed upon concerned the possession of adequate technological resources for online testing, along with the contribution of the pandemic to the spread of this practice, and the possibility of taking the test at times and in places most convenient to the customer (predominant in male respondents). These are attitudes concerned with usability, and were followed by concerns about technical conditions (problems with complex perceptive materials, difficulty to administer remotely, or with clients with frailty or children). The sample interviewed

Figure 1 – Rank of means of the items of the questionnaire in the total sample



Note. Significant differences for gender and age are reported.

Legenda. ® = Reversed item; significant differences for gender: * = M>F, ^ = F>M; significant differences for age: # = older>younger, § = younger>older.

seems to be quite aware of these technical problems. However, there is a high level of agreement on the possibility that online testing can be as reliable as traditional tests, whilst recognising the need for specific preparation, not only for the patients, but also for the psychologists who use this method. However, 25.40% ($N = 31$) of the sample - regardless of gender and age - disagreed with the need of a specific formation.

The scoring for face-to-face testing being preferable, is quite high: the mean is 3.34 on the 5-point scale, and 45.08% of psychologists ($N = 55$) would prefer to administer the test in person, therefore considering online testing as a fallback. The positive attitude is more present among older psychologists, where the percentage of agreement is 48.28%, against 42.19% of the younger ones.

Little importance is attributed to concerns about the possible violation of privacy, the difficulty of correcting any errors, or technical problems that are not considered more serious than in face-to-face testing.

As expected, the statement that "All tests can be used online" garnered the lowest ratings on the scale of agreement. The average score is very low (2.15 on a 5-point scale), but looking more closely at the single answers, one third of the

total (33.33%) are psychologists who agree (only 15, i.e. 12.50%, "completely agree"). The agreement is predominant at the higher age level ($\chi^2 = 12.61$, $df = 3$, $p = .01$).

On the other hand, the limited attribution to customers who trust being able to receive health benefits from online testing is surprising; only 26.23% of psychologists ($N = 32$) agree with this attribution, while the majority are in doubt or even against it, showing a certain distrust on the users' appreciation of the potential benefits (which instead resulted high in the research on users: see Di Nuovo & Narzisi, 2021).

A noticeable variable in our study is the intention to use online testing, which is quite high, but manifests a significant difference by age: youngest group $N = 64$, *Mean* = 4.44, $SD = .81$; older group $N = 58$, *Mean* = 3.88, $SD = .97$; $t_{120} = 3.44$, $p<.001$.

More than half of the younger respondents (59.38%) fully agree with the use of online testing, while only 10.94% refuse or are doubtful. In the group of older psychologists, the percentages are respectively 31.03 of full agreement and 32.76 of doubt or disagreement (the overall χ^2 is 12.74, $df = 3$, $p = .01$). But none of the respondents is completely against using the tests online.

No difference was found regarding gender ($\chi^2 = 1.08$, $df = 3$, $p = .78$).

The overall positivity towards technology assessment (total score derived from the semantic differential) is also, as expected, higher in the younger psychologists: younger group $N = 64$, *Mean* = 54.97, $SD = 4.47$; older group $N = 58$, *Mean* = 52.57, $SD = 6.20$; $t_{120} = 2.47$, $p<.01$.

Correlational analyses

The zero-order correlations between questionnaire items and frequency of previous use of online questionnaires and positive assessment of the technology were calculated using Pearson's coefficient. We report below on the correlations found to be significant for at least $p<.05$, after Bonferroni's correction for multiple comparisons.

The agreement with the statement: "Online testing is possible at the place and time most convenient for the user" correlates both with the frequency of previous use ($r = .40$, $p<.01$) and with the positive evaluation of the technology ($r = .24$, $p<.05$). The positive attitude towards technology correlates also with the agreement on the comprehensibility of the questions asked online ($r = .34$, $p<.01$), the recognition of the input given by Covid to online testing ($r = .24$, $p<.05$), and the need for an adequate setting for the online testing of children ($r = .20$, $p<.05$).

As shown in Table 1, the predictivity of the questionnaire items for the willingness to use online tests was evaluated by a backward stepwise regression analysis.

The most predictive variables of the intention to use, i.e. those selected for the stepwise regression, are the absence of a prejudicial preference for face-to-face testing, the recognition of the usefulness deriving from the pandemic period, and the admission that the online test can also be useful for children (but with attention to specific settings for these age levels). The reduced concern about the difficulty of administering online perceptually complex stimuli is also relevant, along with the risks of privacy violations. Finally, the intention to use online testing is influenced by feeling comfortable in this way of working and the convenience of using the test at the most appropriate times and places for the client.

Multidimensional analysis

To find dimensional relationships between the more relevant variables emerging in the study, the correlation matrix among the variables significant at the stepwise regression analysis, together with the three core variables (Intention of use, Previous use, Positive attitude toward technology), was submitted to of multidimensional scaling in two dimensions, using the Guttman/Lingoes method. Figure 2 shows the graphical results of the analysis.

Dimension 1 contrasts the attitudes towards online testing, showing the positive values on the right side of the graphic with the negative aspects on the left side. The positive aspects (i.e. experience during the pandemic, the possibility of specific settings, and feeling comfortable with online testing) are linked to the positive attitude towards technology and to the intention of use.

Dimension 2 contrasts technical aspects, both positive and negative (upper part of the plot), to the aspects related to the use (lower half of the graphic).

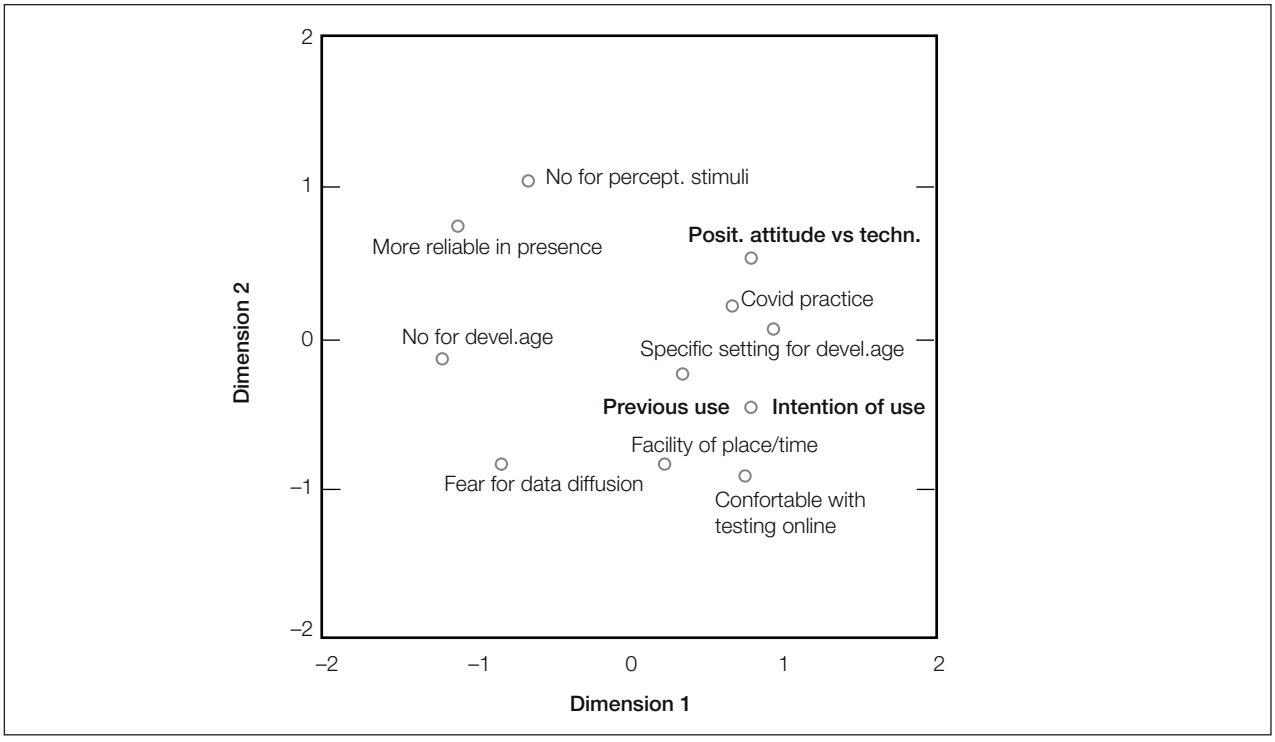
The analysis clearly shows that the technical aspects and usability are intertwined in the perception by the psychologists for defining the dimensions of online testing and the actual intention of use. Previous use of online instruments is neutral for both dimensions, although closely linked to the availability for use.

Table 1 – Backward stepwise regression

	St.coeff.	*p*
® Testing preferable in presence	−.38	<.001
Covid practice as input	.33	<.001
Specific setting needed for developmental age	.20	.04
® No for perceptual stimuli	−.19	.05
® Fear for data diffusion	−.34	.07
® No for developmental age	−.20	.09
Comfortable with online testing	.14	.10
Facility of place/time for users	.12	.10

Note. Dependent variable: intention of use, *p*-to-remove = .15, (R^2 = .47).

Figure 2 – Multidimensional scaling



Note. Alienation of final configuration = .20; proportion of variance (RSQ) = .83.

CONCLUSION

The limitations of this preliminary survey, parallel to the EFPA's international study, were (a) the impossibility of reliably differencing subgroups of the sample, (b) the exploratory approach, and (c) the use of a general questionnaire about online assessment, without asking separately about the different instruments to use online (e.g., cognitive tests, personality inventories, development scales, verbal vs performance tests). A more detailed study should be carried (a) with a larger sample articulated in subgroups of psychologists, already specialized or during their training period, (b) with specific hypotheses about subgroups differences, and (c) with a questionnaire including specific sections on the cognitive, personality, verbal skills and performance variables to measure online. Furthermore, the tool could be improved and generalized administering it by means of different media, i.e. a computer, an app and a robotic assistant.

Despite the preliminary character of the study, the literature review on online testing, and the research data presented, allows some consideration regarding the psychologists' awareness of the difference between remote assessment, or through artificial agents, and the traditional method. The latter - although it can be supported by computerized formats, scoring or interpretations - is however based on the administration of a face-to-face interaction between the psychologist and the user, which long-distance or virtual relationships cannot reproduce in full.

From the results obtained in the study, it can be deduced that the overall attitude of the psychologists toward these innovative modalities of assessment is positive, depending also on the perception of usefulness deriving from the pandemic period. The intention to use online testing is very high, especially in the younger subgroup of the sample. The most predictive variables of the intention to use online testing are the absence of prejudicial refusal of a modality different from face-to-face, and the admission that online testing can also be useful for children and persons with frailty that need extra support or care, but with attention to specific settings for these particular participants. Moreover, the intention to use online testing is influenced by how comfortable the psychologist feels about this way of working, and by the practical convenience of using the test at the most appropriate times and places for the client.

The multidimensional analysis clearly shows that the attitude towards online testing, and the intention to use it, are located in the crossing dimensions of technical (psychometric) aspects and concrete usability.

In conclusion, the general consideration that can be deduced is that psychological testing in telematic mode cannot be considered a simple surrogate for traditional testing, as distance learning or smart working are in educational or work environments. The administration of tests electronically or through artificial agents requires adaptation studies and in many cases a reformulation of the tools that are offered online, taking into account the necessary formal changes as well as the standardization and calibration of the answers obtained.

Finally, for these innovative methods to be a useful advance in psychological assessment, even beyond the pandemic period, the need to prepare specifically validated tools must be underlined, through appropriate psychometric research.

There is also a need to raise awareness amongst psychologists and train those who intend to use it. This training should take place at various levels of university teaching, from basic courses to masters and specialisations; at the same time, professional psychologists could be sensitized to the correct use of the innovative techniques, and to accept them for what they can actually offer.

Funding. This research has been fully supported by the "START to Aid" (209564) project (PIACERI - PIAno di inCEntivi per la RIcerca di Ateneo) of the Department of Humanities, University of Catania (Italy).

Acknowledgments. The authors gratefully thank all participants of this study for the generous contribution of their time.

References

BAIG, M.Z. & KAVAKLI, M. (2019). A survey on psycho-physiological analysis & measurement methods in multimodal systems. *Multimodal Technologies and Interaction*, *3* (2), 37.

BREARLY, T.W., SHURA, R.D., MARTINDALE, S.L., LAZOWSKI, R.A., LUXTON, D.D., SHENAL, B.V. & ROWLAND, J.A. (2017). Neuropsychological test administration by videoconference: A systematic review and meta-analysis. *Neuropsychology Review*, *27* (2), 174-186.

BUCHANAN, T. (2002). Online assessment: Desirable or dangerous? *Professional Psychology: Research and Practice*, *33* (2), 148.

BUKIE, O.F. (2014). Understanding technologies for e-assessment: A systematic review approach. *Journal of Emerging Trends in Computing and Information Sciences*, *5* (12), 936-947.

CONTI, D., COMMODARI, E. & BUONO, S. (2017). Personality factors and acceptability of socially assistive robotics in teachers with and without specialized training for children with disability. *Life Span and Disability*, *20* (2), 251-272.

CONTI, D., DI NUOVO, S., BUONO, S. & DI NUOVO, A. (2017). Robots in education and care of children with developmental disabilities: A study on acceptance by experienced and future professionals. *International Journal of Social Robotics*, *9*, 51-62.

CORTELLESSA, G., FRACASSO, F., SORRENTINO, A., ORLANDINI, A., BERNARDI, G., CORACI, L., DE BENEDICTIS, R. & CESTA, A. (2018). ROBIN, a telepresence robot to support older users monitoring and social inclusion: Development and evaluation. *Telemedicine and E-Health*, *24* (2), 145-154.

CULLUM, C.M., HYNAN, L.S., GROSCH, M., PARIKH, M. & WEINER, M.F. (2014). Teleneuropsychology: Evidence for video teleconference-based neuropsychological assessment. *Journal of the International Neuropsychological Society: JINS*, *20* (10), 1028.

DE GRAAF, M.M.A., BEN ALLOUCH, S. & VAN DIJK, J.A.G.M. (2019). Why would I use this in my home? A model of domestic social robot acceptance. *Human-Computer Interaction*, *34* (2), 115-173.

DI NUOVO, A., VARRASI, S., LUCAS, A., CONTI, D., McNAMARA, J. & SORANZO, A. (2019). Assessment of cognitive skills via human-robot interaction and cloud computing. *Journal of Bionic Engineering, 16*, 526-539.

DI NUOVO, S.F. & NARZISI, V. (2021). Telepsychology: Acceptability of online assessment and treatment in users and professional psychologists. *Mediterranean Journal of Clinical Psychology*, *9* (3).

FEINGOLD POLAK, R., ELISHAY, A., SHAHAR, Y., STEIN, M., EDAN, Y. & LEVY-TZEDEK, S. (2018). Differences between young and old users when interacting with a humanoid robot: A qualitative usability study. *HRI'18 Companion: Conference on ACM/IEEE International Conference on Human-Robot Interaction, March 5-8, 2018*, Chicago, IL, USA, 107-108.

FISKE, A., HENNINGSEN, P. & BUYX, A. (2019). Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *Journal of Medical Internet Research*, *21* (5), e13216.

GRAZIOLI, S., MAURI, M., ROSI, E., VILLA, F., TIZZONI, F., TARABELLONI, A., TRABATTONI, S., MAURI, V., COLOMBO, P. & MOLTENI, M. (2022). Use of machine learning on clinical questionnaires data to support the diagnostic classification of attention deficit hyperactivity disorder: A personalized medicine approach. *European Psychiatry*, *65* (S1), 165-166.

IVANOVIĆ, M. & JAIN, L.C. (2013). *E-learning paradigms and applications: Agent-based Approach (Vol. 528)*. Springer.

OSGOOD, C.E., SUCI, G.J. & TANNENBAUM, P.H. (1957). *The measurement of meaning (Issue 47)*. University of Illinois Press.

PRESTON, J.A. & SHACKELFORD, R. (1999). Improving on-line assessment: An investigation of existing marking methodologies. *Proceedings of the 4th Annual SIGCSE/SIGCUE ITiCSE Conference on Innovation and Technology in Computer Science Education*, 29-32.

RICHES, S., AZEVEDO, L., VORA, A., KALEVA, I., TAYLOR, L., GUAN, P., JEYARAJAGURU, P., McINTOSH, H., PETROU, C. & PISANI, S. (2022). Therapeutic engagement in robot-assisted psychological interventions: A systematic review. *Clinical Psychology & Psychotherapy*, *29* (3), 857-873.

ROSSI, S., CONTI, D., GARRAMONE, F., SANTANGELO, G., STAFFA, M., VARRASI, S. & DI NUOVO, A. (2020). The role of personality factors and empathy in the acceptance and performance of a social robot for psychometric evaluations. *Robotics*, *9* (2), 39.

ROSSI, S., SANTANGELO, G., STAFFA, M., VARRASI, S., CONTI, D. & DI NUOVO, A. (2018). Psychometric evaluation supported by a social robot: Personality factors and technology acceptance. *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 802-807.

SCASSELLATI, B., ADMONI, H. & MATARIĆ, M. (2012). Robots for use in autism research. *Annual Review of Biomedical Engineering*, *14*, 275-294. doi.org/10.1146/annurev-bioeng-071811-150036

SORA, B., NIETO, R., DEL CAMPO, A.M. & ARMAYONES, M. (2021). Acceptance and use of telepsychology from the clients'

perspective: Questionnaire study to document perceived advantages and barriers. *JMIR Mental Health*, *8* (10), e22199.

TAKAHASHI, H., BAN, M. & ASADA, M. (2016). Semantic differential scale method can reveal multi-dimensional aspects of mind perception. *Frontiers in Psychology*, *7*, 1717.

VALLEJO, M.A., JORDÁN, C.M., DÍAZ, M.I., COMECHE, M.I. & ORTEGA, J. (2007). Psychological assessment via the internet: a reliability and validity study of online (vs paper-and-pencil) versions of the General Health Questionnaire – 28 (GHQ-28) and the Symptoms Check-List – 90 – Revised (SCL-90-R). *Journal of Medical Internet Research*, *9* (1), e540.

VARRASI, S., DI NUOVO, S., CONTI, D. & DI NUOVO, A. (2018a). A social robot for cognitive assessment. *HRI'18 Companion: Conference on ACM/IEEE International Conference on Human-Robot Interaction, March 5-8, 2018, Chicago, IL*, USA, 269-270.

VARRASI, S., DI NUOVO, S., CONTI, D. & DI NUOVO, A. (2018b). Social robots as psychometric tools for cognitive assessment: A pilot test. *Springer Proceedings in Advanced Robotics - 10th International Workshop in Human-Friendly Robotics*, 99-112.

VARRASI, S., LUCAS, A., SORANZO, A., McNAMARA, J. & DI NUOVO, A. (2019). IBM cloud services enhance automatic cognitive assessment via human-robot interaction. In G. Carbone, M. Ceccarelli & D. Pisla (Eds.), *New trends in medical and service robotics*. Springer International Publishing.

VENKATESH, V., MORRIS, M., DAVIS, G. & DAVIS, F. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*, 425-478.

VIRKUS, S., LEOSTE, J., MARMOR, K., KASUK, T. & TALISAINEN, A. (2023, ahead-of-print). Telepresence robots from the perspective of psychology and educational sciences. *Information and Learning Sciences*.

WADSWORTH, H.E., DHIMA, K., WOMACK, K.B., HART JR.J., WEINER, M.F., HYNAN, L.S. & CULLUM, C.M. (2018). Validity of teleneuropsychological assessment in older patients with cognitive disorders. *Archives of Clinical Neuropsychology*, *33* (8), 1040-1045.

WAINER, H., DORANS, N.J., FLAUGHER, R., GREEN, B.F. & MISLEVY, R.J. (2000). *Computerized adaptive testing: A primer*. Routledge.

ZHANG, G. & HANSEN, J.P. (2022). Telepresence robots for people with special needs: A systematic review. *International Journal of Human-Computer Interaction*, *38* (17), 1651-1667.

APPENDIX

Questionnaire items

	English version	Italian version
1	*Online testing can be as reliable as traditional testing*	Il questionario online può essere attendibile quanto quelli compilati in modo tradizionale
2	*Online testing is possible at the place and time most convenient for the user*	Chi risponde al questionario online può farlo nel luogo e nel momento della giornata che ritiene più opportuno
3	*I feel comfortable administering an online questionnaire*	Mi sento a mio agio nel far compilare un questionario online
4®	*The answered face to face questionnaire is more reliable than online*	Il questionario compilato in presenza dà risultati più affidabili di quello online
5®	*Some respondents could fear of making mistakes that cannot be corrected*	Qualcuno potrebbe esitare a rispondere alle domande online per paura di commettere errori che non può correggere
6	*Users think they get some health benefits from online testing*	Gli utenti pensano di poter ottenere dei benefici per la loro salute dopo la compilazione del questionario online
7®	*Online answers could be more accurate if the client can ask for explanations*	Se avessero la possibilità di chiedere spiegazioni sul contenuto di alcune domande, le risposte online degli utenti sarebbero più accurate
8	*The online questions are clear and easily understandable*	Le domande online sono scritte in modo chiaro e facilmente comprensibile
9	*I have the necessary technological resources to administer online testing (internet connection, suitable device)*	Ho le risorse necessarie per somministrare questionari online (connessione internet, dispositivo adeguato)
10®	*If the choice is possible, face to face testing is preferable*	Se avessi la possibilità di scegliere, preferirei far compilare il questionario faccia a faccia
11®	*Some respondent could fear that one's data and answers could be disclosed on the Internet*	Chi risponde online può avere paura che i propri dati e le proprie risposte possano essere diffusi su Internet
12	*The normative data of online tests are similar to those administered in presence*	I dati normativi dei test online sono analoghi a quelli della somministrazione in presenza
13	*Online testing requires specific preparation for the psychologist*	L'utilizzo dei test online richiede una specifica preparazione per lo psicologo
14	*Online testing requires specific preparation for the client*	L'utilizzo dei test online richiede una specifica preparazione per il cliente
15®	*Online testing is not suitable for the most clients with frailty*	Per i clienti più fragili è meglio non procedere alla somministrazione di test online
16	*For frail patients, the online test requires an adequate setting*	Per i clienti più fragili si può procedere alla somministrazione del test online se viene introdotto un adeguato setting
17	*All tests can be administered online*	Tutti i test possono essere utilizzati online

continued on next page

continued

	English version	Italian version
18®	*Online testing is not suitable if perceptually complex materials are required*	Test che richiedono l'utilizzo di materiali percettivamente complessi non possono essere utilizzati online
19®	*Online testing is not suitable for clients in developmental age*	Per i clienti in età evolutiva è meglio non procedere alla somministrazione di test online
20	*For clients in the developmental age, an adequate online setting is necessary*	Per i clienti in età evolutiva si può procedere alla somministrazione del test online se viene introdotto un adeguato setting
21	*Covid-19 has enabled new opportunities for online testing*	Il Covid-19 ha fornito l'opportunità di usare nuovi processi digitali quali il testing online

Legenda. ® = reversed scoring.

DOI: 10.26387/bpa.2022.00011

Emotional and health correlates of body dissatisfaction during the menopausal transition

Daniele Saraulli[1], Serena Matroberardino[2], Pietro Spataro[3]

[1] Department of Law, Economics, Politics and Modern Languages, LUMSA University, Rome, Italy
[2] Department of Psychology, Faculty of Medicine and Psychology, Sapienza University of Rome, Italy
[3] Faculty of Economy, Universitas Mercatorum, Rome, Italy

pietro.spataro@unimercatorum.it

ABSTRACT. Lo studio esamina i correlati psicologici ed emotivi dell'insoddisfazione per il proprio corpo in un campione di donne in menopausa. I risultati suggeriscono che alti livelli di insoddisfazione corporea si associano con problemi emotivi, una minore capacità di controllo delle espressioni di rabbia e una diminuzione della salute mentale. Inoltre, le donne esaminate durante il periodo di menopausa (perimenopausa) mostrano gli esiti più negativi, in termini di ansia, depressione e salute mentale, rispetto alle donne esaminate prima dell'inizio del periodo di menopausa o dopo la fine della transizione.

SUMMARY. The present study aimed at investigating the links between body dissatisfaction and other psychological variables pertaining to emotional life and mental health, as a function of menopausal status. To this purpose, we administered the Body Uneasiness Test (BUT), a self-report questionnaire specifically devoted to the examination of body image disorders, to a sample of pre-, peri-, or postmenopausal women aged between 45 and 55 years. In addition, mood states, feelings of anger and general health conditions were assessed with the Profile of Mood States (POMS), the State-Trait Anger Expression Inventory (STAXI) and the Short Form General Health Survey (SF-36), respectively. The results showed that high levels of body dissatisfaction were associated with increases in mood disorders and anger expression, as well as with decreases in mental health and the ability to control angry feelings. Moreover, the regression analyses indicated that these associations remained significant after controlling for age and menopausal status. Body dissatisfaction remained stable across the menopausal period; however, perimenopausal women reported higher scores of depression, fatigue, and anxiety in the POMS, were less able to control angry feelings in the STAXI and had significantly lower scores of physical and mental health in the SF-36, compared to either premenopausal or postmenopausal women. In summary, our results suggest that body dissatisfaction might constitute a significant risk factor for the mental health of midlife women experiencing the menopausal transition.

Keywords: *Menopause, Body image, Body dissatisfaction, Mental health*

INTRODUCTION

In the last decades there has been a growing interest in investigating the consequences of menopause on women's body appearance and self-esteem (Gannon & Stevens, 1998). During the menopausal transition, most women experience substantial physical and hormonal changes that can have a negative impact on their body image perception and satisfaction (Deeks, 2004). Therefore, the present study had two primary aims. First, to determine whether body dissatisfaction was associated with psychological and health variables in midlife women aged 45-55 years. Second, to verify whether pre-, peri- and postmenopausal women differed in their psychological characteristics.

Regarding the first issue, an emerging body of research has begun to examine the psychological correlates of body dissatisfaction during the menopausal transition. Jackson et al. (2014), for example, reported data from the Study of Women's Health Across the Nation (SWAN) Chicago site, showing that women with high body image dissatisfaction or who perceived themselves as unattractive were more likely to report clinically significant levels of depressive symptoms. Similar findings have been obtained by Carrard et al. (2021), who found that body dissatisfaction, importance of appearance and aging anxiety were associated with higher depression scores in women aged 45-65. More generally, there is evidence indicating that body dissatisfaction is associated with numerous undesirable outcomes in midlife women, including high psychological distress, worsened life satisfaction, impaired psychosocial functioning, and lower mental and physical health (Becker, Verzijl, Kilpela, Wilfred & Stewart, 2019; Medeiros de Morais et al., 2017; Mond et al., 2013; Runfola et al., 2013).

With respect to the second issue, few quantitative studies have investigated whether pre-, peri- and postmenopausal women differ in their body image and the overall picture has been largely contradictory (McKinley & Lyon, 2008; Pearce, Thøgersen-Ntoumani & Duda, 2014; Quittkat, Hartmann, Düsing, Buhlmann & Vocks, 2019; Runfola et al., 2013; Tiggemann, 2004). On the one hand, Deeks and McCabe (2001) found that premenopausal women had higher ratings of appearance evaluation and nominated smaller silhouettes from the *Stunkard Body Shape Figure Scale* than women who were perimenopausal and postmenopausal, although these results became no longer significant when the confounding effects of age were considered. Significant differences have been likewise reported by Séjourné, Got, Solans and Raynal (2019), who showed that body dissatisfaction was significantly higher in the perimenopausal group (compared with the premenopausal group), and by McLaren, Hardy and Kuh (2003), who concluded that postmenopausal women and women who started hormone replacement therapy felt more satisfied with their weight than did premenopausal women. On the other hand, a different line of research indicated that body image remained substantially stable during the menopausal transition (Koch, Mansfield, Thurau & Carey, 2005; Mangweth-Matzek et al., 2021). Koch and colleagues (2005), for example, showed that the self-perceived attractiveness of a sample of 307 women between 39 and 56 years did not significantly differ based on their menopausal status.

Controversial results have been also reported regarding the relations between menopausal status and emotional problems such as anxiety and depression (Kurpius, Nicpon & Maresh, 2001; Woods & Mitchell, 1997). Baker, Simpson and Dawson (1997), for instance, showed that perimenopausal women were more prone to experience feelings of anxiety than premenopausal women but argued that mood changes could be a by-product of sleep disruption. More recently, a longitudinal analysis of data from the multi-site Study of Women's Health Across the Nation (SWAN), including 3300 US participants, reported that women with low levels of anxiety at baseline were more likely to report high levels of anxiety symptoms when early or late perimenopausal or postmenopausal, compared to when they were premenopausal, independently of multiple risk factors (upsetting life events, poor perceived health, and vasomotor symptoms: Bromberger et al., 2019, Tang et al., 2019). Yet, a systematic review by Bryant, Judd and Hickey (2012) concluded that anxiety symptom levels were low throughout the menopausal transition and that none of the available studies provided solid data on the prevalence of anxiety disorders that met diagnostic criteria.

A similar picture emerged for depression (Vivian-Taylor & Hickey, 2014). Again, some authors claimed that there is no direct link between menopausal status and depression (e.g., Dennerstein, 1996; Kaufert, Gilbert & Tate, 1992). Specifically, a longitudinal analysis of data derived from the Massachusetts Women's Health Study, with 2565 women aged 45 to 55 years at baseline, showed that the onset of natural menopause was not associated with increased risk of depression, although experiencing a long perimenopausal period (at least 27 months) predicted higher levels of depression (Avis, Brambilla,

McKinlay & Vass, 1994). Other authors, however, reported that the rates of depressive disorders varied significantly across the different menopausal stages (Bromberger et al., 2001, 2007; Tang et al., 2019). Bromberger et al. (2007), for example, found that a woman was more likely to report high levels of depression (CES-D≥16) when she was early perimenopausal, postmenopausal or currently/formerly using hormone therapy (HT), relative to when she was premenopausal.

Considering this background, the present study aimed at investigating the links between body dissatisfaction and other psychological variables pertaining to emotional life and mental health, as a function of menopausal status. We focused on the following issues: a) were individual differences in body dissatisfaction associated with individual differences in mood disturbance, anger expression and general health (Simbar et al., 2020)?, b) was body dissatisfaction significantly related to mood disturbance, anger expression and general health, after removing the effects of age and menopausal status (Deeks & McCabe, 2001)?, and c) did premenopausal, perimenopausal and postmenopausal groups differ in terms of body dissatisfaction, mood disturbance, anger expression and general health (Séjourné et al., 2019)? To address these questions, we administered four self-report measures to a sample of Italian-speaking midlife women between 45 and 55 years. The questionnaires we used were the *Body Uneasiness Test* (*BUT*; a self-report questionnaire specifically devoted to the examination of body image disorders: Cuzzolaro, Vetrone, Marano & Garfinkel, 2006), the *Profile of Mood States* (*POMS*; McNair, Lorr & Droppleman, 1981), the *State-Trait Anger Expression Inventory* (*STAXI*; Spielberger, 1988) and the *Short Form General Health Survey* (*SF-36*; Ware, Snow, Kosinski & Gandek, 1993). Briefly, based on previous evidence, we expected body dissatisfaction to be positively associated with mood disturbance and anger expression (suggesting that women having higher levels of body dissatisfaction should report more negative moods and should be more likely to experience anger feelings, as compared to women having lower levels of body dissatisfaction), but negatively associated with general health (suggesting that women having higher levels of body dissatisfaction should report worse health conditions, as compared to women having lower levels of body dissatisfaction). In addition, these associations should remain significant after removing age differences. As for the differences between menopausal stages, we expected to find an increase in mood disturbance and anger expression, together with a decrease in general health conditions, in the peri-menopausal group (as compared to the pre- and post-menopausal groups). However, we expected body dissatisfaction to remain substantially stable during the menopausal transition (after controlling for age effects: Koch et al., 2005; Mangweth-Matzek et al., 2021).

METHODS

Participants

172 women, with a mean age of 49.6 years (SD = 3.4; range: 45-55 years) and an average of 1.3 children (range: 0-4), were enrolled in the study. Most of them were married (N = 128); the remaining were divorced (N = 24), widowed (N = 5) or single (N = 15). Level of education was distributed as follows: 8 women had a primary-school license, 132 had a secondary-school license and 32 had a university degree. The menopausal status of participants was determined according to self-reports of menstrual bleeding patterns, following the Stages of Reproductive Aging Workshop (STRAW; Soules et al., 2001) and World Health Organization (WHO; Mangweth-Matzek et al., 2021) criteria. Specifically, 76 women were classified as premenopausal (i.e., having regular menses during the past 12 months; age: M = 47.5), 30 as perimenopausal (i.e., reporting the recent onset of amenorrhea or having irregular menses for at least 3 months but less than 12 months; age: M = 49.9) and 66 as postmenopausal (i.e., having no menstruation during the past 12 months; age: M = 51.9). All participants were recruited during periodical examinations at the Department of Obstetrics and Gynaecology of the San Giovanni Calibita Fatebenefratelli Hospital (Rome). After signing the informed consent, participants completed the entire survey in a unique session. The order of the questionnaires was counterbalanced across participants. Menopausal women who were using hormone replacement therapy at the time of testing were excluded from the study. All procedures and instruments were approved by the ethical committee of the San Giovanni Calibita Fatebenefratelli Hospital.

Material

– *Body Uneasiness Test* (*BUT*; Cuzzolaro et al., 2006). The BUT consists of 34 clinical statements assessing the following five independent factors: Weight phobia ("I'm

terrified of putting on weight"), Body image concerns ("I spend a lot of time thinking about some defects in my physical appearance"), Avoidance behaviors ("When I undress, I avoid looking at myself"), Compulsive self-monitoring ("I spend a lot of time in front of the mirror") and Depersonalization ("I feel detached from my body"). A Global Severity Index (GSI) measures the overall amount of body uneasiness. Participants responded on a 6-point scale (from "never" to "always"), indicating the extent to which each item fitted their actual experience. Higher scores indicate greater body dissatisfaction. In the present study, the internal consistencies of the five subscales, evaluated with Cronbach's alpha coefficients, were good, ranging from .60 to .81.

– *Profile of Mood States* (*POMS*; McNair et al., 1981; Italian version: Farnè, Sebellico, Gnugnoli & Corallo, 1991). The POMS assessment yields a profile of transient mood states, on the following six scales: Tension/Anxiety, Anger/Hostility, Depression/Dejection, Fatigue/Inertia, Confusion/Bewilderment and Vigor/Activity. Women rated 58 adjectives on a 5-point scale, indicating the extent to which each term described their mood states in the past week. Higher scores indicate more negative moods (except for the Vigor scale). A Total Mood Disturbance (TMD) was also computed, by adding all subscales and subtracting Vigor (Curran, Andrykowski & Studts, 1995). In the present study, the internal consistencies of the six subscales were good, ranging from .66 to .89.
– *State-Trait Anger Expression Inventory* (*STAXI*; Spielberger, 1988; Italian version: Comunian, 2004). The STAXI is a 44-item inventory which measures the intensity of anger as an emotional state at a particular time (State Anger: S-Anger) and the disposition to experience angry feelings over time as a personality trait (Trait Anger: T-Anger). Two additional subscales were used in the present study. The Anger-Control (AX/Con) scale assesses the tendency to control the expression of anger, whereas the Anger-Expression (AX/Ex) scales provides an overall measure of how often anger is being experienced and expressed. Participants responded on a 4-point scale (from "almost never" to "al-most always"), where higher scores indicated more frequent states of anger. In the present study, the internal consistencies of the four measures were good, ranging from .79 to .87.
– *Short Form General Health Survey* (*SF-36*; Ware et al., 1993; Italian version: Apolone & Mosconi, 1998). The SF-36 yields a profile of functional health and well-being, on the following eight scales: Physical functioning, Role limitations due to physical problems, Bodily pain, General health perceptions, Vitality, Social functioning, Role limitations due to emotional problems and Mental health. Scores range from 0 (lowest well-being) to 100 (highest well-being). Two summary measures were also computed: a physical component (the average of the Physical functioning, Bodily pain, Role limitations due to physical problems, and General health subscales) and a mental component (the average of the Mental health, Vitality, Role limitations due to emotional problems and Social functioning subscales). In the present study, the internal consistencies of the eight scales were good, ranging from .74 to .85.

Statistical analyses

Pearson's correlations were used to assess the bivariate associations between the BUT-GSI, the POMS-TMD, the STAXI scores and the physical and mental components of the SF-36. In addition, a series of hierarchical regressions were performed to determine whether body dissatisfaction predicted women's scores in the POMS, STAXI and SF-36 questionnaires, above and beyond the contributions due to menopausal status and age (Deeks & McCabe, 2001). Age and menopausal status (coded as two dummy variables) were always included in the first step to remove their potential effects, whereas body dissatisfaction was entered in the second step. Specifically, since menopausal status involved three groups (premenopausal, perimenopausal and postmenopausal women), two dummy coded variables were needed to fully represent all categories. Values in the first dummy variable were 1 if the woman belonged to the premenopausal group and 0 if she belonged to the perimenopausal or postmenopausal groups. Values in the second dummy variable were 1 if the woman belonged to the perimenopausal group and 0 if she belonged to the premenopausal or postmenopausal groups. Thus, the postmenopausal group was considered as the reference category in both cases.

Differences among premenopausal, perimenopausal, and postmenopausal women on each questionnaire were initially assessed by multivariate analyses of variance (MANOVAs), followed by univariate ANOVAs and post-

hoc comparisons (with Bonferroni adjustment; Séjourné et al., 2019). It should be noted that, in our conditions, non-parametrical statistics might have been more appropriate to examine group differences, since there were important differences in the number of premenopausal, perimenopausal and postmenopausal women recruited in the three groups. In these conditions, typical ANOVAs are known to have reduced statistical power and reduced robustness to unequal variances. To determine whether this was the case, we performed separate Kruskal-Wallis tests for each dependent variable illustrated in Table 1. The conclusions obtained with these analyses were highly comparable with those obtained with standard ANOVAs. We therefore decided to report only the results of parametrical analyses. This choice was also preferable because it allowed us to control for between-group differences in age, by entering this variable as a covariate.

These analyses were then replicated by including age as a covariate, since the three groups were not matched on this variable [$F_{(2,\ 169)}$ = 52.76, p<.001; pairwise post-hoc comparisons were all significant, indicating a linear increase in age from premenopausal to postmenopausal women: p<.002]. They were instead equivalent on education [$F_{(2,\ 169)}$ = 1.11, p = .33], marital status [married vs unmarried: χ^2 = 2.18, p = .33], and number of children [$F_{(2,\ 169)}$ = .72, p = .49]. Including age as a covariate allowed us to determine whether apparent differences between premenopausal, perimenopausal, and postmenopausal groups were explained or not by age (Deeks & McCabe, 2001).

RESULTS

Bivariate correlations

Table 1 reports Pearson's correlations between age, the BUT-GSI, the POMS-TMD, the four subscales of the STAXI and the physical and mental components of the SF-36. Two points should be noted. First, age was negatively associated with the BUT-GSI and the POMS-TMD, suggesting that body dissatisfaction and mood disturbances were lower in older than in younger women. Second, the BUT-GSI was positively associated with the POMS-TMD and the S-Anger, T-Anger, and AX/Ex subscales of the STAXI, but negatively associated with the AX/Con subscale of the STAXI and the mental component of the SF-36. Thus, women who were less satisfied with their bodies were more likely to exhibit high levels of mood disturbance and to experience high levels of state and trait anger; in addition, they were less able to control the expression of anger and had lower levels of mental health.

Hierarchical regressions

The results of the hierarchical regression analyses are illustrated in Table 2. As can be noted, controlling for age differences, the role of body dissatisfaction was significant in all analyses, except for the SF-36 physical component. Moreover, menopausal status predicted state anger, anger control and the SF-36 mental and physical components. For state anger, data suggest that being premenopausal was associated with a significant increase in the tendency to experience anger as a transient emotional state (relative to being postmenopausal). For anger control, being perimenopausal was associated with a significant decrease in the ability to control the expression of anger. For the SF-36 mental component, being both premenopausal and perimenopausal was associated with a significant decrease in mental health. Lastly, for the SF-36 physical component, being perimenopausal was associated with a significant decrease in physical health. Interestingly, age had no significant effect on any variable, when considered together with menopausal group and body dissatisfaction.

Group differences

Table 3 illustrates the mean scores (and the corresponding standard deviations) achieved by premenopausal, perimenopausal and postmenopausal women in the four questionnaires.

BUT. The initial MANOVA revealed no significant multivariate difference between the three groups [Wilks' λ = .93, $F_{(10,\ 330)}$ = 1.31, p = .22, η^2 = .04]. The follow-up univariate analyses, however, found significant differences on the weight phobia [$F_{(2,\ 169)}$ = 4.10, p = .018, η^2 = .05] and body image concerns subscales [$F_{(2,\ 169)}$ = 3.77, p = .025, η^2 = .04], and marginal differences on the avoidance [$F_{(2,\ 169)}$ = 2.58, p = .079, η^2 = .03] and depersonalization subscales [$F_{(2,\ 169)}$ = 2.40, p = .094, η^2 = .02]. The post-hoc pairwise comparisons indicated that both weight phobia and body image concerns were higher for premenopausal than for post-menopausal women (p = .028 and p = .038, respectively). Notably, these univariate differences were eliminated when age was entered

Table 1 – Pearson's correlations between all variables (N = 172)

Total sample	1	2	3	4	5	6	7	8	9
1. Age	1.00								
2. BUT-GSI	−.16*	1.00							
3. POMS-TMD	−.16*	.67*	1.00						
4. S-Anger	−.06	.24*	.39*	1.00					
5. T-Anger	−.03	.53*	.64*	.37*	1.00				
6. AX/Con	.06	−.31*	−.39*	−.28*	−.48*	1.00			
7. AX/Ex	−.10	.50*	.65*	.47*	.72*	.77*	1.00		
8. SF-36 Physical	−.07	−.07	−.10	.04	−.12	.04	−.06	1.00	
9. SF-36 Mental	.02	−.35**	−.49*	−.19*	−.40*	.33*	−.41*	.45*	1.00

Note. * $p \leq .05$; ** $p < .01$

in the model as a covariate [$F_{(2, 168)} < 1.76$, $p > .17$, $\eta^2 < .02$], except for a marginal difference on the body image concerns subscale [$F_{(2, 168)} = 2.47$, $p = .087$, $\eta^2 = .03$; none of the post-hoc pairwise comparisons reached the significance level: all $p > .16$]. In line with these findings, when considering the overall GSI scores, group differences were significant before covarying for age [$F_{(2, 169)} = 3.29$, $p = .039$, $\eta^2 = .04$; GSI scores were marginally higher for premenopausal than post-menopausal women: $p = .081$], but were eliminated after entering this variable in the model [$F_{(2, 168)} = 1.58$, $p = .21$, $\eta^2 = .01$].

POMS. For the POMS questionnaire, the multivariate difference failed to reach the significance level [Wilks' $\lambda = .89$, $F_{(12, 328)} = 1.49$, $p = .12$, $\eta^2 = .05$]. However, the follow-up univariate analyses revealed significant group differences on the Tension/Anxiety [$F_{(2, 169)} = 5.54$, $p = .005$, $\eta^2 = .06$], Depression/Dejection [$F_{(2, 169)} = 4.41$, $p = .014$, $\eta^2 = .05$], and Fatigue/Inertia [$F_{(2, 169)} = 5.06$, $p = .007$, $\eta^2 = .06$] subscales. In addition, there was a marginal difference on the Anger/Hostility subscale [$F_{(2, 169)} = 2.41$, $p = .093$, $\eta^2 = .03$]. In all cases, the pairwise comparisons indicated higher scores for perimenopausal than postmenopausal women (all $p < .05$). Unlike the BUT, all differences remained significant after covarying for age [for Tension/Anxiety: $F_{(2, 168)} = 3.95$, $p = .021$, $\eta^2 = .05$; for Depression/Dejection: $F_{(2, 168)} = 3.31$, $p = .039$, $\eta^2 = .04$; for Fatigue/Inertia: $F_{(2, 168)} = 3.48$, $p = .033$, $\eta^2 = .04$], except for the Anger/Hostility subscale [$F_{(2, 168)} = 1.22$, $p = .29$, $\eta^2 = .01$]. For the overall TMD scores, group differences were significant in the initial analysis [$F_{(2, 169)} = 4.43$, $p = .013$, $\eta^2 = .05$], but became marginal when covarying for age [$F_{(2, 168)} = 2.84$, $p = .061$, $\eta^2 = .03$]. The post-hoc comparisons confirmed that total mood disturbance was higher for perimenopausal than for postmenopausal women ($p = .057$).

STAXI. In the initial analysis, a significant multivariate difference was found among the three groups [Wilks' $\lambda = .90$, $F_{(8, 332)} = 2.18$, $p = .028$, $\eta^2 = .05$]. The follow-up univariate analyses revealed significant differences on the State-

Table 2 – Hierarchical regressions predicting women's scores in the POMS, STAXI and SF-36 questionnaires (N = 172)

Predicted measure		Predictors	β	t	R^2_{adj}	F change
POMS-TMD	Step 1	Age	−.06	−.79	.04	F = 3.49*
		Premenopausal group	.01	.17		
		Perimenopausal group	.11	1.60		
	Step 2	Body dissatisfaction	.64	11.07**	.44	F = 35.14**
S-Anger	Step 1	Age	.10	1.09	.02	F = 2.25†
		Premenopausal group	.23	2.24*		
		Perimenopausal group	.06	.64		
	Step 2	Body dissatisfaction	.22	2.91**	.06	F = 3.88**
T-Anger	Step 1	Age	.04	.48	.01	F = .67
		Premenopausal group	−.02	−.22		
		Perimenopausal group	.04	.51		
	Step 2	Body dissatisfaction	.53	7.88**	.26	F = 16.23**
AX/Con	Step 1	Age	−.03	−.33	.04	F = 3.49*
		Premenopausal group	−.09	−.98		
		Perimenopausal group	−.22	−2.70**		
	Step 2	Body dissatisfaction	−.28	−3.82**	.11	F = 6.48**
AX/Ex	Step 1	Age	.02	.24	.02	F = 2.23†
		Premenopausal group	.08	.90		
		Perimenopausal group	.12	1.54		
	Step 2	Body dissatisfaction	.48	7.08**	.24	F = 14.71**
SF-36-Mental	Step 1	Age	−.14	−1.51	.04	F = 3.43*
		Premenopausal group	−.19	−1.95*		
		Perimenopausal group	−.21	−2.61**		
	Step 2	Body dissatisfaction	−.33	−4.56**	.14	F = 8.08**
SF-36-Physical	Step 1	Age	−.05	−.49	.03	F = 2.98*
		Premenopausal group	.05	.44		
		Perimenopausal group	−.19	−2.17*		
	Step 2	Body dissatisfaction	−.06	−.81	.43	F = 2.39*

Note. † .06<p<.10; * p≤.05; ** p≤.01

Table 3 – Means and standard deviations for each questionnaire, as a function of menopausal status)

Measures	Menopausal status		
	Premenopause (N = 76)	Perimenopause (N = 30)	Postmenopause (N = 66)
BUT			
Weight phobia	1.30 (1.02)	1.49 (1.27)	.98 (.85)
Body image concerns	.94 (.93)	1.05 (1.04)	.74 (.66)
Avoidance behaviors	.39 (.67)	.55 (.90)	.28 (.41)
Compulsive self-monitoring	.85 (.63)	.74 (.79)	.76 (.58)
Depersonalization	.44 (.62)	.57 (.67)	.35 (.46)
POMS			
Tension/Anxiety	11.33 (6.75)	12.84 (7.84)	9.21 (5.78)
Anger/Hostility	10.35 (9.84)	12.60 (9.37)	8.40 (9.21)
Depression/Dejection	10.55 (11.59)	15.09 (11.66)	7.57 (9.64)
Fatigue/Inertia	8.24 (5.72)	10.42 (6.05)	6.51 (4.20)
Confusion/Bewilderment	7.59 (5.12)	8.91 (4.83)	7.05 (4.64)
Vigor/Activity	15.06 (6.29)	13.62 (5.35)	15.06 (6.26)
STAXI			
State-Anger	13.40 (5.67)	11.42 (5.36)	11.45 (2.77)
Trait-Anger	18.64 (5.60)	19.60 (6.68)	18.22 (5.11)
Anger-Control	22.60 (5.62)	19.62 (5.56)	24.05 (5.12)
Anger-Expression	23.63 (8.99)	26.00 (10.21)	20.91 (9.44)
SF-36			
Physical functioning	84.10 (16.60)	68.55 (23.31)	81.06 (17.64)
Role limitations (Physical)	67.51 (35.44)	59.55 (39.53)	61.74 (40.72)
Bodily pain	70.52 (23.35)	63.05 (22.73)	69.58 (23.53)
General health	61.42 (14.75)	57.04 (17.80)	61.14 (19.85)
Vitality	58.03 (19.60)	51.27 (19.44)	60.98 (19.71)
Social functioning	68.27 (24.40)	60.62 (22.38)	74.21 (20.22)
Role limitations (Emotional)	45.82 (39.92)	46.45 (38.78)	52.83 (40.07)
Mental health	64.67 (20.50)	56.07 (20.01)	66.06 (19.86)

Anger [$F_{(2, 169)}$ = 3.02, p = .051, η^2 = .04], Anger-Control [$F_{(2, 169)}$ = 5.26, p = .006, η^2 = .06], and Anger-Expression [$F_{(2, 169)}$ = 3.33, p = .038, η^2 = .04] subscales. After entering age as a covariate, the differences on the State-Anger and Anger-Control subscales remained significant [$F_{(2, 168)}$ = 3.01, p = .052, η^2 = .04 and $F_{(2, 168)}$ = 4.88, p = .009, η^2 = .05], whereas differences on the Anger-Expression subscale became marginal [$F_{(2, 168)}$ = 2.48, p = .086, η^2 = .03]. Pairwise comparisons indicated that: a) State-Anger was higher in premenopausal than in postmenopausal women (p = .046), and b) Anger-Control was lower in perimenopausal than in post-menopausal women (p = .006). For the Anger-Expression subscale [$F_{(2, 168)}$ = 2.48, p = .086, η^2 = .03], scores were numerically higher in perimenopausal than in postmenopausal women (p = .087).

S*F-36*. In the initial analysis, a significant multivariate effect was found on the SF-36 questionnaire [Wilks' λ = .69, $F_{(16, 326)}$ = 4.12, p<.001, η^2 = .17]. The follow-up univariate analyses confirmed significant differences between the three groups on the Physical functioning [$F_{(2, 169)}$ = 16.76, p<.001, η^2 = .16], Social functioning [$F_{(2, 169)}$ = 9.19, p<.001, η^2 = .10], and Mental health [$F_{(2, 169)}$ = 5.79, p = .004, η^2 = .06] subscales. In addition, there was a marginal significance for the Vigor/Activity subscale [$F_{(2, 169)}$ = 2.77, p = .065, η^2 = .03]. These results remained significant after covarying for age [for Phsysical functioning: $F_{(2, 168)}$ = 16.36, p<.001,η^2 = .16; for Social functioning: $F_{(2, 168)}$ = 9.47, p<.001, η^2 = .10; for Mental health: $F_{(2, 168)}$ = 5.79, p = .004, η^2 = .06], with the exception of the Vigor/Activity subscale, for which differences fell below the significance level [$F_{(2, 168)}$ = 1.12, p = .32, η^2 = .01]. The pairwise comparisons revealed that: a) physical functioning was significantly lower in perimenopausal than in premenopausal and postmenopausal women (both p<.001), b) social functioning was significantly lower in perimenopausal than in postmenopausal women (p<.001) and marginally lower in perimenopausal than in premenopausal women (p = .053), and c) mental health was significantly lower in perimenopausal than in postmenopausal women (p = .002). In addition, two univariate ANOVAs confirmed that the three groups were significantly different on both the physical and mental components of the SF-36 [$F_{(2, 168)}$ = 4.01, p = .020, η^2 = .05 and $F_{(2, 168)}$ = 5.12, p = .007, η^2 = .06]. For the physical component, the post-hoc comparisons indicated that the scores of perimenopausal women were significantly lower than those of premenopausal women (p = .026) and marginally lower than those of postmenopausal women (p = .069). Similarly, for the mental component, the scores of perimenopausal women were significantly lower than those of postmenopausal women (p = .007) and the scores of premenopausal women were marginally lower than those of postmenopausal women (p = .080).

CONCLUSIONS

The present study examined the psychological and health correlates of body dissatisfaction in a sample of 172 Italian women between 45 and 55 years of age – a period of life marked, for most women, by the transition from fertility to menopause. Several interesting results emerged. First, the correlational and regression analyses confirmed that, controlling for age and menopausal status, high levels of body dissatisfaction in midlife women predicted increases in mood disorders and anger expression, as well as decreases in mental health and the ability to control angry feelings. Such conclusions are well in line with similar results observed, among the others, by Ganem, de Heer and Morera (2009) and Simbar et al. (2020). Ganem and colleagues (2009) conducted a survey on 174 US university students ranging in age from 17 to 39 years and found that body dissatisfaction was a significant negative predictor of life satisfaction, self-esteem, depression and psychological well-being. Interestingly, with exception of the last variable, the body dissatisfaction by sex interactions were also significant, suggesting a stronger negative impact of body dissatisfaction on women (rather than on men). The more recent study by Simbar et al. (2020) focused specifically on midlife women experiencing the menopausal transition. Results showed that women who were more satisfied with their bodies were less likely to exhibit depression or anxiety problems, even after controlling for several potentially confounding variables (age, duration of menopause, occupation, housing situation and the adequacy of monthly household income). Thus, the common recommendation emerging from this field of research is that psychological interventions in midlife women should be specifically focused at improving acceptance of the physiological changes occurring to their bodies during the menopausal transition and promoting a correct, positive perception of self-body image (Simbar et al., 2020).

Second, our analyses showed that the overall levels of body dissatisfaction, as measured by the BUT-GSI, did not vary among pre-, peri- and postmenopausal women, after

eliminating age differences. As discussed above, the relation between body image and the menopausal transition is far from being clear, with different studies highlighting both negative and positive consequences (Pearce et al., 2014). Our data lean clearly towards the idea that body image remains fairly constant during the menopausal period and are therefore consistent with previous evidence reported, among the others, by Koch et al. (2005) and Mangweth-Matzek et al. (2021). In particular, the latter study demonstrated that premenopausal, perimenopausal, and postmenopausal women (classified with the same method adopted in the present study) reported similar levels of eating and body-image disorders. In contrast, menopausal symptomatology was robustly associated with shape and weight concerns, such that women suffering more severe symptoms were also more likely to be dissatisfied with their bodies. The authors concluded that menopausal stage may be less informative than menopausal symptoms as a predictor of eating and body image disorders. In this respect, our study was limited by the fact that we relied solely on the STRAW classification and did not evaluate menopausal symptomatology. On the other hand, a positive aspect is that we removed from our multivariate and univariate analyses the potentially confounding effects of aging – a procedure which has not been always applied in previous studies (e.g., Séjourné et al., 2019). By doing so, we showed that apparent differences in weight phobia and body image concerns fell below the significance level after controlling for age. The finding that the importance of body shape, weight and appearance decreased in older (postmenopausal) women, as compared to younger (premenopausal) women, was indeed expected (McLaren et al., 2003, Tiggemann, 2004) and could led researcher to incorrectly attribute group differences to menopausal stage, rather than to aging.

A third relevant finding was that, controlling for age, perimenopausal women: a) felt more depressed, fatigued and anxious in the POMS, b) were less able to control angry feelings and tended to express them more frequently in the STAXI, and c) had significantly lower scores of physical functioning, social functioning and mental health in the SF-36, compared to either premenopausal or postmenopausal women. These results corroborate previous findings obtained by other groups, showing that perimenopause represents a particularly difficult period for midlife women. Specifically, Avis et al. (1994) showed that experiencing a long period of perimenopause was associated with a greater risk of depression. Regarding health problems, Mishra, Brown and Dobson (2003) reported that women whose perimenopausal stage had lasted two years experienced larger declines in physical functioning than all other groups. As illustrated by Robinson (2001), at least three different hypotheses have been put forward to explain the increased occurrence of mood disorders in perimenopausal women (Woods & Mitchell, 1997). The first hypothesis assumes that depression would result from fluctuations in gonadal hormones levels leading to decreases in neural transmitters (Halbreich & Kahn, 2001). For the second hypothesis, mood disorders would be the consequence of concomitant increases in the frequency of vasomotor symptoms, such as night sweats and hot flashes (Kurpius et al., 2001); lastly, for the third hypothesis, depression in perimenopausal women would be linked to negative attitudes and expectations concerning menopause (Lyons, 2000; McKinley & Lyon, 2008). Furthermore, the role of stressful life events occurring during the menopausal years and the lack of social support cannot be ignored (Kaufert et al., 1992; Kurpius et al., 2001). The correlational nature of our study does not allow us to adjudicate between these explanations; however, the present findings converge with previous ones in suggesting that changes occurring during the perimenopausal period may be a source of heightened stress, as also demonstrated by increases in counseling referrals for psychological problems (Hamilton, Parry & Blumenthal, 1988).

Clearly, the present study has limitations that must be taken into account. First, the cross-sectional nature of our methodology did not allow us to conclude that body dissatisfaction was causally related to the emotional troubles exhibited by menopausal women. Second, the overall sample size was relatively low and the distribution of women across the three menopausal stage was uneven (i.e., perimenopausal women were underrepresented). Third, we did not measure several variables that are known to have a significant impact on women's mental health, such as appearance-related attitudes toward menopause and aging (e.g., McKinley & Lyon, 2008). Despite these limitations, our results suggest that body dissatisfaction might represent a prominent factor of emotional distress in midlife women, over and above the influence of age and menopausal stage. Additional longitudinal research will help us to further assess the nature of the relationship between body image and psychological wellbeing in women experiencing the menopausal transition.

References

APOLONE, G. & MOSCONI, P. (1998). The Italian SF-36 Health Survey: Translation, validation and norming. *Journal of Clinical Epidemiology*, 51 (11), 1025-1036.

AVIS, N.E., BRAMBILLA, D., McKINLAY, S.M. & VASS, K. (1994). A longitudinal analysis of the association between menopause and depression. Results from the Massachusetts Women's Health Study. *Annals of Epidemiology*, *4*, 214-220.

AVIS, N.E., ORY, M., MATTHEWS, K.A., SCHOCKEN, M., BROMBERGER, J. & COLVIN, A. (2003). Health-related quality of life in a multiethnic sample of middle-aged women: Study of Women's Health Across the Nation (SWAN). *Medical Care*, *41*, 1262-1276.

BAKER, A., SIMPSON, S. & DAWSON, D. (1997). Sleep disruption and mood changes associated with menopause. *Journal of Psychosomatic Research*, *43*, 359-369.

BECKER, C.B., VERZIJL, C.L., KILPELA, L.S., WILFRED, S.A. & STEWART, T. (2019). Body image in adult women: Associations with health behaviors, quality of life, and functional impairment. *Journal of Health Psychology*, *24* (11), 1536-1547.

BROMBERGER, J.T., MATTHEWS, K.A., SCHOTT, L.L., BROCKWELL, S., AVIS, N.E., KRAVITZ, H.M., ... & RANDOLPH Jr, J.F. (2007). Depressive symptoms during the menopausal transition: the Study of Women's Health Across the Nation (SWAN). *Journal of Affective Disorders*, *103* (1-3), 267-272.

BROMBERGER, J.T., MEYER, P.M., KRAVITZ, H.M., SOMMER, B., CORDAL, A., POWELL, L., GANZ, P.A. & SUTTON-TYRRELL, K. (2001). Psychologic distress and natural menopause: A multiethnic community study. *American Journal of Public Health*, *91*, 1435-1442.

BROMBERGER, J.T., SCHOTT, L.L., AVIS, N.E., CRAWFORD, S.L., HARLOW, S.D., JOFFE, H., ... & MATTHEWS, K.A. (2019). Psychosocial and health-related risk factors for depressive symptom trajectories among midlife women over 15 years: Study of Women's Health Across the Nation (SWAN). *Psychological Medicine*, *49* (2), 250-259.

BRYANT, C., JUDD, F.K. & HICKEY, M. (2012). Anxiety during the menopausal transition: A systematic review. *Journal of Affective Disorders*, 139 (2), 141-148.

CARRARD, I., ARGYRIDES, M., IOANNOU, X., KVALEM, I.L., WALDHERR, K., HARCOURT, D. & McARDLE, S. (2021). Associations between body dissatisfaction, importance of appearance, and aging anxiety with depression, and appearance-related behaviors in women in mid-life. *Journal of Women & Aging*, 33 (1), 70-83.

COMUNIAN, A.L. (2004). STAXI-2: State-*Trait Anger Expression Inventory – 2: Manuale*. Firenze, Italy: O.S. Organizzazioni Speciali.

CUZZOLARO, M., VETRONE, G., MARANO, G. & GARFINKEL, P.E. (2006). The Body Uneasiness Test (BUT): Development and validation of a new body image assessment scale. *Eating and Weight Disorders*, *11*, 1–13.

DEEKS, A.A. & McCABE, M.P. (2001). Menopausal stage and age perceptions of body image. *Psychology and Health*, *16*, 367-379.

DENNERSTEIN, L. (1996). Well-being, symptoms and the menopausal transition. *Maturitas*, *23*, 147-157.

FARNÈ, M., SEBELLICO, A., GNUGNOLI, D. & CORALLO, A. (1991). POMS: *Profile of Mood States. Adattamento italiano*. Firenze, Italy: O.S. Organizzazioni Speciali.

GANEM, P.A., DE HEER, H. & MORERA, O.F. (2009). Does body dissatisfaction predict mental health outcomes in a sample of predominantly Hispanic college students? *Personality and Individual Differences*, *46* (4), 557-561.

GANNON, L. & STEVENS, J. (1998). Portraits of menopause in the mass media. *Women Health*, *27*, 1-15.

HALBREICH, U. & KAHN, L.S. (2001). Role of estrogen in the aetiology and treatment of mood disorders. *CNS Drugs*, *15* (10), 797-817.

HAMILTON, J.A., PARRY, B.L. & BLUMENTHAL, S.J. (1988). The menstrual cycle in context I: Affective syndromes associated with reproductive hormonal changes. *Journal of Clinical Psychiatry*, 49, 474-480.

JACKSON, K.L., JANSSEN, I., APPELHANS, B.M., KAZLAUSKAITE, R., KARAVOLOS, K., DUGAN, S.A., ... & KRAVITZ, H.M. (2014). Body image satisfaction and depression in midlife women: the Study of Women's Health Across the Nation (SWAN). *Archives of Women's Mental Health*, *17* (3), 177-187.

KAUFERT, P.A., GILBERT, P. & TATE, R. (1992). The Manitoba Project: A re-examination of the link between menopause and depression. *Maturitas*, *14* (2), 143-155.

KESSLER, R.C., MCGONAGLE, K.A., SWARTZ, M., BLAZER, D.G. & NELSON, C.B. (1993). Sex and depression in the National Comorbidity Survey 1: Lifetime prevalence, chronicity, and recurrence. *Journal of Affective Disorders*, *29*, 85-96.

KOCH, P.B., MANSFIELD, P.K., THURAU, D. & CAREY, M. (2005). "Feeling frumpy": The relationships between body image and sexual response changes in midlife women. *Journal of Sex Research*, *42* (3), 215-223.

KURPIUS, S.E.R., NICPON, M.F. & MARESH, S.E. (2001). Mood, marriage, and menopause. *Journal of Counseling Psychology*, *48*, 77-84.

LYONS, A.C. (2000). Examining media representations: Benefits for health psychology. *Journal of Health Psychology*, *5* (3), 349-358.

MANGWETH-MATZEK, B., RUPP, C.I., VEDOVA, S., DUNST, V., HENNECKE, P., DANIAUX, M. & POPE, H.G. (2021). Disorders of eating and body image during the menopausal transition: Associations with menopausal stage and with menopausal symptomatology. *Eating and Weight Disorders*, *26* (8), 2763-2769.

McKINLEY, N.M. & LYON, L.A. (2008). Menopausal attitudes, objectified body consciousness, aging anxiety, and body esteem: European American women's body experiences in midlife. *Body Image*, *5*, 375-380.

McLAREN, L., HARDY, R. & KUH, D. (2003). Women's body satisfaction at midlife and lifetime body size: A prospective study. *Health Psychology*, *22*, 370-377.

McNAIR, D.M., LORR, M. & DROPPLEMAN, L.F. (1981). *Manual for the Profile of Moods States*. San Diego, CA: Educational and Industrial Testing Service.

MEDEIROS DE MORAIS, M.S., ANDRADE DO NASCIMENTO, R., VIEIRA, M.C.A., MOREIRA, M.A., CÂMARA, S.M.A.D., CAMPOS CAVALCANTI MACIEL, Á. & ALMEIDA, M.D.G. (2017). Does body image perception relate to quality of life in middle-aged women? *PloS One*, *12* (9), e0184031.

MISHRA, G.D., BROWN, W.J. & DOBSON, A.J. (2003). Physical and mental health: Changes during menopause transition. *Quality of Life Research*, *12*, 405-412.

MOND, J., MITCHISON, D., LATNER, J., HAY, P., OWEN, C. & RODGERS, B. (2013). Quality of life impairment associated with body dissatisfaction in a general population sample of women. *BMC Public Health*, 13 (1), 1-11.

PEARCE, G., THØGERSEN-NTOUMANI, C. & DUDA, J. (2014). Body image during the menopausal transition: A systematic scoping review. *Health Psychology Review*, *8* (4), 473-489.

QUITTKAT, H.L., HARTMANN, A.S., DÜSING, R., BUHLMANN, U. & VOCKS, S. (2019). Body dissatisfaction, importance of appearance, and body appreciation in men and women over the lifespan. *Frontiers in Psychiatry*, *10*, 864.

ROBINSON, G.E. (2001). Psychotic and mood disorders associated with the perimenopausal period. *CNS Drugs*, *15* (3), 175-184.

RUNFOLA, C.D., VON HOLLE, A., PEAT, C.M., GAGNE, D.A., BROWNLEY, K.A., HOFMEIER, S.M. & BULIK, C.M. (2013). Characteristics of women with body size satisfaction at midlife: Results of the Gender and Body Image (GABI) study. *Journal of Women & Aging*, *25* (4), 287-304.

SÉJOURNÉ, N., GOT, F., SOLANS, C. & RAYNAL, P. (2019). Body image, satisfaction with sexual life, self-esteem, and anxiodepressive symptoms: A comparative study between premenopausal, perimenopausal, and postmenopausal women. *Journal of Women & Aging*, *31* (1), 18-29.

SIMBAR, M., NAZARPOUR, S., ALAVI MAJD, H., DODEL ANDARVAR, K., JAFARI TORKAMANI, Z. & ALSADAT RAHNEMAEI, F. (2020). Is body image a predictor of women's depression and anxiety in postmenopausal women? *BMC Psychiatry*, *20* (1), 1-8.

SOULES, M.R., SHERMAN, S., PARROTT, E., REBAR, R., SANTORO, N., UTIAN, W. & WOODS, N. (2001). Executive summary: Stages of reproductive aging workshop (STRAW). *Climacteric*, *4* (4), 267-272.

SPIELBERGER, C.D. (1988). *Manual for the State-Trait Anger Expression Inventory (STAXI)*. Odessa, FL: Psychological Assessment Resources.

STEVENS, C. & TIGGEMANN, M. (1998). Women's body figure preferences across the life span. *Journal of Genetic Psychology*, *159*, 94-102.

TANG, R., LUO, M., LI, J., PENG, Y., WANG, Y., LIU, B., ... & CHEN, R. (2019). Symptoms of anxiety and depression among Chinese women transitioning through menopause: Findings from a prospective community-based cohort study. *Fertility and Sterility*, *112* (6), 1160-1171.

TIGGEMANN, M. (2004). Body image across the adult life span: Stability and change. *Body Image*, *1*, 29-41.

VIVIAN-TAYLOR, J. & HICKEY, M. (2014). Menopause and depression: Is there a link? *Maturitas*, *79* (2), 142-146.

WARE, J.E., SNOW, K.K., KOSINSKI, M. & GANDEK, B. (1993). *SF-36 health survey manual and interpretation guide*. Boston, MA: The Health Institute, New England Medical Centre.

WOODS, N.F. & MITCHELL, E.S. (1997). Pathways to depressed mood for midlife women: Observations from the Seattle midlife women's health. *Research in Nursing and Health*, *20*, 119-129.

WOODS, N.F, SMITH-DIJULIO, K., PERCIVAL, D.B., TAO, E.Y., MARIELLA, A. & MITCHELL, S. (2008). Depressed mood during the menopausal transition and early postmenopause: observations from the Seattle Midlife Women's Health Study. *Menopause*, *15*, 223–232.

DOI: 10.26387/bpa.2023.00003

Management strategies for teacher performance evaluation in Colombian schools

Ronald Miguel Linero-Racines[1], Aldair Ricardo Morales-Cuadro[1], Noris Navarro-Yepes[1], Carolina Mercado-Porras[2], Roberto Carlos Buelvas-Salas[3]

[1] Department of Social Sciences, Psychology Program
[2] Department of Humanities, Corporación Universidad de la Costa, CUC, Colombia
[3] Institución Educativa Concejo Municipal

robertbuelvas16@hotmail.com

ABSTRACT. L'obiettivo di questo studio è stato quello di analizzare la relazione tra le strategie manageriali attuate dai direttori scolastici e le valutazioni delle prestazioni didattiche nelle scuole del comune di Soledad, Atlántico, Colombia. Lo strumento di indagine consisteva in un questionario di tipo Likert composto da 66 item, con la scala di misurazione: (5) sempre, (4) quasi sempre, (3) qualche volta, (2) quasi mai e (1) mai. La popolazione era composta da 14 direttori scolastici e 32 insegnanti. È stata riscontrata una correlazione statisticamente significativa tra le strategie di gestione e le valutazioni delle prestazioni degli insegnanti, con una correlazione tra le variabili pari a .82 al livello di significatività di .05. Si conclude che, nella misura in cui le strategie di gestione sono adeguatamente utilizzate, si potrà ottenere un'adeguata implementazione dei criteri e dei processi di valutazione delle prestazioni degli insegnanti.

SUMMARY. The objective of this study was to analyze the relationship between the managerial strategies implemented by school directors and the teaching performance evaluations at schools in the municipality of Soledad, Atlántico, Colombia. The study is quantitative, correlational descriptive, with an ex post facto, cross-sectional, non-experimental design. The survey instrument consisted of a Likert-type questionnaire, whose contents were validated by a panel of five subject-matter experts and a Cronbach's alpha reliability test (rtt = .978). The instrument is made up of 66 items, with the following measurement scale: (5) always, (4) almost always, (3) sometimes, (2) almost never and (1) never. The population consisted of 14 directors and 32 teachers. As a result, shortcomings were found in management strategies and teacher performance evaluations. A statistically significant correlation was found between management strategies and teacher performance evaluations, with a correlation between the variables of .82 at the .05 level of significance. In this sense, it is concluded that, to the extent that management strategies are adequately used, an adequate implementation of the criteria and processes for evaluating teaching performance will be achieved.

Keywords: *Management strategies, Teacher performance, Management personnel*

INTRODUCTION

In Latin America, school directors face several difficulties when it comes to conducting the evaluation of teaching performance at public and private schools, due to the absence of a useful and verifiable model to assess the practice of teachers in the classroom in an effective and constructive way to ensure quality education. At the same time, the quality of education, and therefore the evaluation of the different educational processes, instances and actors, is a growing and persistent concern in diverse educational systems, both nationally and internationally.

In modern organizations, transformation is a necessity since globalism and competitiveness in an ever more complex world produce constant changes. It is therefore essential to implement management strategies that incorporate innovative elements that promote a new culture among the staff, in which diverse ways of acting and behaviors will characterize and determine the quality of the organization and its development.

In this context, the application and use of management strategies that empower directors to decide and act with responsibility and commitment for the achievement of objectives and goals is beneficial for the strengthening of organizations.

In this regard, school administrators must ensure the quality of the institution, which implies that the director must project a series of values, skills, and attributes of resiliency that allow him/her to interact with the academic community. Furthermore, from his role, he must guide, evaluate and direct the teaching and administrative staff to achieve the institutional goals. He must also ensure that training is provided to the teaching staff, to promote the development of competent leaders who can face the challenges of education today (Zea-Vallejo, 2020).

THEORETICAL FOUNDATIONS

Types of management strategies

Strategies are the actions that must be conducted to maintain and support the achievement of the objectives of an organization and its work units, and thus make the expected results of strategic projects a reality (García Núñez, 2018). In this sense, the school director, by adopting a strategy as a procedure to cope with the changes that arise in the educational system, will be able to make viable the alternatives that are considered necessary to assimilate and project managerial, academic, and cultural changes in the school environment (Bravo & Salazar, 2017).

Chiavenato (2013, 2014) states that management strategies involve long-term goals, the adoption of courses of action and the allocation of resources necessary to achieve the objectives. Management strategies, according to Rueda Beltrán, Schmelkes & Díaz-Barriga (2014) and Mosquera Maldonado (2020) are based on objective criteria and analysis according to the information available, which requires permanent knowledge of social and cultural changes in the company's external environment. This promotes the commitment, discipline, and willingness of the work teams. González-Díaz & Serrano Polo (2017) propose certain managerial strategies in schools, such as implementing cross-sectional educational projects and training for innovation, to propitiate the development of plans and programs that incorporate individual and collective initiatives, to produce positive changes in the future.

Chiavenato (2014) mentions that to manage the interdependencies between the elements of the work environment, each organization must use different strategies depending on each situation, which may involve coercion and contingencies, requiring the sum of the members of the organization. Mosquera Maldonado (2020) considers that any authentic managerial action must be based on an exhaustive review of the process and of the institution; otherwise, it will only amount to a simple set of instructions.

In the educational context, strategies are based on external opportunities and threats, as well as internal strengths and weaknesses, to effectively arrange the management functions. Liaison strategies are established based on the network of relationships between work teams, which are essential to determine any actions that must be taken in the organization (Parra-Martínez, 2017). In this process, directors must know the institution beyond the limits of their function to understand the interrelationships between the different units and to promote the cooperation of all the actors involved in the educational process (Chiavenato, 2013; Jara, Sánchez & Cox, 2019).

Based on the above, the main function of the school director is to integrate the school with the community and to build strong moral ties among all members of the educational community. To foster this network of relationships, it is necessary to form a community team made up of

management, teachers, students, administrators, workers, civil associations, community leaders and other community stakeholders. A preparatory meeting should include key informants and the institution's personnel, to address the realities of education and the community (Enriquez & Calderon, 2017).

The establishment of work teams in schools requires effective management skills, and especially a display of strong leadership by the director during this process, given that these teams often will not operate on their initiative, although occasionally the teams may be self-directed if one of their members is a persuasive and effective leader (Jara et al., 2019).

Another strategy that is highly effective is delegation. This strategy refers to the ability to entrust to others the performance of tasks for a group purpose, which involves assigning responsibility and authority to subordinates to achieve greater fluidity in work processes (Chiavenato, 2013; Lugo & Villasmil Ferrer, 2019).

Consensus-building is yet another type of management strategy. It involves a commitment to values and far-reaching projects both by the institution and the community members, to reach agreements at the operational level. However, it should not be assumed that the consensus is derived directly from the common values; it is rather a process of clarification and dialogue that enables the actors to settle their differences (Moreno Trejo, 2019).

Therefore, the basic function of the school director's management strategy is to improve the results of the teaching process, i.e., to achieve the ever-increasing improvement of the quality of the final product of the mentioned process, in an organic, systematic, organized, and continuous way, throughout the educational process (Chiavenato, 2014; Robbins & Coulter, 2014; Sotelo & Arrieta, 2017). Thus, knowledge management has been defined as an art in which information and intellectual assets are transformed into lasting value for an organization and its customers.

Authors such as Alles (2015) not only define and study the subject of competencies, but also classify them as follows: basic/generic competencies (capabilities required to achieve personal desires, live harmonizing with society and know how to act at work or at a professional level) and specific competencies (skills specific to a given profession or occupation, having a high level of specialization achieved through technical programs, training for work and university studies).

The competencies of teachers, the efficiency of their performance and the development of educational and curricular programs are key to the improvement of educational quality. It should be noted that the teachers' performance is favored by collaboration with each other, since it helps them to detect and make decisions about the strengths and weaknesses of their practice intending to make continuous improvements (Obreque, Hernández-Mosqueira, Troncoso-Peña, Agredo & Salvatierra, 2019).

The evaluation of teaching performance is the formulation of judgments on standards, structures, processes, and products to make any corrections that are necessary and convenient for the most efficient achievement of institutional objectives. Thus, it can be understood as a necessary process within an institution (Galaz Ruiz, Jiménez-Vásquez & Díaz-Barriga, 2019). Organizations need to know how employees are performing their jobs, to identify who effectively adds value and who does not. For this purpose, performance evaluations are conducted, which enable to assign ratings to employees to differentiate between effective and ineffective employees.

Colombian law regulates the annual teacher performance evaluations, establishing rules on 14 aspects of performance to be assessed: construction and development of the Institutional Educational Project (PEI), compliance with educational standards and policies, knowledge and valuation of students, pedagogical foundation, work planning, pedagogical strategies, strategies for participation, evaluation and improvement, innovation, institutional commitment, interpersonal relations, conflict mediation, teamwork, and leadership (MEN, 2016).

Under this premise, the Colombian Ministry of National Education (MEN, 2016) set out the objectives of performance evaluations, such as encouraging commitments to professional development, performance and continuous training. In this manner, the didactic sequences learned by the teacher and those that he/she implements go hand in hand with the series of sequenced activities focused on learning. (Ávila-Camacho, Juárez-Hernández, Arreola-González & Palmares- Villarreal, 2019; González-Díaz & Serrano Polo, 2017; Guzmán, 2016; Rockwell, 2018).

METHOD

This is a quantitative study conducted with an *ex post facto* design of cross-sectional measurement aimed at analyzing the relationship between variables such as the

managerial strategies employed by the management staff of schools and the processes to evaluate teaching performance by teachers, employed by the management staff in the schools of the municipality of Soledad (Barranquilla-Colombia) (Hernández-Sampieri & Torres, 2018; Montero & León, 2007).

Population and sample

Population: fourteen (14) directors and thirty-two (32) teachers assigned to the four (4) middle schools of the municipality of Soledad, Atlántico, Colombia.

Technique and instruments

This study used a structured survey as technique and a Likert-type scale as instrument. For content validity, the criteria of 5 subject-matter experts were taken into considered, after the review and recommendation of the expert peers, the variables were operationalized as can be depicted on Table 1.

After reviewing and validating the original data collection instrument, the pertinent modifications were made based on the observations, thus generating a definitive version of the instrument to be applied to the population under study. Following are the most used alternatives and values: (5) always, (4) almost always, (3) occasionally, (2) almost never, (1) never.

The final instrument has 66 items and five multiple-choice alternatives. A pilot test was applied by selecting 15 study subjects outside the research sample; the questionnaire was administered to them at their workplace.

After the pilot evaluation, reliability was evaluated using Cronbach's alpha, with a value of .97. The above indicates that the scale is dependable for managerial and teaching staff in the Colombian population, as shown in the Table 2.

Statistical analysis

The statistical data and tables were processed using the statistical package SPSS version 21; the results are presented in percentages and the analysis and interpretation is performed employing comparison of percentages, based on the summary of the responses given by the directors (DIR) and teachers (DOC) for each instrument item. The data analysis is descriptive. The association of variables was calculated with Rho Spearman, which takes the value of +1 when there is equality of ranks in the two variables, and –1 when there are exactly opposite ranks.

Procedure

- Theoretical review of the variables: in this phase we proceeded to review the bibliography of the variables of classic and current theoretical references, to be able to create the items corresponding to the subscales.
- Design of the items, subscales, types of management strategies, strategic competencies, and criteria of the teacher performance evaluation.
- Analysis of the items through a quantitative analysis, to then tabulate the responses obtained from the instrument.
- Analysis of internal and external consistency, calculating the values of central tendency and dispersion.
- Application of the instrument to the selected population: in this phase, public and private educational institutions will be chosen, where a visit will be made to explain what the project consists of, as well as to socialize with the teaching staff and school administrators about informed consent as depicted on Table 3.
- Analysis of results: in relation to the statistical treatment, the statistical application software SPSS version 21 and the application program Microsoft Excel 2010 were used to perform descriptive statistics, using the analysis of the distribution of absolute and percentage frequency of the data and the association of variables with Spearman's Rho. After generating the results, we proceeded to conduct a socialization workshop type sensitization in the facilities of the I.E.D for 1 hour and 30 minutes on the data found in the sample study, this workshop will be directed to teachers and teachers' managers of the District Educational Institutions.

RESULTS

Table 4 shows that for the variable Management strategies in the dimension Types, the directors (DIR) report that they use liaison strategies almost always (55.56%), always (22.22%) and sometimes (22.22%). On the other hand, the teachers

Table 1 – General objective*

Specific objectives	Variable	Dimension	Indicators
To characterize the managerial strategies used by the directive personnel of the Educational Institutions of the Municipality of Soledad, Atlántico, Colombia.		Types of management strategies	Linking strategies Delegation strategies Management strategies
To identify the strategic competencies of the principal of the Educational Institutions of the Municipality of Soledad, Atlántico, Colombia.	*Management strategies*	Strategic competencies	*Cardinal competencies* Commitment Ethics Organizational awareness Initiative Perseverance *Specific managerial competencies* Team development Leadership for change Strategic thinking Quality of work Planning skills Effective communication
To describe the criteria of the evaluation of the teaching performance applied by the management staff of the educational institutions of the Municipality of Soledad, Atlántico, Colombia.		Teacher performance evaluation criteria	Feasibility Precision Utility Ethics
To identify the processes to evaluate the teaching performance of the teachers, used by the directive personnel in the Educational Institutions of the Municipality of Soledad, Atlántico, Colombia.	*Teacher performance evaluation*	Processes for performance evaluation	Job definition Performance Measurement Feedback
Relating management strategies and teacher performance evaluation in educational institutions in the municipality of Soledad, Atlántico, Colombia.	They arise from the applications of Spearman's correlation coefficient		

Note. * To establish the relationship between the managerial strategies implemented by the management personnel and the evaluation of the teaching performance in the Educational Institutions located in the Municipality of Soledad.
Source: Buelvas-Salas, Mercado-Porras, Navarro-Yepes, Morales-Cuadro & Linero-Racines, 2021.

Table 2 – Cronbach's alpha interpretation scale

Interval relative frequency (%)	Range	Category Level of presence /Dominance/Management
.81-1.00	I	Very high
.61-.80	II	High
.41-.60	III	Under
.21-.40	IV	Very low
.01-.20	V	Extremely low

Note. Source: Buelvas-Salas, Mercado-Porras, Navarro-Yepes, Morales-Cuadro & Linero-Racines, 2021.

Table 3 – Distribution of the target population

Secondary Education Institutions	Directors	Teachers
Institución Técnico Industrial el Milagroso (Soledad, Atlántico, Colombia)	6	12
Colegio Metropolitano de Soledad, 2000	4	10
Colegio Madre Marcelina	4	10
Total	**14**	**32**

Note. Source: Buelvas-Salas, Mercado-Porras, Navarro-Yepes, Morales-Cuadro & Linero-Racines, 2021.

Table 4 – Dimension: Types of Management strategies

	Always		Almost always		Sometimes		Almost never		Never	
Indicators	DIR	DOC	DIR	DOC	DIR	DOC	DIR	DOC	DIR	DOC
	%	%	%	%	%	%	%	%	%	%
Liaison strategies	22.22	4.23	55.56	28.04	22.22	44.97	0	20.38	0	2.38
Delegation strategies	0	1.85	33.33	26.98	44.45	47.62	22.22	23.02	0	0.53
Management strategies	5.56	2.38	11.11	22.75	72.22	47.88	11.11	24.87	0	2.12
Average %	9.26	2.82	33.33	25.92	46.3	46.82	11.11	22.76	0	1.68

Legenda. DIR = directors; DOC = teachers.
Note. Source: Buelvas-Salas, Mercado-Porras, Navarro-Yepes, Morales-Cuadro & Linero-Racines, 2021.

(DOC) believe that the director uses this strategy sometimes (44.97%), almost always (28.04%), almost never (20.38%) and always (4.23%).

Regarding the indicator Delegation strategies, the directors claim that they use this management strategy sometimes (44.45%), and almost always (33.33%), but none of the directors say that they always apply delegation strategies. The teachers tend to agree with the directors, stating that the directors use delegation strategies sometimes (47.62%), almost always (26.98%) and almost never (23.02%).

Regarding the Management-building strategies indicator, most directors (72.22%) indicated that they sometimes use this type of strategy, and 11.11% stated that they almost always use it; the same percentage claimed that they almost never used it and only 5.56% opted for the option always. In the case of teachers, 47.88% stated that the directors put consensus-building strategies into practice sometimes, 24.87% almost never, 22.75% almost always, 2.38% always and 2.12% never.

It can be observed that 46.3% of the directors believe they use the types of managerial strategies sometimes, which is similar to 46.82% of the teachers. In this sense Puentes (2017) suggests that, since they are collaborating with people, the directors' guiding principle should be to instill the desire for improvement by recognizing the teachers' efforts and achievements in terms of the quality of their work, to improve in this way overall teaching quality at the institution.

Table 5 shows the percentages of responses given by directors (DIR) and teachers (DOC) for the indicators of the Key competencies dimension. Regarding the Commitment indicator, it is observed that the majority of directors, with 55.56%, think that they sometimes manage this key competency, an opinion that is shared by most of the teachers, with 47.09%. Moreover, 22.22% of directors report that they always manage this competency, and an additional 22.22% say they almost always do so, thus showing a positive trend for this indicator for the directors. The teachers do not display a unified tendency, given that their answers for this indicator

Table 5 – Dimension: Key competencies

	Always		Almost always		Sometimes		Almost never		Never	
Indicators	DIR	DOC	DIR	DOC	DIR	DOC	DIR	DOC	DIR	DOC
	%	%	%	%	%	%	%	%	%	%
Commitment	22.22	4.24	22.22	26.19	55.56	47.09	0	21.16	0	1.32
Ethics	22.22	5.29	27.78	31.48	33.33	42.33	16.67	19.05	0	1.85
Innovation	6.66	4.23	0	28.84	77.78	46.3	0	18.25	5.56	2.38
Organizational awareness	11.11	7.94	27.78	33.07	44.44	38.36	5.56	19.05	11.11	1.58
Initiative	44.44	8.99	33.33	33.07	11.11	40.21	5.56	16.14	5.56	1.59
Perseverance	44.44	7.67	0	35.45	55.56	40.48	0	14.55	0	1.85
Average %	26.85	6.39	18.52	31.35	46.3	42.46	4.63	18.03	3.71	1.76

Legenda. DIR = directors; DOC = teachers.
Note. Source: Buelvas-Salas, Mercado-Porras, Navarro-Yepes, Morales-Cuadro & Linero-Racines, 2021.

are distributed into almost always 26.19%, almost never 21.16%, always 4.24% and never 1.32%.

Regarding Organizational awareness as a key competency, 44.44% of the directors report that it is sometimes managed, 27.78% said almost always, 11.11% said always and 11.11% said never; whereas 38.36% of the teachers report that organizational awareness is sometimes managed, followed by 33.07% who said almost always and 19.05% almost never. For 44.44% of the directors, initiative is always managed as a key competence, followed by 33.33% who say it is almost always managed, indicating a favorable tendency towards this indicator for the directors. For teachers, 40.21% maintain that sometimes initiative as a key competency is put into practice, while 33.07% presume almost always managed.

Most directors, 55.56%, say that perseverance is sometimes put into practice, and 44.44% of directors claim it is always put into practice. The teachers say that sometimes (40.48%) perseverance is a key competency, and 35.45% say that it is almost always one.

Regarding the Team development indicator, 44.44% of the directors think that this specific competency is sometimes used as a management strategy, while 38.89% think that it is always used, 11.11% think that it is almost never used, and only 5.56% think that it is almost always used. For 42.86% of the teachers, team development is sometimes practiced, 32.8% think it is almost always used, 15.08% think it is almost never used, 7.41% think it is always used and 1.85% think it is never used.

In relation to the Leadership for change indicator, 33.34% of the directors believe that this specific competency is almost never handled, and an equal distribution of 22.22% is observed for the answers always, almost always and sometimes; for teachers, 41.8% think that sometimes leadership for change is put into practice as a specific competency followed by 33.86% whose opinion is that it is almost always put into practice.

Regarding Strategic thinking, 44.44% of the directors think that it is almost always managed as a specific competency, 27.78% think that it is almost always managed and 16.67% think that it is always managed. For 33.34% of the directors, the quality of work is sometimes considered as a specific competency, followed by an equal distribution of 22.22% for the answers always, almost always and almost never; whereas 44.18% of the teachers think that sometimes the quality of work is considered, followed by 27.78% who think it is almost always and 19.58% who think it is almost never considered.

Regarding Planning capacity, most of the directors think that sometimes, 55.56%, this specific competence is managed, 27.77% think that it is always managed and 11.11% think that it is almost always managed; in relation to the teachers, 39.95% think that it is sometimes, and 27.51% think that it is almost always managed.

For 38.89% of the directors, effective communication is sometimes a specific competency put into practice as a management strategy, and for 22.22% it is almost always practiced; the same opinion is held by the teachers who, with 41.27%, think that effective communication is sometimes managed, while 25.4% think that it is almost always managed. On average, both directors and teachers, 39.82% and 42.15% respectively, believe that sometimes the indicators for specific managerial competencies are put into practice.

Table 6 shows that 27.78% of the directors believe that Feasibility is almost never a criterion for teacher performance evaluation, followed by 22.22% of directors who believe that it is never, 22.22% sometimes and 22.22% almost always a criterion for performance evaluation; for 41.8% of teachers, feasibility is sometimes considered, while 30.16% of teachers believe that it is almost always and 18.78% believe that it is almost never a criterion.

Regarding Accuracy as a criterion for performance evaluation, 44.44% of directors think that it is almost never, while 27.78% think that it is almost always and 22.22% always; on the other hand, 42.06% of teachers think that sometimes accuracy is a criterion for teacher performance evaluation, followed by 26.19% who think that it is almost always and 21.97% almost never a criterion, indicating a large difference of perceptions between directors and teachers.

Regarding the criterion of Usefulness, 44.44% of the directors agree that this criterion is almost always used to evaluate teacher performance, while 38.89% think that it is almost never and 16.67% always; 38.89% of the teachers think that sometimes the criterion of usefulness is used, 28.56% think that it is almost always and 23.02% think that it is almost never used.

Regarding the indicator of Ethics as a performance evaluation criterion, the majority of the directors, 72.22%, maintain that this criterion is almost always used, 16.67% presume that it is sometimes and 11.11% that it is always used, indicating a favorable tendency towards this evaluation criterion for the directors; regarding the teachers, 47.09% believe that sometimes ethics is considered a criterion for the evaluation of teaching performance, 27.51% maintain that it is

Table 6 – Dimension: Performance evaluation criteria

	Always		Almost always		Sometimes		Almost never		Never	
Indicators	DIR	DOC	DIR	DOC	DIR	DOC	DIR	DOC	DIR	DOC
	%	%	%	%	%	%	%	%	%	%
Feasibility	5.56	6.35	22.22	30.16	22.22	41.8	27.78	18.78	22.22	2.91
Accuracy	22.22	6.08	27.78	26.19	5.56	42.06	44.44	21.97	0	3.7
Usefulness	16.67	5.56	44.44	28.56	0	38.89	38.89	23.02	0	3.97
Ethics	11.11	3.17	72.22	27.51	16.67	47.09	0	17.2	0	5.03
Average %	13.89	5.29	41.67	28.11	11.11	42.46	27.78	20.24	5.56	3.9

Legenda. DIR = directors; DOC = teachers.
Note. Source: Buelvas-Salas, Mercado-Porras, Navarro-Yepes, Morales-Cuadro & Linero-Racines, 2021.

almost always and 17.2% that it is almost never a criterion. In terms of overall percentages, on average 41.67% of the directors believe that these indicators are almost always considered as criteria for evaluating teacher performance, while 42.46% of the teachers believe that sometimes these indicators are part of the criteria for evaluating teacher performance.

Regarding the criteria for the evaluation of teaching performance, Hernández-Mosqueda, Tobón-Tobón and Guerrero-Rosas (2016) state that the success of an evaluation largely depends on formulating criteria to guide both the evaluator and the evaluated, for them to know where they should direct their efforts and the reasons why they should perform in a specific way within the organization. In the case of the schools under study, it can be seen that the directors are not monitoring the above evaluation criteria, so they are not contributing to the control of the staff's development in accordance with their duties and competencies required for organizational success.

Table 7 shows the percentage results of the responses for both directors and teachers to measure the indicators of the dimension "processes for evaluating teacher performance" of the teacher performance evaluation variable. It is observed that most directors (55.55%) maintain that the definition of work is almost always a process for evaluating teacher performance, whereas 16.67% say it is sometimes, 11.11% say it is always, and only 16.67% say it is almost never used. A favorable trend is evident for the directors regarding the indicator Job definition; while for the teachers it is observed that 42.33% believe that it is sometimes, 28.83% almost always, and 20.11% almost never used.

Regarding Performance measurement, the majority of directors, 50%, believe that performance measurement is sometimes a process considered for teacher performance evaluation, while 27.78% believe that it is almost always and 16.66% that it is always used, indicating a favorable trend for directors regarding performance measurement. 44.44% of teachers believe that it is sometimes, 32.28% almost always and 15.61% almost never considered.

Regarding Feedback as a process for evaluating teacher performance, 44.44% of directors believe that this process is almost always, 38.89% sometimes and 16.67% almost never used; 42.87% of teachers believe that it is used sometimes, 29.89% almost always and 18.78% almost never. In terms of overall percentages, both directors and teachers, with 35.19% and 43.21% respectively, think that on average sometimes these processes (work definition, performance measurement and feedback) are considered for the evaluation of teacher performance.

Table 7 – Dimension: Processes for evaluating teacher performance

	Always		Almost always		Sometimes		Almost never		Never	
Indicators	DIR	DOC	DIR	DOC	DIR	DOC	DIR	DOC	DIR	DOC
	%	%	%	%	%	%	%	%	%	%
Job definition	11.11	4.23	55.55	28.83	16.67	42.33	16.67	20.11	0	4.5
Performance measurement	16.66	3.97	27.78	32.28	50.00	44.44	5.56	15.61	0	3.7
Feedback	0	4.23	44.44	29.89	38.89	42.87	16.67	18.78	0	4.23
Average %	9.26	4.14	42.59	30.33	35.19	43.21	12.97	18.17	0	4.14

Legenda. DIR = directors; DOC = teachers.
Note. Source: Buelvas-Salas, Mercado-Porras, Navarro-Yepes, Morales-Cuadro & Linero-Racines, 2021.

The results coincide with Núñez Rojas and Díaz Castillo (2017), who state that it is convenient to explain to employees what their obligations and duties will be within the organization, and that these must be accepted by them. Then the director will be able to evaluate by measuring the performance of each individual according to the predefined guidelines, and finally the feedback of the results will be made to employees, referring to both deficiencies and achievements.

To establish the degree of the relationship between the variables, a non-parametric correlation coefficient was used, in this case Spearman's Rho coefficient, which at .829 is an acceptable value, significant at the alpha level of .05 and represents a remarkably high correlation between the variables.

DISCUSSION AND CONCLUSIONS

It was found that school directors display weaknesses in the use of liaison strategies for the effects of enhancing school-community integration. Regarding management strategies, deficient implementation was found of strategies that facilitate the acquisition of knowledge by teachers, which means that the management strategies and educational activities do not guarantee positive changes in the institution. The key competencies of the teachers were found to be weak, given that there are failures in how the directors exercise the routes for the fulfillment of the goals, the values promoted by the institution and the contribution of suggestions to improve the efficiency of the organization.

Another weakness found was in the implementation of the key competencies mentioned above. Weaknesses were also found in the establishment of criteria for the evaluation of teaching performance, failures in the implementation of the criteria for the evaluation of teaching performance, and failures in the implementation of the feasibility, accuracy, usefulness, and ethics criteria. Likewise, the way in which the evaluation of teaching performance is conducted has considerable flaws, given that the knowledge of teachers about their duties within the institution is taken for granted.

A high correlation was found between the variables of management strategies and teacher performance evaluations. This implies that to the extent that management strategies are used appropriately, the criteria and processes for evaluating teaching performance will be properly implemented. Therefore, it is recommended to improve management, evaluation and integrating strategies, which allows strengthening institutional effectiveness and interpersonal relationships among teachers.

Limitations

During the field work, it was possible to evidence the shortcomings presented by the teachers' managers regarding the implementation of accuracy, usefulness and ethics, as well as in the actions to evaluate teachers, in this sense, teachers presented some resistance to the application of the instrument, therefore, when interviewing them it was possible to evidence the communication problems between teachers' managers and teachers, especially about their commitments within the institution, which could affect organizational awareness, initiative, and insistence on teachers in not having clarity for managers on the aforementioned practices.

Recommendations

Considering the conclusions of the study, a series of pertinent recommendations are derived to comply with the principles of usefulness and recognition of the subjects observed, as well as for the organizations considered as a field of action in the realization of the study:

- Promote strategies of liaison, delegation, and management based on educational innovations that reorient the operation of each of the actions carried out by the staff selecting each of them, according to the objectives, as well as the goals,set, to ensure that the institution is of progress, excellence, and educational quality with the best teaching team to ensure an adequate performance evaluation.
- To focus the actions on the strengthening of cardinal and specific managerial competencies, which strengthen an adequate evaluation of the teaching performance, directing, clarifying, and guiding the actions of teachers to encourage staff to a better development of the pedagogical practice.
- To elaborate action strategies that allow for the reorientation of the criteria and processes of teacher performance evaluation conducted by the directors of the institutions consulted so that it may contribute to the personal and professional growth and development of educators, to institutional development, and to school and teaching improvement.

References

ÁVILA-CAMACHO, M., JUÁREZ-HERNÁNDEZ, L, ARREOLA-GONZÁLEZ, A. & PALMARES-VILLARREAL, O. (2019). Construcción y validación de un instrumento de valoración del desempeño docente en la ejecución de una secuencia didáctica. *Revista de Investigación en Educación*, *17* (2), 122-142.

BRAVO, M. & SALAZAR, J. (2017). El clima laboral y su incidencia en el desempeño del personal docente de una escuela de educación básica en Ecuador. *PODIUM*, 131-143.

CHIAVENATO, I. (2013). *Organizational behavior: The dynamics of success in organizations*. Mexico: McGraw Hill.

CHIAVENATO, I. (2014). *Administración en los nuevos tiempos*. Bogotá, Colombia: McGraw Hill.

GALAZ RUIZ, A., JIMÉNEZ-VÁSQUEZ, M. & DÍAZ-BARRIGA, Á. (2019). Teacher performance evaluation in Chile and Mexico: Antecedentes, convergencias y consecuencias de una política global de estandarización. *Perfiles Educativos*, *4* (163), 177-199.

GARCÍA NÚÑEZ, S. (2018). *The process of entrepreneurial constitution and its relationship with the definition of strategies for the establishment of socio-productive relationships, in the textile and clothing industry in Aguascalientes*. Master's thesis, Universidad Autónoma de Aguascalientes, Aguascalientes (Mexico).

GONZÁLEZ-DÍAZ, R. & SERRANO POLO, E. (2017). Managerial strategies for innovation in public educational institutions. *Journal of Latin American Science*, *1* (1), 1-23.

GUZMÁN, J. (2016). ¿Qué y cómo evaluar el desempeño docente? Una propuesta basada en los factores que favorecen el aprendizaje. *Propósitos y Representaciones*, *4* (2), 285-358. doi: http://dx.doi.org/10.20511/pyr2016.v4n2.124

HERNÁNDEZ-MOSQUEDA, J., TOBÓN-TOBÓN, S. & GUERRERO-ROSAS, G. (2016). Towards a comprehensive performance evaluation: Socioformative rubrics. *Ra Ximhai*, *12* (6), 359-376.

HERNÁNDEZ-SAMPIERI, R. & TORRES, CH. (2018). *Metodología de la investigación*, *vol. 4*. Mexico: McGraw-Hill Interamericana.

JARA, C., SÁNCHEZ, M. & COX, C. (2019). Educational leadership and citizenship training: Vision and practices of actors. *Calidad en la Educación*, *51*, 350. doi.org/10.31619/caledu.n51.687

LUGO, N. & VILLASMIL FERRER, J.R. (2019). Managerial leadership as a factor of improvement in educational quality. *EPISTEME KOINONIA, 2* (4), 4. doi.org/10.35381/e.k.v2i4.521

MEN - MINISTRY OF NATIONAL EDUCATION (2016). *Review of national policies in education. Education in Colombia. Organization for Economic Co-operation and Development (OECD)*.

MONTERO, I. & LEÓN, O. (2007). Guía para nombrar los estudios de investigación en psicología. *International Journal of Clinical and Health Psychology*, *7* (3), 847-862.

MORENO TREJO, Y.M. (2019). Educational management versus job satisfaction of the current teacher: An analytical look. *Scientific Magazine*, *4* (12), 369-380. doi.org/10.29394/scientific.issn.2542-2987.2019.4.12.20.369-380

MOSQUERA MALDONADO, M.N. (2020). Managerial strategies for the improvement of human talent management and school resources. *Science and Education*, *1* (3), 6-19. doi.org/10.48169/ecuatesis/0103202014

NÚÑEZ ROJAS, N. & DÍAZ CASTILLO, D. (2017). Perfil por competencias gerenciales en directivos de instituciones educativas. *Estudios Pedagógicos (Valdivia)*, *43* (2), 237-252.

OBREQUE, A., HERNÁNDEZ-MOSQUEIRA, C., TRONCOSO-PEÑA, S., AGREDO, M. & SALVATIERRA, M. (2019). Evaluación del desempeño docente en Chile: Percepción de profesores mal evaluados. *Cadernos de Pesquisa*, *49* (172), 144-163. doi.org/10.1590/198053145792

PARRA-MARTÍNEZ, J. (2017). La gestión eficaz en educación y su importancia en acción gerencial educativa de Colombia. *Aibi Revista de Investigación, Administración e Ingeniería*, *5* (2), 16-22.

PUENTES, J. (2017). Estrategias gerenciales para el reconocimiento del desempeño laboral docente. *Mundo FESC*, *7* (14), 42-56.

ROCKWELL, E. (2018). The complexity of teaching work and the challenges of its evaluation: International results and national processes of educational reform. *Cuadernos de Educación*, *16* (16).

RUEDA BELTRÁN, M., SCHMELKES, S. & DÍAZ-BARRIGA, Á. (2014). La evaluación educativa. Presentación del número especial Perfiles Educativos 2013. La evaluación en la educación superior. *Perfiles Educativos*, *36* (145), 190-204.

SOTELO, J. & ARRIETA, D. (2017). *El capital humano como factor clave para la calidad en las organizaciones*. Mexico: Mc Graw Hill.

ZEA-VALLEJO, D.A. (2020). Organizational behavior of the educational manager in community participation. *Prohominum*, *2* (1), 63-72. doi.org/10.47606/acven/ph0004

DOI: 10.26387/bpa.2023.00001

Measuring the need for cognition: Structural analysis and measurement of invariance of the short version of the Need for Cognition Scale in Italian adolescents

Elisabetta Lombardi[1], Emanuele Aloi[2], Christian Tarchi[3], Daniela Traficante[2]

[1] PhD Associate Professor in Developmental and Educational Psychology, Faculty of Psychology, eCampus University
[2] Department of Psychology, Catholic University of Milan, Italy
[3] Department of "Formazione, lingue, intercultura, letterature e psicologia", University of Florence, Italy

elisabetta.lombardi@uniecampus.it

ABSTRACT. Lo studio di validazione si propone di fornire una versione breve italiana per adolescenti della scala del *Need for Cognition*, strumento che misura le differenze individuali nella motivazione ad apprezzare e a impegnarsi in attività cognitive che richiedono uno sforzo. La scala è composta da 18 item ed è stata somministrata a 473 studenti italiani della scuola secondaria. L'analisi fattoriale confermativa ha dimostrato che la scala presenta due fattori correlati che misurano due diverse dimensioni di motivazione, ovvero l'avvicinamento e l'evitamento delle attività cognitive impegnative. La scala è invariante anche per il genere e per la modalità di somministrazione (online e carta-matita). Grazie alle sue buone proprietà psicometriche, la scala si mostra come uno strumento utile sia in ambito educativo che di ricerca.

SUMMARY. This study aims to adapt and validate the Need for Cognition Scale – Short Version (NFCS; Cacioppo & Petty, 1982) in Italian adolescents. This instrument measures individual differences in the motivation and enjoyment in being involved in effortful cognitive activities. The Italian version of the scale, translated and adapted from the original version, is composed of 18 items and was administered to secondary school students. The confirmatory factorial analysis proved that the scale had two correlated factors measuring two different dimensions of motivations, namely approach and avoidance of effortful cognitive activities. The scale is also invariant for gender and for the administration measurement (on-line and paper-pencil). Results also indicated that the NFCS had good reliability indices and satisfactory convergent and discriminant validity. Thanks to its good psychometric properties, the Need for Cognition Scale has been proven to be a useful tool in both educational and research areas.

Keywords: *Need for cognition, Invariance measurement, Adolescence*

INTRODUCTION

The Need for Cognition (NFC) is a personality variable that reflects the extent to which individuals are inclined to demanding cognitive tasks (Cacioppo & Petty, 1982). The construct was first identified in 1955 by Cohen's research group (Cohen, Stotland & Wolfe, 1955) and, since then, research has extensively documented how the need for cognition influences various cognitive and behavioural factors (Petty, Briñol, Loersch & McCaslin, 2009).

In particular, it has been shown that - concerning problem-solving and decision-making activities - high scores in tests that evaluate NFC are predictive of a high search for additional information and a high generation of possible solutions before making a decision (Petty et al., 2009). Furthermore, subjects who show high levels of NFC, not only have a high chance of approaching problems through cognitively expensive but effective processes, but also demonstrate great awareness of their thought processes and are more likely to correctly evaluate the validity of reasoning (Petty, Briñol & Tormala., 2002). Finally, those with high scores in the construct tend to actively seek new information (Verplanken, Hazenberg & Palenewen, 1992), prefer complex tasks to simple ones (Cacioppo & Petty, 1982), and show good performance in various cognitive activities, such as comprehension of the text (Dai & Wang, 2007) or the decision-making processes (Levin, Hunekeb & Jasper, 2000).

Overall, the literature seems to agree that high levels of NFC are associated with an increased inclination to debate, reflection, thought/opinion pondering and problem-solving. Individuals who exhibit this characteristic are therefore inclined to use analytical/rational elaboration processes. Instead, individuals with a low NFC tend to exhibit opposite tendencies, processing information in a more heuristic way. The individuals low in the need for cognition elaborate on incoming information less than those with a high NFC (Petty & Cacioppo, 1986) and tend to engage in fallacies or biases. For example, framing biases in decision making would more strongly affect individuals with low levels of need for cognition: low NFC persons varied their responses based on the framing of the problem while high NFC persons gave consistent responses independent of framing (Smith & Levin, 1996). Cacioppo and Petty (1982) described the construct as a stable individual difference and developed the *Need for Cognition Scale* (*NFCS*), for its evaluation. More specifically, the two authors have developed, at first, a scale with 34 items. The psychometric properties revealed a high internal consistency characterized by a single dominant factor as resulted from the Principal Component Analysis (PCA) with a Cronbach alpha of .84. Subsequently, Cacioppo and colleagues (Cacioppo, Petty & Feng Kao, 1984) developed a short version with 18 items, based on those items with the highest factor loadings. PCA extracted a single factor that explained the 37% of variance with a high level of consistency (Cronbach alpha of .90).

Many studies have used the *Need for Cognition Scale* either in the original version or in a translated version over the last fifteen years. Most studies used the short version. Many of these studies reported high reliability coefficients (between .75 and .90). The goals of the studies were quite various, as the scale has been used to investigate the construct in different research fields: in school and learning processes (Amichai-Hamburger, Kaynar & Fine., 2007; Bertrams & Dickhäuser, 2009; Bors, Vigneau & Lalande, 2006), in advertising and marketing field (Kuo, Horng, Lin & Lee, 2012), in social cognition and decision-making ability (Carnevale, Inbar & Lerner, 2011; Harman, 2011), in prejudice and stereotypes studies (Cárdaba, Briñol Turnes, Horcajo & Petty, 2013). The variety of applications of the Need for Cognition Scale shows how the need for cognition is relevant to detect psychological dimensions and to study behaviors in social life. Most of the studies reporting significant associations between need for cognition and constructs such as academic achievement, dogmatism, rational decisions, intellectual engagement, openness to experience, emotional stability, and goal orientation also highlights the convergent validity of the scale. Furthermore, a study has investigated the role of gender and geographic birthplace, showing that the scale is gender-neutral and shows similar monofactorial structures in both Europe and North America (Cacioppo, Petty, Feinstein & Jarvis, 1996).

Despite numerous data supporting the validity and reliability of the items (Cacioppo et al., 1996), there has been some controversy over the underlying factorial structure. In fact, while Petty and Cacioppo have always stated that there was only one dominant factor, the analyses of other authors seem to indicate the presence of further factors. Overall, the one-dimensional model has been supported by authors (Hevey et al., 2012; Sadowski, 1993), who have argued that the one-dimensional model is more parsimonious than other models with two or three factors. This evidence led to the introduction of a monofactorial version with 6 items (Lins de Holanda Coelho, Hanel & Wolf, 2020). Similar results were

reported for the translated versions of the scale. Culhane, Morera and Hosch (2004) found a single-factor solution for the Hispanic version of the scale and the Dutch scale translation also shows a one-dimensional structure with 15 final items (Verplanken et al., 1992). However, the fact that all items of the scale load positively on a first unrotated factor does not preclude the emergence of two or more interpretable factors, after rotation (Fabrigar, Wegener, MacCallum & Strahan, 1999). This solution has become widely accepted in recent years, starting from the study by Stark and colleagues (Stark, Bentley, Lowther & Shaw, 1991), which proposed a bi-factorial structure with an identification of two dimensions: one referred to items reflecting an approach to cognitive effortful activities, and the other including items reflecting the opposite attitude, i.e. an avoidance of the cognitive activities. Forsterlee and Ho (1999) performed a PCA with oblique rotation on the 18-item NFC and reported a two-factor solution with highly correlated dimensions (r = .52): the approach and the avoidance dimensions. More recently, literature (Bors et al., 2006; Zhang, Noor & Savalei, 2016) also reported a two-factor model differentiating the approach and the avoidance factors. Furthermore, Bors and colleagues (2006), in their validation of the French version, revealed that only the avoidance dimension was predictive of academic success, supporting the idea that the approach and the avoidance are separate constructs of the Need for Cognition. About the Italian context, the recent work by Aquino and colleagues (Aquino, Picconi & Alparone, 2018), that validated the *Need for Cognition Scale* for young adults and adults (19-36 years-old), proved that the scale had two correlated factors measuring two different kinds of motivations: approach and avoidance of effortful cognitive activities. The distinction between the approach and the avoidance dimensions has already been widely accepted in psychological research, so the literature supports the idea that these two kinds of motivations should be addressed separately (for a review, Maio, Verplanken & Haddock, 2018).

As for the administration of the scale to adolescents, the short version of *Need for Cognition Scale* was validated in a few samples of secondary school students. Although previous studies, which explored the underlying Need for Cognition structure with secondary school students (Bors et al., 2006; Preckel, 2014), suggesting that the underlying factor structure of the Need for Cognition could better be explained by a unidimensional model, it confirmed the need to explore a trait-method model with two factors for positively and for negatively worded items. More recently, Georgiou and Kyza (2018) administered the scale to 177 Greek-Cypriot high school students (mean age = 15.35 years). They tested and compared 9 models and found the best fits in the unidimensional model with correlated errors among the positively worded items (final items were 14, χ^2 = 163.30, df = 56, CFI = .959, RMSEA = .06, SRMR = .031). However, even the two-factor model with two correlated factors defined by the polarity of items showed a good fit indices (χ^2 = 132.63, df = 56, CFI = .964, RMSEA = .051, SRMR = .031). In sum, the factor structure of the *Need for Cognition Scale* is still debated. The literature reported the factor structures relate to i) a unidimensional Need for Cognition factor model, which in most cases sacrifices some items; ii) a trait-method model, which consider the effect of the positively and negatively worded items; and iii) two factors model which defines two kinds of motivation through the polarity of items.

Thus, despite these various approaches to measuring this construct, the need for cognition has rarely been studied in adolescence despite its relationship to learning and school performance (Luong et al., 2017), as well as its relevance to the motivational aspects of learning (Wigfield & Eccles, 2000) Therefore, the urge for knowledge that drives the need for cognition seems very important when considering the age of adolescence, which is a fundamental phase of life for the developmental acquisition of thinking that underlies the appropriate decision-making and judgment formation (De Haan, 2010; Lombardi, Di Dio, Castelli, Massaro & Marchetti, 2017; Taimur & Sattar, 2019). Moreover, this would appear significant when considering that adolescents, even as they face adulthood, still often employ imprecise and superficial logical processes when analyzing a problem or understanding complex phenomena, including socially relevant ones (Berti et al., 2017). Thus, starting from the work by Aquino and colleagues (2018) on an Italian sample of young adults and adults and considering the few studies on adolescents, the present study intends to provide the Italian validation of the questionnaire for adolescence, which will also contribute to the literature about the underlying factor structure of the *Need for Cognition Scale*.

The goal is to explore the psychometric characteristics of the short version of *Need for Cognition Scale* for Italian secondary school students (14-18 years). For this purpose, we intend to:

- verify the factorial structure of the NFCS for Italian adolescents by confirmatory factor analysis;
- investigate measurement invariance regarding gender and

the administered measurement (paper-pencil *vs* online) by employing multigroup confirmatory factor analysis.

METHOD

Participants and procedure

In order to assess the factor structure of the scale, 473 Italian students (261 males) participated in the study, from schools in the Northern of Italy. The age is between 14 and 18 years (mean age = 15.6 years, *SD* = .9). Of the participants, 74% have a family of Italian origin, the remaining is distributed between European and extra-European countries. Additionally, a sub-sample of 318 participants also completed the scales necessary to assess the convergent and divergent validity of the *Need for Cognition Scale*. After receiving the school-manager's approval to carry out the research, the caregivers and the students were informed on the aim and procedure of the study. Parents provided a written consent for their children's participation in the study and students gave informed written consent to the study, according to the General Data Protection Regulation (GDPR 2016/79, 25/05/2018). The present study was approved by the Scientific and Ethics Committee of the Department of Psychology of Catholic University of Milan, in accordance with the Helsinki Declaration. The Italian version for adolescence of the NFCS was assessed both via an online procedure and a pen-pencil procedure. The questionnaires were presented to the students through face-to-face meetings in the classroom (*N* = 390, 82.5% of the total, 55% are males, mean age = 15.3), or online (*N* = 83, 17.5% of the total, 60% are males, mean age = 15.7). On such occasions there have been given careful instructions on mode for compilation, delivering the questionnaires in a unique moment. All participants completed the Italian translation version of 18-items of the *Need for Cognition Scale* (Cacioppo et al., 1984), provided by the Department of Education and Psychology - University of Florence and is similar to the version proposed for young adults by Aquino et al. (2018).

Measures

The following three questionnaires were administered:

– The *Need for Cognition Scale – Short Version* (*NFCS*; Cacioppo et al., 1984), originally composed of 18 items, which investigates the need for cognition. Bilinguals performed back-translations, blind to the content of the original English words. The original English items and the Italian adaptation of the NFCS are reported in the Appendix. Based on the empirical data in the literature and the theoretical reference model, the items are distributed on two factors, which measures the approach (e.g., "I find satisfaction in deliberating hard and for long hours", α = .82) and the avoidance (e.g., "I only think as hard as I have to", α = .72) dimensions. The proposed questionnaire consists of 18 statements for each of which must be answered on a 5-point Likert scale (from 1 = does not describe me at all, to 5 = It describes me very well). Higher scores reflect a higher adhesion with the statement. Items 3, 4, 5, 7, 8, 9, 12, 16,17 were kept as items reversed as in the original version. About the avoidance dimension, a higher score indicated a minor tendency to avoid cognitive situations. The items of NFCS, the descriptive statistics and reliability of the measure are shown in Table 1.

– The *Big Five Inventory* (*BFI*), which was proposed to verify the convergent validity of the NFCS. The Italian version (Ubbiali, Chiorri, Hampton & Donati, 2013), is a questionnaire for adults and young adults that investigates personality according to the 5-factor theory. The personality factors are: Extraversion (8 items), Agreeableness (9 items), Conscientiousness (9 items), Emotional stability (8 items), Open-mindedness (10 items). The proposed questionnaire consists of 44 statements on a 5-step Likert- type response scale (from 1 = strongly disagree to 5 = strongly agree).

– The *Need for Cognitive Closure Scale* was proposed to verify the divergent validity of the NFCS (Webster & Kruglanski, 1994). The Italian version measures the "desire on the part of the individual for a definitive and certain answer to a question/problem and reluctance to ambiguity" (Pierro et al., 1995). The proposed questionnaire consists of 42 items divided into five scales: 1) preference for order and structure in environment; 2) predictability of future contexts; 3) decisiveness of judgments and choices; 4) affective discomfort occasioned by ambiguity; 5) closed-mindedness. Each item must be answered on a 6-step Likert-type response scale (from 1 = strongly disagree to 6 = strongly agree). The final score was computed as the mean of the items.

Table 1 – Need for Cognition Scale (NFCS): Descriptive statistics and reliability (18 items)

				If item dropped
	Mean	***SD***	**Item-rest correlation**	**Cronbach's α**
NFCS1 - I prefer complex to simple problems	2.79	1.221	.473	.805
NFCS2 - I like to have the responsibility of handling a situation that requires a lot of thinking	3.01	1.160	.618	.796
NFCS3r - Thinking is not my idea of fun	3.24	1.084	.385	.810
NFCS4r - I would rather do something that requires little thought than something that is sure to challenge my thinking abilities	3.07	1.131	.494	.804
NFCS5r - I try to anticipate and avoid situations where there is a likely chance, I will have to think in depth about something	3.23	1.066	.367	.811
NFCS6 - I find satisfaction in deliberating hard and for long hours	2.93	1.076	.554	.801
NFCS7r - I only think as hard as I have to	3.34	1.133	.467	.806
NFCS8r - I prefer to think about small, daily projects to long-term ones	3.13	1.129	.360	.812
NFCS9r - I like tasks that require little thought once I've learned them	2.73	1.100	.308	.815
NFCS10 - The idea of relying on thought to make my way to the top appeals to me	3.76	.932	.432	.808
NFCS11- I really enjoy a task that involves coming up with new solutions to problems	2.88	1.202	.499	.803
NFCS12r - Learning new ways to think doesn't excite me very much	3.25	1.074	.234	.818
NFCS13 - I prefer my life to be filled with puzzles that I must solve	2.44	1.117	.414	.809
NFCS14 - The notion of thinking abstractly is appealing to me	3.10	1.197	.426	.808
NFCS15 - I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought	2.60	1.102	.423	.808
NFCS16r - I feel relief rather than satisfaction after completing a task that required a lot of mental effort	2.80	1.090	.154	.823
NFC17r - It's enough for me that something gets the job done; I don't care how or why it works	3.16	1.192	.299	.816
NFCS18 - I usually end up deliberating about issues even when they do not affect me personally	3.48	1.148	.349	.812

Data analysis

The preliminary analyses were performed with the support of IBS SPSS Statistics for Apple Macintosh OSX, Version 22.0, in order to check the normal distribution by calculating mean, standard deviation, and indices of skewness and kurtosis and verified the possible presence of outliers. The analysis of skewness and kurtosis indicated that the distance from normality was not severe (the indices were between −1.04 and 1.60), so no variable transformations were deemed necessary.

Confirmatory Factor Analysis (CFA) was conducted using SEM with Mplus 7.11 software (Muthén & Muthén, 1998-2015). To evaluate the CFA models and to test invariances, goodness of fit was estimated by Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI), Standardized Root Mean Square Residual (SRMR). A Robust Maximum Likelihood (MLR) method of estimation was applied to test the hypothesized model. The χ^2 should be non-significant in order to consider the CFA model as fitting the observed data; however, since it is largely affected by sample size (Hu & Bentler, 1999), we examined other fit indices (Hu & Bentler, 1998). Models with good fit present a RMSEA<.08 and CFI>.90 (Bentler, 1990), whereas models with excellent fit present a RMSEA<.05 and CFI>.95 (Hu & Bentler, 1998). Akaike Information Criterion (AIC) was used to compare the relative fit of models, with lower AIC values indicating superior model fit (Marsh, Hau & Wen, 2004).

Multigroup analyses (MGCFA) were carried out to verify the invariance of this factorial structure across gender and type of administration starting from the model supported by the CFA (Brown, 2015). Competing models were compared with regard to their model fit by performing i) χ^2 difference tests ($\Delta\chi^2$ and Satorra-Bentler scaled χ^2; Satorra & Bentler, 2001) and ii) the difference between comparative fit indices (ΔCFI). More specifically, we compared the scale structure (configural invariance), the items' factor loading (metric invariance), the items' intercepts (scalar invariance) as well as their residual variance (uniqueness invariance). If the chi-square obtained as difference among the two models' χ^2 is not significant, the two models are not significantly different and, thus, invariance is met. Because this test is "sensitive" to sample size (Cheung & Rensvold, 2002), we also tested the ΔCFI model comparison. A negative ΔCFI value lower than −.01 (e.g., ΔCFI = −.02) would indicate a lack of invariance (Cheung & Rensvold, 2002), while when CFI increases (i.e., ΔCFI>0) that may occur due to the changes in degree of freedom, there are no concerns about the invariance, because a larger CFI indicates better fit (Dimitrov, 2010).

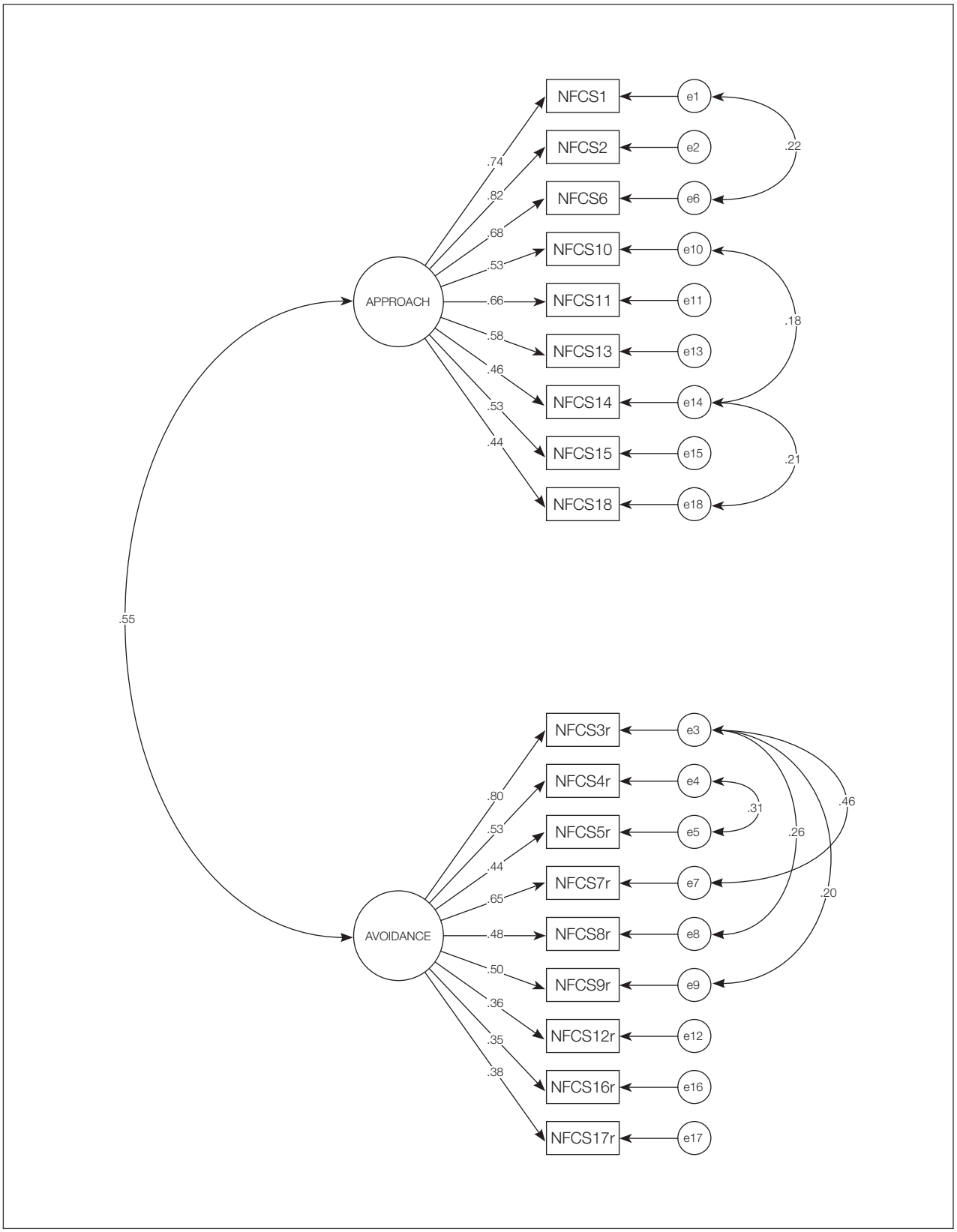
The reliability was evaluated using Cronbach's alpha (Cronbach, 1951) and the convergent and divergent validity were assessed using the *Big Five Inventory* and the *Need for Closure Scale*. In particular, statistically significant positive relationships with the score on the Open-mindedness scale of the BFI and statistically significant negative relationships with the Need for Closure score were expected. Furthermore, descriptive statistics of the different dimensions and correlations for all variables of interest were calculated.

RESULTS

According to the literature suggesting a single dimension of the scale (e.g., Cacioppo et al., 1984; Hevey et al., 2012), a CFA was carried out, performed through the robust maximum likelihood method. Modification indices were also checked to assess the degree to which the hypothesized model was adequately described. Correlated errors are specified when items share a portion of variance. CFA revealed that the uni-dimensional model had acceptable fit indices but worse compared to the bidimensional model. The uni-dimensional model fit indices are acceptable: $\chi^2_{(136)} = 567.55$; RMSEA = .082, 90% CI [.07;.09], CFI = .74; SRMR = .088; AIC = 24755.27. On the contrary, the bidimensional model showed a good fit: $\chi^2_{(135)} = 331.62$; RMSEA = .056, 90% CI [.05;.06], CFI = .85; SRMR = .105; AIC = 24576.29 which is in line with the results shown by Aquino and colleagues (2018) (RMSEA = .058: 90% CI [.04;.07], CFI = .91; NNFI = .90; SRMR = .06). The modification indices analysis suggested adding covariance between the errors (freeing up errors covariances was allowed because they are part of the same latent variable) in order to improve fit indices, showing an excellent fit: $\chi^2_{(127)} = 199.37$; RMSEA = .035, 90% CI [.02;.04], CFI = .95; SRMR = .054; AIC = 24431.10. The AIC index confirmed that this solution better fitted the data. As shown in Figure 1 that depicts the bidimensional solution, all factor loadings were statistically significant at $p<.001$ and ranged from .35 to .82. CFA upheld that the factors were related to each other, $r = .55$. Correlated errors are performed as suggested by the modification indices.

Measurement invariance refers to the ability of an instrument to measure the same construct across different

Figure 1 – Confirmative Factorial Analysis of the Italian Need for Cognition Scale – Short Version for adolescence with standardized regression weights



groups and offers confidence in the application of the construct across groups and also allows for verification that scores that have been derived from a measure of the construct can be directly compared (Brown, 2015). Structural equation modelling offers a robust and established methodological framework within which to assess measurement invariance. Multigroup invariance analyses (MGCFA) were performed in order to verify the NFCS equivalence across gender (male *vs* female) and type of administration (paper-pencil vs online). We firstly verified the validity of the baseline model both for gender and type of administration, as suggested literature (Brown, 2015). About type of administration, baseline models showed acceptable fit indexes both for online (RMSEA = .027, 90% CI [.00;.06], CFI = .96; SRMR = .094) and paper pencil administrations (RMSEA = .037, 90% CI [.026;.047], CFI = .94; SRMR = .081). About gender, baseline models showed a good fit indexes both for male (RMSEA = .060, 90% CI [.051;.081], CFI = .81; SRMR = .09) and female (RMSEA = .061, 90% CI [.052;.077], CFI = .82; SRMR = .07). Subsequently, using multi-group analysis, groups were compared according to four types of measurement invariance (configural, weak, strong, strict invariance). These steps of invariance testing have to be taken from weakest (configural) to strongest (strict invariance). Given the sample size (N>300), a decrease in CFI equal to or greater than .01, along with an increase in RMSEA equal to or greater than .015, is considered a substantial decrease in model fit (Chen, 2007). As shown in Table 2 and in Table 3, results from the analysis of the multi-group measurement invariance supported the presence of strict invariance across gender and type of administration, showing the evidence of generalizability of the NFCS across gender and modality of administration (paper-pencil vs online).

About the convergent and divergent validity, Table 4 shows the correlations of the approach and avoidance dimension of NFCS with other measures, namely *Big Five Inventory and Need for Closure Scale*.

The pattern of correlations between *Big Five Inventory* and *Need for Cognition Scale* shown in Table 4 is consistent with the literature. In particular, significant positive relationships have been demonstrated between the scores obtained at the NFCS and those obtained on the Open-mindedness and Conscientiousness scales. Table 5 shows the correlations with the *Need for Cognitive Closure Scale*, in particular the negative correlation of the Closed-mindedness scale with the 2 dimensions of the *Need for Cognition Scale* confirms the different dimensions investigated by these two questionnaires. In particular, with respect the correlations about the open mindedness components of *Big Five Inventory* and Closed-mindedness as measured by the *Need for Cognitive Closure Scale*, show the validity of the *Need for Cognition Scal*e dimensions.

DISCUSSION

In this study, we have developed the Italian version for adolescents of the *Need for Cognition Scale – Short Version* that can be used with students between 14 and 18 years. In particular, the aims of the present study were: i) to verify the factorial structure of the *Need for Cognition Scale – Short Version* for Italian adolescents, ii) to investigate measurement invariance regarding gender and the administered measurement.

About our first goal, the confirmatory factor analysis suggested an excellent fit of bifactorial solution with a differentiation between the approach to cognitive effortful activities and the avoidance of situations requiring a lot of thinking, as the previous studies show, both in the Italian context (Aquino et al., 2018) and with the secondary school students (Bors et al., 2006). Both the approach (Cronbach's alpha = .82) and the avoidance dimensions (Cronbach's alpha = .72) showed good internal consistency, preserving all items of the original short form. These findings contribute to the body of literature, which is still debated, regarding the structure of the scale, bearing important implications on the topic of data collection through rating scales with both negative and positive items, underlying the two factors model which defines two kinds of motivation through the polarity of items. Furthermore, related to our second objective, we show that the *Need for Cognition Scale – Short version* for Italian adolescents is invariant for gender and for the administration measurement, indicating that the items assess the same attitude factors for both female and male adolescents. Therefore, any differences between males and females can be attributed to actual variations in the responses to some items, and not to a differential functioning of the questionnaire, as claimed by other studies (Cacioppo et al., 1996; Lins de Holanda Coelho et al., 2020). Moreover, the results from online administration (used because of the pandemic Covid19) are similar to the results from paper-pencil administration, indicating that the items, even though

Table 2 – Measurement invariance for NFCS between male and female

				RMSEA							
Measurement Invariance - gender	***df***	χ^2	**CFI**	**Estimate**	**95% CI**	**SRMR**	**Δ*df***	**SBΔχ^2**	***p***	**ΔRMSEA**	**ΔCFI**
Configural	250	319.114	.952	.034	.022-.045	.054					
Weak (*vs* configural)	266	341.836	.947	.035	.023-.045	.064	16	22.85	.118	.001	.005
Strong (*vs* weak)	282	365.276	.936	.037	.026-.047	.068	16	23.81	.094	.002	.011
Strict (*vs* strong)	300	390.686	.919	.045	.036-.054	.072	18	25.79	.105	.008	.017

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

Table 3 – Measurement invariance for NFCS between paper-pencil and online administration

				RMSEA							
Measurement Invariance - administration	***df***	χ^2	**CFI**	**Estimate**	**95% CI**	**SRMR**	**Δ*df***	**SBΔχ^2**	***p***	**ΔRMSEA**	**ΔCFI**
Configural	250	295.696	.968	.026	.011-.040	.051					
Weak (*vs* configural)	266	311.058	.967	.027	.010-.039	.058	16	15.14	.514	.001	.001
Strong (*vs* weak)	282	324.933	.965	.028	.009-.039	.059	16	14.21	.583	.001	.002
Strict (*vs* strong)	300	337.193	.964	.029	.008-.041	.062	18	10.26	.923	.001	.001

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

Table 4 – Correlation coefficients between the NFCS and the factors of Big Five Inventory

	Conscientiousness	Open-mindedness	Emotional stability	Extraversion	Agreeableness
NFCS Approach	.195**	.480***	.063	.061	.100
NFCS Avoidance	.189**	−.316***	.137*	−.009	.099

Note. $N = 318$.
* $p<.05$, ** $p<.01$, *** $p<.001$

Table 5 – Correlation coefficients between the NFCS and the factors of Need for Cognitive Closure Scale

	Decisiveness	Predictability	Affective discomfort	Preference for order	Closed-mindedness
NFCS Approach	.097	−.121 *	.189***	.078	−.320***
NFCS Avoidance	.151**	−.076	.074	−.010	.345 ***

Note. $N = 318$.
* $p<.05$, ** $p<.01$, *** $p<.001$

administered in different modalities, assess attitude factors in a similar way. This finding has implications especially in light of the latest trends regarding online ways of administration.

Furthermore, we investigated the validity of the scale using scale statistics for internal consistency and correlations with conceptually related measures. The scale shows good psychometric properties with respect both to the reliability and to the measure of construct validity. In particular, the strong significant correlations with the scale of the Open-mindedness of the *Big Five Inventory* and the Closed-mindedness of the Need for Closure questionnaire prove that *Need for Cognition Scale – Short Version* can be considered a measure of the inclination to demanding cognitive tasks also in Italian adolescents. This result underlines the tendency of people with a high need for cognition that enjoys themselves in dealing with complex problems, be curious, and engage in new ideas, showing to evaluate and process incoming information as do those that are highly open-minded (e.g. Mussel, 2010), as corroborates both positive and negative significant correlations with Open-mindedness factor (respectively with Approach and Avoidance dimension of *Need for Cognition Scale*). Furthermore, the correlations show that the need for cognition has correlated aspects with conscientiousness, which includes aspects, such as carefulness, organization, control, and dutifulness. This result underlines the overlapping and positively correlated elements between the two constructs, as shown by previous literature (e.g. Fleischhauer et al., 2010; Furnham & Thorne, 2013). Together, these findings confirmed the convergent and divergent validity of *Need for Cognition Scale*. Furthermore, none of the correlation coefficients was equal to or greater than .70, underlines there isn't overlap with other constructs.

CONCLUSIONS, LIMITATIONS AND FUTURE DIRECTIONS.

To sum up, the present study resulted in 18-item version of the *Need for Cognition Scale – Short Versio*n for Italian adolescents, as a validated measurement of cognitive motivation, which can be used with Italian-speaking students attending secondary school. The scale has characteristics of simplicity and speed in the compilation that make it particularly suitable for research in typical and atypical populations. Results showed that bi-dimensional model shows a good fit, which is improved to excellent after correlating the residuals. This is relevant in view of the fact that the factorial structure of the Need for Cognition is controversial. In order to better specify the structure of the model, the applied modifications should be confirmed in future studies. Regarding the limitations of this study, according to the literature (Bors et al., 2006), socio-economic status and school outcomes should be considered in future research. Moreover, other questionnaires investigating the convergent validity of the NFC should be explored for this age group; in fact, the low to moderate correlations we found suggest that the questionnaires used are probably weaker at capturing these dimensions in adolescence. Additionally, considering a sample only from Northern Italy could be considered a limitation of the study. Beyond these limits, the Need for Cognition questionnaire can provide useful data in studies concerning the interaction between personality characteristics and learning contexts in adolescence period of life. It can also be applied in longitudinal studies, aimed at investigating the relationship between the different aspects of personality with the choice of academic pathways and the academic performance in a prospective way.

References

AMICHAI-HAMBURGER, Y., KAYNAR, O. & FINE, A. (2007). The effects of need for cognition on Internet use. *Computers in Human Behavior*, *23* (1), 880-891. doi: 10.1016/j.chb.2006.03.002

AQUINO, A., PICCONI, L. & ALPARONE, F.R. (2018). Validation of the Italian version of the Need for Cognition Scale – Short Version. *BPA-Applied Psychology Bulletin*, *66* (283).

BENTLER, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107* (2), 238-246. doi:10.1037/0033-2909.107.2.238

BERTI, A.E., AJELLO, A.M., APREA, C., CASTELLI, I., LOMBARDI, E., MARCHETTI, A., MASSARO, D., SAPPA, V. & VALLE A. (2017). Adolescents' and young adults' naïve understandings of the economic crisis. *Europe's Journal of Psychology*. doi: 10.5964/ejop.v13i1.1187

BERTRAMS, A. & DICKHÄUSER, O. (2009). High-school students' need for cognition, self-control capacity, and school achievement: Testing a mediation hypothesis. *Learning and Individual Differences*, *19* (1), 135-138. doi: 10.1016/j.lindif.2008.06.005

BORS, D.A., VIGNEAU, F. & LALANDE, F. (2006). Measuring the need for cognition: Item polarity, dimensionality, and the relation with ability. *Personality and Individual Differences*, *40* (4), 819-828. doi: 10.1016/j.paid.2005.09.007

BROWN, T.A. (2015). *Confirmatory Factor Analysis for applied research (2nd ed.)*. London: The Guilford Press.

CACIOPPO, J.T. & PETTY, R.E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42* (1), 116. doi: 10.1037/0022-3514.42.1.116

CACIOPPO, J.T., PETTY, R.E., FEINSTEIN, J.A. & JARVIS, W.B.G.

(1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, *119* (2), 197-253. doi: 10.1037/0033-2909.119.2.197

CACIOPPO, J.T., PETTY, R.E. & FENG KAO, C. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48* (3), 306-307. doi: 10.1207/s15327752jpa4803_13

CÁRDABA, M.A.M., BRIÑOL, P., HORCAJO, J. & PETTY, R.E. (2013). The effect of need for cognition on the stability of prejudiced attitudes toward South American immigrants. *Psicothema*, *25*, 73-78. doi: 10.7334/psicothema2012.107. PMID: 23336547

CARNEVALE, J.J., INBAR, Y. & LERNER, J.S. (2011). Individual differences in need for cognition and decision-making competence among leaders. *Personality and Individual Differences*, *51* (3), 274-278. doi: 10.1016/j.paid.2010.07.002

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 10.1007/BF02310555

CHEUNG, G.W. & RENSVOLD, R.B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9* (2), 233-255. doi: 10.1207/S15328007SEM0902_5

COHEN, A.R., STOTLAND, E. & WOLFE, D.M. (1955). An experimental investigation of need for cognition. *The Journal of Abnormal and Social Psychology*, *51* (2), 291-294. doi: 10.1037/h0042761

CRONBACH, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297-334

CULHANE, S.E., MORERA, O.F. & HOSCH, H.M. (2004). The factor structure of the need for cognition short form in a Hispanic sample. *The Journal of Psychology*, *138* (1), 77-90. doi: 10.3200/JRLP.138.1.77-90

DAI, D.Y. & WANG, X. (2007). The role of need for cognition and reader beliefs in text comprehension and interest development. *Contemporary Educational Psychology*, *32* (3), 332-347. doi: 10.1016/j.cedpsych.2006.05.002

DE HAAN, G. (2010). The development of ESD-related competencies in supportive institutional frameworks. *International Review of Education*, *56* (2), 315-328. doi: 10.1007/s11159-010-9157-9

DIMITROV, D.M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*, *43* (2), 121-149. doi:10.1177/0748175610373459

FABRIGAR, L.R., WEGENER, D.T., MacCALLUM, R.C. & STRAHAN, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4* (3), 272-299. doi: 10.1037/1082-989X.4.3.272

FLEISCHHAUER, M., ENGE, S., BROCKE, B., ULLRICH, J., STROBEL, A. & STROBEL, A. (2010). Same or different? Clarifying the relationship of need for cognition to personality and intelligence. *Personality and Social Psychology Bulletin*, *36* (1), 82-96. doi:10.1177/0146167209351886

FORSTERLEE, R. & HO, R. (1999). An examination of the short form of the Need for Cognition Scale applied in an Australian sample. *Educational and Psychological Measurement*, *59* (3), 471-480. doi: 10.1177/00131649921969983

FURNHAM, A. & THORNE, J.D. (2013). Need for cognition: Its dimensionality and personality and intelligence correlates. *Journal of Individual Differences*, *34*, 230-240. doi: 10.1027/1614-0001/a000119

GEORGIOU, Y. & KYZA, E.A. (2018). Translation, adaptation, and validation of the Need for Cognition Scale – Short Form in the Greek language for secondary school students. *Journal of Psychoeducational Assessment, 36* (5), 523-531. doi: 10.1177/0734282916686005

HARMAN, J.L. (2011). Individual differences in need for cognition and decision making in the Iowa Gambling Task. *Personality and Individual Differences*, *51* (2), 112-116. doi: 10.1016/j.paid.2011.03.021

HEVEY, D., THOMAS, K., PERTL, M., MAHER, L., CRAIG, A. & NI CHUINNEAGAIN, S. (2012). Method effects and the Need for Cognition Scale. *International Journal of Educational and Psychological Assessment, 12* (1), 20-33.

HU, L.T. & BENTLER, P.M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3* (4), 424-453. doi:10.1037/1082-989X.3.4.424

HU, L.T. & BENTLER, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55. doi:10.1080/10705519909540118

KUO, J.C., HORNG, D.J., LIN, C.L. & LEE, S.H. (2012). The causal relationship between need for cognition and advertising recall. *Social Behavior and Personality: An International Journal*, *40* (6), 1025-1032. doi: 10.2224/sbp.2012.40.6.1025

LEVIN, I.P., HUNEKE, M.E. & JASPER, J.D. (2000). Information processing at successive stages of decision making: Need for cognition and inclusion-exclusion effects. *Organizational Behavior and Human Decision Processes, 82* (2), 171-193. doi: 10.1006/obhd.2000.2881

LINS DE HOLANDA COELHO, G., HANEL, P.H.P. & WOLF, L.J. (2020). The very efficient assessment of need for cognition: Developing a six-item version. *Assessment*, *27* (8), 1870-1885. doi:10.1177/1073191118793208

LOMBARDI, E., DI DIO, C., CASTELLI, I., MASSARO, D. & MARCHETTI, A. (2017). Prospective thinking and decision making in primary school age children. *Heliyon*, *3* (6), e00323. doi:10.1016/j.heliyon.2017.e00323

LUONG, C., STROBEL, A., WOLLSCHLÄGER, R., GREIFF, S., VAINIKAINEN, M.P. & PRECKEL, F. (2017). Need for cognition in children and adolescents: Behavioral correlates and relations to academic achievement and potential. *Learning and Individual Differences*, *53*, 103-113. doi:10.1016/j.lindif.2016.10.019

MAIO, G.R. & HADDOCK, G. (2010). *The psychology of attitudes and attitude change*. Thousand Oaks, CA: Sage.

MAIO, G.R., VERPLANKEN, B. & HADDOCK, G. (2018). *The psychology of attitudes and attitude change*. Sage.

MARSH, H.W., HAU, K.T. & WEN, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, *11* (3), 320-341. doi: 10.1207/s15328007sem1103_2

MARTÍN CÁRDABA, M.Á., BRIÑOL TURNES, P., HORCAJO, J. & PETTY, R.E. (2013). The effect of need for cognition on the stability of prejudiced attitudes toward South American immigrants. *Psicothema*, *25* (1), 73-78. doi: 10.7334/psicothema2012.107

MUSSEL, P. (2010). Epistemic curiosity and related constructs: Lacking evidence of discriminant validity-science direct. *Personality and Individual Differences*, *49*, 506-510. doi: 10.1016/j.paid.2010.05.014

MUTHÉN L.K. & MUTHÉN B.O. (1998-2015). *Mplus User's Guide (Version 7)*. Los Angeles, CA: Muthén & Muthén. doi: 10.1007/s12519-009-0049-7

PETTY, R.E., BRIÑOL, P., LOERSCH, C. & McCASLIN, M.J. (2009). The need for cognition. In M.R. Leary & R.H. Hoyle (Eds.), *Handbook of individual differences in social behavior*. Guilford Press.

PETTY, R.E., BRIÑOL, P. & TORMALA, Z.L. (2002). Thought confidence as a determinant of persuasion: The self-validation hypothesis. Journal of *Personality and Social Psychology, 82* (5), 722-741. doi: 10.1037/0022-3514.82.5.722

PETTY, R.E. & CACIOPPO, J.T. (Eds.) (1986). The elaboration likelihood model of persuasion. In R.E. Petty & J.T. Cacioppo, *Communication and persuasion*, New York, NY: Springer.

PIERRO, A., MANNETTI, L., CONVERSO, D., GARSIA, V., MIGLIETTA, A., RAVENNA, M. & RUBINI, M. (1995). Caratteristiche strutturali della versione italiana della scala di bisogno di chiusura cognitiva. *TPM*, *2* (3-4), 125-141.

PRECKEL, F. (2014). Assessing need for cognition in early adolescence: Validation of a German adaption of the Cacioppo/Petty Scale. *European Journal of Psychological Assessment, 30* (1), 65-72. doi: 10.1027/1015-5759/a000170

SADOWSKI, C.J. (1993). An examination of the short Need for Cognition Scale. *The Journal of Psychology: Interdisciplinary and Applied*, *127* (4), 451-454. doi: 10.1080/00223980.1993.9915581

SATORRA, A. & BENTLER, P.M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*, 507-514. doi:10.1007/BF02296192

SMITH, S.M. & LEVIN, I.P. (1996). Need for cognition and choice framing effects. *Journal of Behavioral Decision Making*, *9*, 283-290.

STARK, J.S., BENTLEY, R.J., LOWTHER, M.A. & SHAW, K.M. (1991). The student goals exploration: Reliability and concurrent validity. *Educational and Psychological Measurement*, *51*, 413-422. doi:10.1177/0013164491512015

TABACHNICK, B.G. & FIDELL, S. (2007). *Using multivariate statistics*. Boston, MA: Pearson/Allyn & Bacon.

TAIMUR, S. & SATTAR, H. (2019). Education for sustainable development and critical thinking competency. In W. Leal Filho, A. Azul, L. Brandli, P. Özuyar & T. Wall (Eds), *Quality education. Encyclopedia of the UN sustainable development goals.* Springer, Cham. Doi:10.1007/978-3-319-69902-8_64-1

UBBIALI, A., CHIORRI, C., HAMPTON, P. & DONATI, D. (2013). Psychometric properties of the Italian adaptation of the Big Five Inventory (BFI). *Bollettino di Psicologia Applicata*, *266*, 37-48.

VERPLANKEN, B., HAZENBERG, P.T. & PALENEWEN, G.R. (1992). Need for cognition and external information search effort. *Journal of Research in Personality*, 26 (2), 128-136. doi: 10.1016/0092-6566(92)90049-A

WEBSTER, D.M., & KRUGLANSKI, A.W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology, 67* (6), 1049-1062. doi:10.1037/0022-3514.67.6.1049

WIGFIELD, A. & ECCLES, J.S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology, 25* (1), 68-81. doi: 10.1006/ceps.1999.1015

ZHANG, X., NOOR, R. & SAVALEI, V. (2016). Examining the effect of reverse worded items on the factor structure of the Need for Cognition Scale. *PLoS ONE*, *11* (6): e0157795. doi:10.1371/journal.pone.0157795

APPENDIX

Translated (and original items) of the Italian version of Need for Cognition Scale – Short Version for Adolescence

NFCS1 - Preferisco i problemi complicati ai semplici (I prefer complex to simple problems)
NFCS2 - Mi piace dover gestire una situazione che richiede molto da pensare (I like to have the responsibility of handling a situation that requires a lot of thinking)
NFCS6 - Io provo soddisfazione quando c'è da riflettere molto e a lungo (I find satisfaction in deliberating hard and for long hours)
NFCS10 - Mi attrae l'idea di fare affidamento sulla mia capacità di pensare per arrivare in alto (The idea of relying on thought to make my way to the top appeals to me)
NFCS11 - Mi piacciono i compiti in cui si devono trovare nuove soluzioni ai problemi (I really enjoy a task that involves coming up with new solutions to problems)
NFCS13 - Preferisco che la mia vita sia piena di dilemmi da risolvere (I prefer my life to be filled with puzzles that I must solve)
NFCS14 - Mi attrae l'idea di pensare in modo astratto (The notion of thinking abstractly is appealing to me)
NFCS15 - Preferisco compiti intellettuali, difficili ed importanti a quelli abbastanza importanti ma che non richiedono tanto da pensare (I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought)
NFCS18 - Di solito finisco col riflettere su problemi anche quando non mi riguardano personalmente (I usually end up deliberating about issues even when they do not affect me personally)
NFCS3r - Pensare non è l'idea che ho di divertimento (Thinking is not my idea of fun)
NFCS4r - Preferirei fare qualcosa che non richiede molto da pensare, piuttosto che qualcosa che sfidi le mie abilità di pensiero (I would rather do something that requires little thought than something that is sure to challenge my thinking abilities)
NFCS5r - Tendo a prevedere ed evitare situazioni in cui sia molto probabile dover pensare profondamente su qualcosa (I try to anticipate and avoid situations where there is a likely chance, I will have to think in depth about something)
NFCS7r - Io mi impegno a ragionare solo lo stretto necessario (I only think as hard as I have to)
NFCS8r - Preferisco pensare a piccoli progetti quotidiani piuttosto che a quelli a lungo termine (I prefer to think about small, daily projects to long-term ones)
NFCS9r - Mi piacciono i compiti che, una volta appresi, richiedono poco da pensare (I like tasks that require little thought once I've learned them)
NFCS12r – L'idea di imparare nuovi modi di pensare non mi eccita molto (Learning new ways to think doesn't excite me very much)
NFCS16r - Dopo aver completato un compito che richiede un grande sforzo mentale mi sento sollevato, piuttosto che soddisfatto (I feel relief rather than satisfaction after completing a task that required a lot of mental effort)
NFCS17r - Mi basta trovare un qualsiasi modo per terminare il lavoro; non mi importa come o perché funzioni (It's enough for me that something gets the job done; I don't care how or why it works)

Legenda. r = reverse items.
Note. Original items are in brackets.