

19.00 €

vol. LXV. September-December 2017

N° 280

Four-monthly Journal
ISSN 0006-6761

BPA



Bollettino di Psicologia Applicata

APPLIED PSYCHOLOGY BULLETIN

Indexed in PsycINFO® – Scopus Bibliographic Database



Review



Experiences & Tools



Research



GIUNTIO.S.
PSYCHOMETRICS

Scientific Director Alessandro Zennaro

CONTENTS

■ Review

- Measuring empathy: A literature review of available tools 2
Guste Ilgunaite, Luciano Giromini, Marzia Di Girolamo

◆ Research

- Can video games be an innovative tool to assess personality traits of the Millennial generation? An exploratory research 29
Elena Formica, Elena Gaiffi, Mario Magnani, Andrea Mancini, Ezio Scatolini, Manuele Ulivieri

▲ Experiences & Tools

- Validation and psychometric properties of the Italian Copenhagen Psychosocial Questionnaire II - short version 48
Ilaria Setti, Angelo d'Errico, Daniela Di Cuonzo, Elena Fiabane, Piergiorgio Argentero

- The CBQ-p: A confirmatory study on factor structure and convergent validity with psychotic-like experiences and cognitions in adolescents and young adults 58
Andrea Pozza, Davide Dèttore

- Raven's Standard Progressive Matrices: Contribution to Italian standardization for subjects between ages 6 and 18 70
Laura Picone, Arturo Orsini, Lina Pezzuti

Measuring empathy: A literature review of available tools

Guste Ilgunaite, Luciano Giromini, Marzia Di Girolamo

Department of Psychology, University of Turin

• **ABSTRACT.** L'empatia è definita come la capacità di un individuo di capire come si sente l'altro, acquisirne lo stato d'animo e stargli emotivamente vicino. Lo scopo di questo studio è stato, attraverso la ricerca della letteratura e la definizione dell'empatia, quello di evidenziare l'importanza del costrutto, evidenziando le metodologie più usate negli ultimi 15 anni. È emerso che la tipologia di metodi più usata sono i questionari self-report ma che esistono altri strumenti per misurare l'empatia, i quali non sono di facile impiego a causa della scarsa esplicazione sull'uso o sulla categoria di strumento a cui sono riferiti.

• **SUMMARY.** Empathy is described as a complex construct that develops the whole life of a person. It is defined as the capability of a person to understand the other person's feelings, to be able to feel the same way the other person does and to be capable to take action to resolve the problem of that individual. There are three major categories of empathy, i.e., affective, cognitive and compassionate. All are very important to clinical psychology, interpersonal relationships, and psychological assessment. The aim of this study was to conduct a literature review to describe the most widely used instruments to measure empathy in the range of the last fifteen years. The results showed that there are different approaches to measuring empathy, with the most popular ones being self-reports.

Keywords: Empathy, Measure, Self-report, Review

INTRODUCTION

Empathy is a complex construct that is defined as a process that changes more or less the whole life of the individual (Zillmann, 1991). It starts to develop from the minute that the person is born till the end of his life (Zahn-Waxler, Robinson & Emde, 1992). However, there is no proof that empathy increases with age, even if some developmental changes of it are typically seen in childhood (O'Brien, Konrath, Gruhn & Hagen, 2012) – this change is determined by the increase

of cognitive capabilities – like the capability to take other's perspective and decide how to act in a certain situation. Also, no cross-sectional study suggests an age-related increase in empathy, while the only systematic longitudinal study available indicates that self-reported empathy may decline with age, but quite modestly. Given this state of art of the literature, we still do not know whether empathy shows long-term modifications and, if so, whether long-term change in empathy depends on people's age or other person characteristics, such as a cognitive decline due to elderly (Grühn, Rebucal, Diehl, Lumley & Labouvie-Vief, 2008).

However, not always a person can choose how to react, sometimes it is an automatic response that a person has from the beginning of his life (Wellman, Cross & Watson, 2001). There are also different factors that influence the development of empathy. For example, the temperament a child is born with (Cornell & Frick, 2007), environmental, cultural factors and mental capabilities. Empathy is defined as a capability of a person to understand how another person feels, be able to feel those feelings together or be able to find a way to help solve problems that caused those feelings (de Waal, 2009). All these factors depend on different types of empathy. Affective empathy refers to the ability of a person to perceive and share other individual's emotional states and feelings (de Waal, 2009). Emotional empathy is the one that a person is born with – in other words, it is an automatic emotional response to the environmental stimulus (Martin & Clark, 1982). If a person is capable to understand how another person is feeling, or how his behavior might influence another person's feelings – that would be cognitive empathy (Decety & Lamm, 2006). Said differently, cognitive empathy is a skill that human beings develop throughout their whole life span – from the personal experiences and different types of emotions. It is learned from daily life, and in adulthood it allows a person to decide what type of response to adopt depending on the situation, or not to show any reaction at all (Batson, Ahmad & Stocks, 2004). If a person is capable to understand another person's feelings because of the situation he is in, and is able to try to find a way to resolve that problem – that would be compassionate empathy (Borg, Brenner & Berry, 2014). Thus, compassionate empathy is a more sophisticated level, which develops with age. In adulthood, one should be able to evaluate the situation, and to take an actual action to resolve it (Goleman, 2007). Empathy is considered to be the most important element of the relationship between patient and a person who works in the clinical environment (Hojat et al., 2002), and it is a crucial element to be considered during psychological assessment – as a lack of empathy characterizes several psychopathological conditions such as narcissism, antisocial disorder, and psychopathy. Despite the relevance of empathy to psychological assessment and clinical psychology, the literature on this construct is not very well organized, and for this reason, it is often difficult to find what a professional is looking for, especially when one needs to decide which instrument(s) to adopt for his/her clinical or research purposes. To fill this gap, the aim of the current article was to analyze and summarize literature of last fifteen years, so as to put all the most popular ways to measure

empathy in one place. In this way, a researcher can see what is available for him/her and can choose the type of measurement that is most likely best for him/her. As such, our goal was to find all tools of measurement, describe their advantages and disadvantages, define the structure of each instrument and, of course, describe its psychometric characteristics. Moreover, we also intended to describe the tools that appear doubtful, and that maybe would not be so good to use when measuring empathy.

METHOD

This literature review was made during October, 2015 and January, 2016. Two different databases were utilized, i.e., ProQuest and PubMed. The steps of this search were planned ahead of time. Firstly, by brainstorming, two lists of key-words were produced. In the first of these lists, some synonyms of “empathy” or related expressions were formulated; in the second, a list of words describing “measurement” was generated. The synonyms or expressions related to “empathy” utilized for the first list were: “empathetic,” “empathic,” and “empathizing”. In Appendix I, it is reported the second list of the words that were used to search for measurement.

For each search in both databases, these two categories were combined by using the option of “AND”, so that in each search one word from the empathy group and one from the measurement group were inputted. This procedure was repeated for all possible combinations, i.e., each and every key-word from both groups was searched together. The words that found at least 1 correspondence with empathy semantic group are shown in Table 1.

Secondly, some inclusion criteria were established. Specifically, it was decided to use articles in the time frame from 2000 and 2015, only in English language, only with availability of full texts, and only articles published in academic journals (the dissertations or theses were excluded). This choice aimed at restricting the field to the most psychometrically sound and widely used tools.

Initially it was also considered to search for key-words both between titles and abstracts of the articles, but millions of studies were found, and for this reason it was very difficult to understand which articles were useful and which ones were not of our interest. Therefore, ultimately it was chosen to limit our search to key-words only in the titles.

Later on, when all suitable for the criteria of the research

Table 1 – The list of key-words and frequency between each other

| | Measure | Frequency | | | Measure | Frequency | |
|-------------|---------------|-----------|-------------|---------------|------------|-----------|--------|
| | | PsycInfo | PubMed | | | PsycInfo | PubMed |
| Empathetic | Evaluation | 0 | 1 | Empathy | Scale | 1 | 0 |
| | Function | 0 | 1 | | Testing | 0 | 1 |
| | Measuring | 0 | 3 | | Tool | 0 | 0 |
| | Scale | 0 | 2 | | Assess | 0 | 2 |
| | Self-report | 1 | 1 | | Assessing | 7 | 12 |
| | Test | 0 | 1 | | Assessment | 1 | 15 |
| Empathic | Assessing | 2 | 4 | Degree | 1 | 5 | |
| | Assessment | 0 | 2 | Detection | 0 | 4 | |
| | Degree | 0 | 1 | Evaluating | 1 | 3 | |
| | Detection | 0 | 2 | Evaluation | 5 | 16 | |
| | Evaluating | 0 | 1 | Examination | 4 | 5 | |
| | Evaluation | 0 | 1 | Examining | 4 | 7 | |
| | Examining | 1 | 3 | Function | 2 | 7 | |
| | Functioning | 1 | 2 | Functioning | 5 | 22 | |
| | Identifying | 0 | 1 | Identify | 0 | 1 | |
| | Interview | 0 | 1 | Identifying | 0 | 1 | |
| | Level | 1 | 0 | Index | 10 | 7 | |
| | Measure | 1 | 3 | Instrument | 0 | 1 | |
| | Measuring | 0 | 3 | Interview | 0 | 2 | |
| | Performance | 0 | 1 | Level | 4 | 11 | |
| | Questionnaire | 0 | 2 | Measure | 8 | 23 | |
| | Rating | 1 | 0 | Measuring | 4 | 8 | |
| | Scale | 0 | 2 | Performance | 2 | 13 | |
| | Score | 1 | 1 | Questionnaire | 3 | 6 | |
| | Scoring | 0 | 0 | Quotient | 8 | 9 | |
| | Self-report | 0 | 3 | Rating | 3 | 0 | |
| Survey | 0 | 0 | Scale | 8 | 48 | | |
| Task | 0 | 1 | Score | 2 | 2 | | |
| Test | 1 | 1 | Self-report | 2 | 7 | | |
| Testing | 0 | 1 | Survey | 1 | 8 | | |
| Empathizing | Assessment | 1 | 0 | Task | 1 | 9 | |
| | Level | 0 | 1 | Test | 2 | 7 | |
| | Measuring | 0 | 1 | Testing | 4 | 7 | |
| | Quotient | 1 | 0 | Tool | 0 | 6 | |

articles were found, some categorization was made to make the research easier to follow. It was decided to save the used database, authors of the articles, the title, the abstract (if it was missing while saving, it had to be found and copied by hand), year of publication and the title of the publication. Duplicates were deleted and it was chosen to use articles that were considering adults only. Moreover, instruments assessing empathy were searched mainly in the titles, but this was not the only way to generate our list of all described measures – abstracts and whole articles were considered as well. In other words, if an abstract or an article was referencing to a specific measure it had to be included in the list of all instruments as well. Furthermore, because our goal was to present a list with the most widely investigated and/or utilized measures of empathy, if an article described more than one empathy tool, that article would eventually contribute to our list multiple times, i.e., all tools described in that article would appear in our final list. After reading all the articles, it was seen that the majority of the studies were about self-report methods and only a small part of the articles did not enter to that category. Thus, it was decided to organize the results into two chief sections – *self-reports* and *other*. That is, all the studies that were not about *self-report* methods (included those in which the methodological approach was not described clearly) were put together into one separate group, i.e., *other*.

RESULTS

After performing a literature search by the structured methods explained before, 436 articles were found in total from both databases (i.e., ProQuest and PubMed). As noted above, duplicates and articles with empathy measures not focused on adults were then excluded from the list. Thus, the working list was eventually reduced to 252 articles.

Next, we defined how to organize and categorize the instruments retrieved from those articles by reading the full texts one more time. Most of the studies were about self-report methods to measure empathy; others were considered to be performance-based and/or rating scales. More in detail, the resultant categorization was as follows: 223 articles for self-report articles, 20 for performance-based methods and 11 for rating scales methods.

Subsequently, a few additional adjustments were made. For example, our initial categorization was mainly based on

our reading of the abstracts of the articles. However, when we read the whole texts, we realized that some of the articles that we initially considered to be involving performance-based tools or rating scales, were actually using self-report methods, both performance-based and self-report, or both self-report and rating scale methods; or they used more than one method to measure empathy in one article. For this reason, the number of articles in our categorization is different from the number of total studies, as some studies were eventually included in more than one category (e.g., both self-report and rating scale).

Ultimately, it was decided to organize the results into two different types of methods to measure empathy: Those for which it was clear that they were describing *self-report* methods, and those for which the label *others* would be more appropriate, in that it was not clear what method it was used, or it was simply not a self-report method. This decision was taken because there were not enough studies for each non-self-report method to justify an additional classification category. Within the *self-report* group articles, 64 different methods to measure empathy were found. Within the group of *other* methods, there were 8 different approaches. After calculating the percentage of occurrence of each method out of all studies, we decided to further describe in this article only those tools for which the percentage of occurrence among the retrieved articles was over 1% (see Table 2). Below, a brief description of all these selected instruments follows.

Most widely used methods to measure empathy-self-report

As noted above, the big majority of the found studies were consisted of self-report measurement tools to assess empathy. Out of all the studies of self-report articles, it was chosen to further inspect only those which were over 1%. After calculating the percentage it was seen that in this group enter 14 different instruments. Each of them will be shortly presented below. In Table 3 we present all the instruments that were found in the articles of self-report methods: only the first 14, i.e., those that exceed 1% of occurrence, will be described below.

- *Interpersonal Reactivity Index (IRI)* (Davis, 1980, 1983). The IRI is a 28-item, self-report instrument to measure different reactions and personal experiences of one individual while observing the other (Davis, 1983). The

Table 2 – Summary of empathy measures and frequency of use among selected articles

| Instrument | | Times | % all studies (N = 252) |
|-------------|---------------------------------------------------------------|-------|----------------------------|
| Self-report | IRI | 43 | 17.10% |
| | Jefferson Scale of Physician Empathy | 33 | 13.10% |
| | Empathy Quotient | 25 | 9.90% |
| | Consultation and Relational Empathy | 11 | 4.40% |
| | Jefferson Scale of Empathy | 10 | 4.00% |
| | Likert scale | 8 | 3.20% |
| | QCAE | 6 | 2.40% |
| | Empathy Concern Scale | 5 | 2.00% |
| | Jefferson Scale of Patient's Perceptions of Physician Empathy | 5 | 2.00% |
| | Jefferson Scale of Empathy–Health Profession Students version | 5 | 2.00% |
| | Empathy–Based Stories | 3 | 1.20% |
| | Basic Empathy Scale | 3 | 1.20% |
| | Questionnaire Measure of Emotional Empathy | 3 | 1.20% |
| | Toronto Empathy Questionnaire | 3 | 1.20% |
| Other | fMRI activation paradigm | 8 | 3.20% |
| | EMG | 4 | 1.60% |
| | Reading the Mind in the Eyes | 4 | 1.60% |
| | EEG activity | 3 | 1.20% |

Note. Times = number of times this instrument was met in the papers.

questions have to be answered on 5-point Likert scale from 1 = “Does not describe me well” to 5 = “Describes me very well” (Davis, 1983; Davis & Oathout, 1987). It was designed to measure different empathic tendencies: a) *Perspective Taking* (PT); b) *Fantasy* (FS); c) *Empathic Concern* (EC); d) *Personal Distress* (PD). Every each of them is made up of seven various items. The homogeneity of the different scales of IRI are quite good, the Cronbach's α coefficients are ranging from .68 to .79. The previous

studies also showed that IRI subscales of PT and FS are related to cognitive empathy and that different subscales of this instrument vary in between cognitive and emotional characteristics of empathy in adults with autism (Rogers, Dziobek, Hassenstab, Wolf & Convit, 2007).

- *Jefferson Scale of Physician Empathy* (JSPE; Hojat et al., 2001, 2002, 2003). JSPE is a 20-item, self-report questionnaire that measures different components of empathy between physicians in patient-care environment. Questions had to

Table 3 – Percentage of the use of all self-report empathy measures included in this study

| Instrument | Times | Percentage | |
|------------------------------------------------------------------------------|-------|--------------------------|--------------------------|
| | | Self-report (n = 223) | All studies (N = 252) |
| IRI | 43 | 19.30% | 17.10% |
| Jefferson Scale of Physician Empathy | 33 | 14.80% | 13.10% |
| Empathy Quotient | 25 | 11.20% | 9.90% |
| Consultation and Relational Empathy | 11 | 4.90% | 4.40% |
| Jefferson Scale of Empathy | 10 | 4.50% | 4.00% |
| Likert scale | 8 | 3.60% | 3.20% |
| QCAE | 6 | 2.70% | 2.40% |
| Empathy Concern Scale | 5 | 2.20% | 2.00% |
| Jefferson Scale of Patient's Perceptions of Physician Empathy | 5 | 2.20% | 2.00% |
| Jefferson Scale of Empathy–Health Profession Students version | 5 | 2.20% | 2.00% |
| Empathy–Based Stories (MES) | 3 | 1.30% | 1.20% |
| Basic Empathy Scale | 3 | 1.30% | 1.20% |
| Questionnaire Measure of Emotional Empathy | 3 | 1.30% | 1.20% |
| Toronto Empathy Questionnaire | 3 | 1.30% | 1.20% |
| The Empathy Scale of the Impulsiveness–Venturesomeness–Empathy Questionnaire | 2 | .90% | .80% |
| Scale of Ethnocultural Empathy | 2 | .90% | .80% |
| Emotional Empathy Scale | 2 | .90% | .80% |
| The Empathy (E) scale | 2 | .90% | .80% |
| Interpersonal Reactivity Index for Couples | 2 | .90% | .80% |
| Victim Empathy Response Assessment | 2 | .90% | .80% |
| Hogan Empathy Scale | 2 | .90% | .80% |
| Balanced Emotional Empathy Scale | 2 | .90% | .80% |
| Reynolds Empathy Measure | 2 | .90% | .80% |
| Empathy Index | 2 | .90% | .80% |
| Parental Empathy Measure | 2 | .90% | .80% |
| Qualitative Short Survey | 2 | .90% | .80% |
| Test of Emotional Perception | 1 | .40% | .40% |
| The Perceived Empathic Self–Efficacy Scale | 1 | .40% | .40% |
| Global Rating of Empathy scale | 1 | .40% | .40% |
| Therapist Empathy Scale | 1 | .40% | .40% |
| The Schutte Emotional Intelligence scale | 1 | .40% | .40% |

continued on next page

continued

| Instrument | Times | Percentage | |
|--------------------------------------------------------|-------|--------------------------|--------------------------|
| | | Self-report (n = 223) | All studies (N = 252) |
| Empathy for Pain Scale | 1 | .40% | .40% |
| Empathic Tendency Scale | 1 | .40% | .40% |
| Fictional IRI | 1 | .40% | .40% |
| Empathy for Infant Pain video program | 1 | .40% | .40% |
| Affective and Cognitive measure of Empathy | 1 | .40% | .40% |
| JSE S-Version | 1 | .40% | .40% |
| Empathetic Care Scale | 1 | .40% | .40% |
| Empathy Assessment Index | 1 | .40% | .40% |
| Toronto Composite Empathy Scale | 1 | .40% | .40% |
| Barrett-Lennard empathy subscale | 1 | .40% | .40% |
| Quotient of Empathic Abilities | 1 | .40% | .40% |
| Scale of Ethnocultural Empathy | 1 | .40% | .40% |
| Interaction Response Scale for Palliative Care Nursing | 1 | .40% | .40% |
| EMPATHY | 1 | .40% | .40% |
| Mehrabian Emotional Empath Scale | 1 | .40% | .40% |
| EMPATHIC questionnaire | 1 | .40% | .40% |
| Questionnaire Measure of Emotional Empathy | 1 | .40% | .40% |
| How I Feel in Different Situations Scale | 1 | .40% | .40% |
| Emotional Recognition subscale | 1 | .40% | .40% |
| Penner's Prosocial Personality Battery | 1 | .40% | .40% |
| Emotional Empathy Tendency scale | 1 | .40% | .40% |
| Empathy Response Index | 1 | .40% | .40% |
| Empathy for others pain | 1 | .40% | .40% |
| Intellectual empathy | 1 | .40% | .40% |
| Emotional Perspective-Taking scale | 1 | .40% | .40% |
| "Big Three" | 1 | .40% | .40% |
| Emotion Specific Empathy questionnaire | 1 | .40% | .40% |
| Objective Structured Clinical Examination | 1 | .40% | .40% |
| Empathic Communication Coding System | 1 | .40% | .40% |
| Empathic Opportunity | 1 | .40% | .40% |
| Kiersma-Chen Empathy Scale | 1 | .40% | .40% |
| Empathic Accuracy (EA) Task | 1 | .40% | .40% |
| Relationship Evaluation Questionnaire (RELATE) | 1 | .40% | .40% |

Note. Times = number of times this instrument was met in the papers

- be answered by Likert scale from 1 to 7, where 1 means strongly disagree and 7 means strongly agree and are divided into two types: one half is positively worded and the other one is negatively worded. The higher score on this test a person gets, the higher level of empathy it shows (Yang & Yang, 2013). This instrument is translated into 25 languages so the reliability of it varies, but it is almost always significantly high: Cronbach's α is in the range of .80, the test-re-test reliability coefficient is .65 (Hojat & LaNoue, 2014). In the previous studies, it was found the correlation between IRI and JSPE ($r = .45, p < .01$) but it was not significantly high (Hojat, Mangione, Gregory, Kane & Gonnella, 2005).
- *Empathy Quotient (EQ)*; Baron-Cohen & Wheelwright, 2004). EQ is a 60-item self-report tool to measure empathy. Questions of this questionnaire have to be answered on a scale from 0 to 2, where the higher score means a higher level of empathy. The EQ measures three different factors – *cognitive empathy*, *emotional reactivity* and *social skills* (Lawrence, Shaw, Baker, Baron-Cohen & David, 2004). The Cronbach's α varies from the language the questionnaire is translated to, but more or less is always in the range of .85, so it is significantly quite meaningful (Melchers, Montag, Markett & Reuter, 2015). The test-retest results in previous studies show high temporal stability ($r = .97, p < .01$) for all of the items (Melchers et al., 2015). Also, the medium correlation between IRI's *perspective taking* (PT) and *empathic concern* (EC) scales and EQ were found (Melchers et al., 2015). That shows that EQ is very useful to measure cognitive empathy.
 - *Consultation and Relational Empathy (CARE)*; Mercer & Reynolds, 2002). CARE is a self-report 10-item questionnaire that was developed to measure the consultations based on the standardized and generic patient-reported assessment (Wirtz, Boeckerb, Forkmann & Neumann, 2011). It contains 4 different components: emotional, ethical, behavioral and cognitive. The answers had to be chosen in the scale from 1 to 5, where 1 is poor and 5 is excellent (Wirtz et al., 2011). The original version of CARE Cronbach's α is significantly high, the value is .93 (Mercer et al., 2004). The previous studies also showed correlation between CARE and Reynolds Empathy Scale (RES) which was strong ($r = .85, p < .001$), and strong correlation between CARE and Barrett-Lennard Empathy Subscale (BLESS) ($r = .84, p < .001$) (Mercer et al., 2004).
 - *Jefferson Scale of Empathy (JSE)*; Hojat et al., 2001). JSE is a self-report 20-item instrument that was developed to value empathy in the environment of medical education and the care of the patients. The questionnaire has to be answered by Likert scale of 7 points from strongly disagree (1) to strongly agree (7). JSE is divided into three components: *perspective taking* (that is seen as a very important part of empathy), *compassionate care* (that is seen as an essential level of patient-physician relationship) and the third one is *walking in patient's shoes* (Hojat & LaNoue, 2014). The reliability of JSE is significantly quite high (Cronbach's α value is .8 that varies depending on the language that the questionnaire is translated to, from .7 to .8) (Hojat & LaNoue, 2014)
 - *Likert scale R* (Likert, 1932). In this research Likert scale was used in various forms, where the scales varied from 5 to 10 points, and in different contexts. However, there were a few articles where it was explained the type of Likert scale they were using, but not the questionnaire or the instrument to measure empathy. In any case, Likert scale was used as a self-report instrument, to answer the questions that specific authors developed for that study. However, in none of these studies the name of the questionnaire was mentioned. For this reason, there were some doubts on whether to put it together with self-report instruments or to put it together with other instruments to measure empathy. Eventually, it was decided to leave it together with the self-report measurement instruments because it was used to measure personal experienced empathy.
 - *The Questionnaire of Cognitive and Affective Empathy (QCAE)*; Reniers, Corcoran, Drake, Shryane & Vollm, 2011). The QCAE is a self-report, 31-item questionnaire that was developed to measure affective and cognitive empathy (Reniers et al., 2011). Questionnaire consist of two components mentioned before and is divided into five different subscales. The component of cognitive empathy comprises two subscales: a) *Perspective Taking* (that consists 10 items), which lets asses to see how one person is able to see the situation from another person's perspective; b) *Online Simulation* (that consists 9 items), which lets see how another person is able to understand and mentally represent how another person is feeling. Other three subscales measures the affective empathy: a) *Emotion Contagion* (that consists 4 items), which lets see how the person is able to reflect self-oriented emotions while noting the emotional states of others; b) *Proximal Responsivity* (that consists 4 items), which measures one

person's emotional reaction to the moods of another person, who is physically or emotionally close to him; c) *Peripheral Responsivity* (that consists 4 items), which lets to measure people's emotional reaction to the state of moods of another person, who is not close to them or they do not know that person at all. Every subscale has to be measured by the Likert type scale of 4 points (Reniers et al., 2011). The previous studies showed that the inter reliability between QCAE subscales were quite sensible (α 's varied between .75 and .91) except of peripheral responsivity response that was lower, $\alpha = .42$ (Michaels et al., 2014; Reniers et al., 2011). Also, the medium correlations between QCAE and IRI subscales were found. The QCAE cognitive empathy subscales revealed the highest correlations with IRI PT subscale ($r = .63, p < .001$), medium correlations with IRI FS and EC subscales ($r = .39, p < .05$), and negative correlation with IRI PD subscale ($r = -.33, p < .05$) (Michaels et al., 2014).

- *Empathy Concern Scale* (EC; Davis, 1980, 1983). Empathy Concern Scale is one of the 4 subscales in the questionnaire of IRI. Sometimes it is used as a separate instrument but maintains the same characteristics of the main tool. In particular, it measures the ability of the person to express feelings of empathy and concern for other people who are in unfortunate situations (affective empathy). Previous studies showed that the reliability of this scale is not so high – the Cronbach's α is .52 (Leong, Cano, Wurm, Lumley & Corley, 2015).
- *Jefferson Scale of Patient's Perceptions of Physician Empathy* (JSPPPE; Hojat et al., 2003). The JSPPPE is a short, self-report tool of 5 items. It is unidimensional scale that measures only one factor: empathic engagement. Response options are given in 5 points Likert scale, where 1 means strongly disagree and 5 means strongly agree. JSPPPE is given for the patients to complete, to measure the empathy of the physicians from the perspective of the patients. The Cronbach's α is .58, which in general is a low result, but considering the fact that this questionnaire has only 5 items, it might be satisfactory. Previous studies showed a very low correlation between JSPPPE and Jefferson Scale of Physician Empathy ($r = .24, p = .22$) which was not considered as significantly important (Kane, Gotto, Mangione, West & Hojat, 2007).
- *Jefferson Scale of Empathy-Health Profession Students version* (JSE-HPS; Hojat, 2007). JSE-HPS is a self-report questionnaire that was developed to measure the level of

empathy between students in healthcare environment. It is a 20-item measuring instrument that is divided into three sub-factors: a) *Perspective Taking*; b) *Compassionate Care* (or emotional engagement); c) *Standing in the Patient's Shoes* (Hojat et al., 2002). It has to be answered in the 7 point Likert scale (1 is strongly disagree, 7 is strongly agree). The Cronbach's α through all the subscales vary between .80 and .89. The previous studies showed that students, whose grades are better, are more likely to have a higher level of empathy (Montanari et al., 2015).

- *Empathy-Based Stories* (MES; Eskola, 1998). The MES is a qualitative research method that is used in different fields (i.e. sociology, pedagogy, social psychology). The aim of this instrument is to see different points of view of different people, to see their personal experiences in emotionally heavy situations. This instrument works by giving a participant a sheet of paper with a short story on the top of the paper, those stories might be very different, can be used in different fields and depends on the examiner (Juntunen & Saarti, 2000). After a person gets a paper he has to finish the story that was started on the same paper and can write it only on one side of the sheet. This instrument usually is used in the situation where the opinion of the participant is very important (for example, when an institution wants to improve their offered services). This instrument is useful because a person is completely free to write everything he wants: his observations on the situation mentioned in the paper, can concentrate on the future and give his honest opinion (positive or negative) or emotions, and can choose the highlights of the situation. To analyze the data of this instrument, matrixes are used. In every study, where this measurement is used, matrixes are different, so it depends completely on the examiner what he wants to test or improve (Juntunen & Saarti, 2000). The very important factor using this measure tool is that in certain situations, the self-esteem of the person does influence on the results. The more he/she is confident, the better stories can be written. Sometimes, when MES is used in one certain area it can be useful to bring some people who do not have anything in common with the area to write stories. It can be very helpful when analyzing the results.
- *Basic Empathy Scale* (BES; Jolliffe & Farrington, 2006). BES is a self-report 20-items instrument which was developed to measure affective and cognitive empathy. The questionnaire is divided into two types of subscales: 9-item *Cognitive Empathy Subscale* (which measures the understanding

of other people's emotions – a person understands why another person is happy) and 11-item *Affective Empathy Scale* (which measures the emotional part of the person – an individual becomes sad while watching a sad movie). Each item from both subscales has to be answered by a Likert type scale of 5 points, the higher number of points means a higher level of empathy. The Cronbach's α for the affective empathy subscale is .81, and for the cognitive empathy subscale it is .84. Also, it has shown a good model fit (Baldner & McGinley, 2014). Previous studies have shown that the correlation between an Affective Empathy Subscales of BES and Interpersonal Reactivity Index are higher ($r = .51 - .64$) than Cognitive Empathy Subscales ($r = .31 - .49$) (Baldner & McGinley, 2014).

- *Questionnaire Measure of Emotional Empathy (QMEE;* Mehrabian & Epstein, 1972). The QMEE is a self-report measurement tool that was developed to measure emotional empathy. It is a 33-item instrument that has to be answered by a 9 point ratings from -4 to 4, where -4 means a strong disagreement and +4 means a strong agreement. QMEE has 7 subscales: a) *vulnerability to emotional contamination*; b) *appreciation of the feelings of others that a person does not know*; c) *strong emotional response*; d) *tendency to be emotionally touched by other people positive emotional capabilities*; e) *tendency to be emotionally touched by other people negative emotional capabilities*; f) *sympathetic capacities*; g) *wish to be in contact with people that have problems* (Mehrabian & Epstein, 1972). The reliabilities of the subscales are fair to adequate – the Cronbach's α s vary from .63 to .80 (Lyons & Hazler, 2002).
- *Toronto Empathy Questionnaire (TEQ;* Spreng et al., 2009). TEQ is a self-report instrument that was developed to measure different types of empathy. It is a 16-item measure that has only one empathy scale (just like empathy quotient) (Baldner & McGinley, 2014). It has to be answered by a Likert type scale of 5 points, where more points mean a higher level of empathy. This instrument has equal number of positively and negatively worded items (so when the item is worded negatively, the response to it has to be also reversed). It was also proven that TEQ is a good model fit (Baldner & McGinley, 2014) and its reliability is adequate (Cronbach's α is .88). Previous studies found that there is a positive correlation between TEQ and IRI *Empathic Concern Scale* ($r = .74, p < .001$), also, a low correlation between TEQ and IRI *Perspective Taking* was found ($r = .35, p < .001$) (Baldner & McGinley, 2014).

Most widely used methods to measure empathy: other

As it was explained before, we created the category other that includes performance based or rating scales methods for the empathy assessment, because we found few methods that measure empathy in a different way from self-report questionnaire. Furthermore, for the self-report category, we decided to describe only the instruments that were present in the studies over the 1%. However, the total number of the instruments present in the other category is 8, as it can be seen in Table 4. So, because they are a few number of tools, we decided to describe all the instruments present in this group.

- *fMRI activation paradigm* (Vollm et al., 2006). The fMRI activation paradigm is used to expose activation areas related with empathy processing. It is a visual activation paradigm, which consist of a series of cartoons (which represent different short stories in every block of pictures). The cartoons can contain a two type stories – Physical and Empathy. At the beginning of this procedure, a series of questions are given to participants (that helps to see the same mental construct corresponding in all participants). The scenarios of each series are made that the characters of cartoons continue their story in upcoming pictures. Because of this reason, a participant has to be empathizing for the protagonist of the story. The questions contain a text asking about what a person thinks will happen next in the picture, or what s/he thinks that the protagonist of the cartoon is feeling at that particular moment. Each of the blocks (two pictures telling the same story at the time) is shown for only four seconds in the upper part of the screen, then, for other four seconds, the possible endings of that story (other two pictures) are shown at the bottom part of the screen. The participant has to choose the answer to the question that is given before, and only one of the two possibilities is right. It is considered, that more correct answers a person gives, a higher level of empathy it has; also, by doing fMRI study, it is seen which part of the brain is active while doing a part regarding empathy (Kim et al., 2010).
- *Electromyography (EMG)*. This instrument is used to capture the electrical activity of facial muscles, because it is believed that facial muscle reactions are related to emotional reactions (Ekman & Friesen, 1975; Tomkins, 1991). The intention of this tool is to catch all the facial reaction in the face-to-face situations. Moreover, the time of the reaction is also important: faster a person shows

Table 4 – Percentage of the use of all Other empathy measures included in this study

| | Times | % in Other (n = 31) | % in All studies (N = 252) |
|--------------------------------------|-------|------------------------|-------------------------------|
| fMRI activation paradigm | 8 | 25.80% | 3.20% |
| EMG | 4 | 12.90% | 1.60% |
| Reading the Mind in the Eyes Test | 4 | 12.90% | 1.60% |
| EEG activity | 3 | 9.70% | 1.20% |
| Performance-Based Measure of Empathy | 2 | 6.50% | .80% |
| Multifaceted Empathy Test | 2 | 6.50% | .80% |
| Story-Based Empathy Task | 1 | 3.20% | .40% |
| Social Relations Model | 1 | 3.20% | .40% |

Note. Times = number of times this instrument was met in the papers.

his reaction, a better understanding of other's feelings are considered that the person has. However, in the previous studies it was found that reactions can be learned or controlled by the person, so it is not always a good way to measure empathic responses (Sonnby-Borgstrom, Jonsson & Svensson, 2003).

- *Reading the Mind in the Eyes Test (RMET;* Baron-Cohen et al., 1997). The RMET is a performance based instrument that was developed to assess the ability of the person to read another person's emotions based on only looking to the line pictures of the eyes. This tool lets to see and measure the mental states of oneself and others. It contains 36 pictures of the eyes, that represents different emotions, the person is instructed to choose one out of four offered words (that describes different emotions), that they personally think signifies the current emotion in the photo. This measure is usually used with people who has Asperger Syndrome, as they have major difficulties on recognizing the emotions of other people. The medium response for this test is 26.2, or 21.9 for people with Asperger Syndrome or High-functioning Autism. A notable issue about this test is that the words given to choose from might be not clear or familiar to the person, so that also might be a

consequence of the lower result. The previous studies have showed that no correlation exist between RMET and IQ ($r = .09, p = .6$) (Baron-Cohen, Whellwright, Hill, Raste & Plumb, 2001).

- *Electroencephalography (EEG) activity.* The EEG method is an instrument that helps to see the different level of empathy in different people, as EEG responses vary by doing requested task while being recorded. This method works by showing different types of pictures on the computer to the participant. After he is attached to EEG apparatus, it is recorded by doing this protocol: a) for 3 minutes the person is recorded in the resting state with his eyes closed; b) different pictures that induce positive emotions are shown on the computer screen for 1,5 minutes; c) different pictures with neutral stimuli are shown for one and a half minutes; d) erotically colored pictures are shown for 1,5 minutes; e) pictures inducing negative stimuli are shown for another one and a half minutes; f) lastly, the EEG is recorded again in the resting state for 3 minutes with the eyes closed. In between of each series some grey-colored pictures with meaningless context are presented for 1,5 minutes. After that EEG was recorded a participant is asked to value every block

- of pictures that they have seen before by the scales of 9 points (the one that gave most pleasure, that were neutral, positive, most unpleasant or erotic). Then it is asked to value the pictures in the same way that were maximally activating, most calming or neutral. Persons that received most points on emotionally active pictures were considered that are more empathic than those who got lower scores. Also, by doing EEG measure, it is possible to see which part of the brain activates when a person watches different pictures that represent different stimuli (Demidova, Dubovik, Kravchenko & Makarchouk, 2014).
- *Performance-Based Measure of Empathy* (Derntl et al., 2009). The Performance-Based Measure of Empathy is a tool of three tasks that a person has to take on a computer, which includes: a) facial affect perception; b) affective responsiveness; c) emotional perspective taking. Tasks are time registered (the reaction time is one of the measure of the test), and the two forced-choice responses are given, so an accuracy of the responses are also considered (Derntl et al., 2009). A person has to recognize the emotion in the shown picture and choose the answer as fast as possible, the less time it takes and the better accuracy a person gets, a higher level of empathy and higher understanding about emotional states of other person is considered that the participant has (Derntl et al., 2009).
 - *Multifaceted Empathy Test (MET)* (Dziobek & Heekeren, 2008). The MET is a rating scale type of instrument to measure cognitive and affective empathy. It consists of different series of the photographs, where mostly people are in emotionally stimulating situations. In the part of the test where the cognitive empathy is measured, it is asked for a person to try to name the mental states of the persons who are in the given photographs. After this part a person is informed about the correct answers that he gave. Later on, to measure the emotional empathy, it is asked for the participants to rate their personal emotional reactions that they experience while looking in to the given pictures. The MET consists of 23 pairs of different pictures (that involves one context and one person picture), while looking to those pictures an individual is asked some questions (Dziobek & Heekeren, 2008). For the pictures that consists a context, a person is asked to rate his level of excitement by using *Self-Assessment Manikin (SAM)* (Lang et al., 1997). It is a visual-analogue scale that has a rating scale from 0 to 9 (0 = very calm; 9 = very excited). For the cognitive part of the test, a participant is asked to name the mental states in depicted persons, which they have to choose one out of four given in the test. The reliability of the MET varies from medium to high: Cronbach's α ranges from .71 to .92. Moreover, the correlations between MET cognitive empathy and IRI *Perspective Taking* scales were found ($r = .28, p < .01$), but were not significantly meaningful, and the correlation between MET affective empathy and IRI *Empathic Concern* ($r = .63, p < .01$) were found (Dziobek & Heekeren, 2008).
 - *Story-Based Empathy Task (SET)* (Dodich et al., 2015). SET is a non verbal task that was developed to measure an intention and emotion acknowledgment. This task is based on original cartoons, it takes about 15–20 minutes to accomplish this test and is consisted of two main experimental conditions and one control condition (Dodich et al., 2015). The main conditions are: a) identifying intentions (SET-IA); b) emotional states (SET-EA); the control condition is the inference of causality reaction (SET-CI) that is based on the personal knowledge of the individual of physical properties of objects and human bodies (Dodich et al., 2015, p.1908). Each of these conditions are composed of six pictures and then it is asked to choose a possible ending for that story given in the pictures (but the possible endings are given only later). Each of the parts can be valued maximum of 6 points (1 point per every correct choice), so the total score and the best possible task performance is 18 points. More points a participant gets, higher level of understanding about other's feelings it is considered that he has (Dodich et al., 2015).
 - *Social Relations Model (SRM)* (Kenny & La Voie, 1988). SRM is a rating scale type of measure that was developed to describe dyadic relationships when the components are assessed on a continuous scale. SRM is divided into three components: a) perceiver (helps to understand how the participant sees other people); b) target (helps to understand how the person itself is seen by other people); c) relationship (helps to understand how a perceiver sees the target) (Kenny, Mohr & Levesque, 2001). Two ways to use this model can be used – round robin or block. Round robin is a way that is mostly used in interpersonal perception studies (every member of the group has to rate or judge every other person in the group). The results of round robin are calculated by using the computer program SOREMO. Another way of the model is block (a group is divided into two smaller subgroups, and each person from each subgroup has to rate or judge every person from other subgroup). SRM

is an instrument that helps a participant to understand, if other people do understand his emotions and helps to see, if a participant itself understands other individual's emotions (Kenny et al., 2001).

CONCLUSIONS

Despite the increasing interest in the construct of empathy, both from a clinical and from a research perspectives, the literature describing different methods to measure empathy is not well organized. To facilitate the work of researchers and practitioners willing to measure empathy, the current article describes the most frequently utilized instruments available to date. The results of our literature search showed that the most popular ways to measure empathy are self-report style instruments (e.g., IRI, Jefferson's Scale of Empathy, Empathy Quotient). This result is not too surprising, given that self-reports are easy to use, faster to analyze and often produce valid and reliable scores. On the other hand, it should be noted that self-report are subject to social desirability, and it is controversial to ask a person to evaluate his or her empathetic abilities, given that people with poor empathetic skills are likely to not be aware about their ability to perceive, understand and share the emotional states of the others. In line with this position, a number of other methods to measure empathy were also found (e.g., fMRI paradigm).

Like all studies, this literature review is also not without limitations. Firstly, we mainly relied on the titles of the articles (although we also read the abstract and the full text of all of the selected articles). In the future studies, by adding abstracts and key-words it might be that some additional information could be found. Furthermore, in the future studies it might be considered to use different key-words, because it is not possible to check if the key-words that were used in this study were exhaustive. Likewise, future studies might inspect older articles. Also, the most popular instrument between other type of measurements was fMRI, which is not actually a measurement tool that can exactly measure a level or type of empathy, which would be useful

in the field of clinical psychology. Lastly, comparing the frequency of citation of each of these measures may not be the best approach to identify which instruments are actually the most used ones, or can be considered as the best ones to use to measure empathy. For example, if an instrument was developed long time ago, but then it got discarded because it was not very good to measure empathy, our review would probably still list that instrument as one of the most cited ones. Moreover, in this article it is possible to recognize which instrument is good for the assessment of both types of empathy, which is good for the affective one (e.g. the implicit variables) and which is better for the cognitive one; however, our review does not suggest whether some measuring instruments are better for one professional or another. It represents the overall view of all possible tools that are available, by describing the positive and negative sides of it, but it is up to the professional to decide whether to use one instrument or another. For example, self-report measures can be very successfully used as the tool to assess empathy as a trait, where other types of instruments can be used better in situations where the valuation of other people is needed. This study can help researchers to choose an instrument, as he/she has a full view of what is out there. However, even if this article is useful to provide an overall view of most used instruments, it does not provide an organized literature that is able to inform any professionals on what measure is suitable for his/hers work.

Despite these limitations, our study still has a merit to be the first one to organize the literature on empathy measures by the most used ones in the last fifteen years. Instruments to measure empathy and statistical usage of them were never presented together and grouped like this in one study before. The aim of this study was also to summarize all the instruments that are given during last 15 years, especially for the researchers that are willing to investigate empathy capabilities. This literature review might be useful for those who are exploring empathy and are searching for new types of instruments to measure it.

References

- BALDNER, C. & MCGINLEY, J.J. (2014). Correlational and exploratory factor analyses (EFA) of commonly used empathy questionnaires: New insights. *Motivation and Emotion*, 38 (5), 727–744.
- BARON-COHEN, S., WHELLWRIGHT, S., HILL, J., RASTE, Y. & PLUMB, I. (2001). The “Reading the Mind in the Eyes Test revised version: A study with normal adults, and adults with Asperger Syndrome or High-functioning Autism. *Journal of Child Psychology*, 42 (2), 241–251.
- BATSON, C.D., AHMAD, N. & STOCKS, E.L. (2004). Benefits and liabilities of empathy-induced altruism. In A.G. Miller (Ed.), *The social psychology of good and evil*. New York: Guilford Press.
- BORG, M.B. Jr., BRENNER, G.H. & BERRY, D. (2014). A New Understanding of Compassionate Empathy. *Psychology today*.
- CORNELL, A.H. & FRICK, P.J. (2007). The moderating effects of parenting styles in the association between behavioral inhibition and parent-reported guilt and empathy in preschool children. *Journal of Clinical Child and Adolescent Psychology*, 36, 305–318.
- DAVIS, M.H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44 (1), 113–126.
- DAVIS, M.H. & OATHOUT, H.A. (1987). Maintenance of satisfaction in romantic relationship: Empathy and relational competence. *Journal of Personality and Social Psychology*, 53, 397–410.
- DECETY, J. & LAMM, C. (2006). Human empathy through the lens of social neuroscience. *Scientific World Journal*, 6, 1146–1163.
- DEMIDOVA, K.Y., DUBOVIK, V.V., KRAVCHENKO, V.I. & MAKARCHOUK, N.E. (2014). EEG Activity during Viewing of Neutral and Emotionally Colored Pictures by Men and Women with Different Levels of Empathy. *Neurophysiology*, 46 (2).
- DERNTL, B., FINKELMEYER, A., TOYGAR, T.K., HULSMANN, A., SCHNEIDER, F., FALKENBERG, D.I. & HABEL, U. (2009). Generalized deficit in all core components of empathy in schizophrenia. *Schizophrenia Research*, 108 (1–3), 197–206.
- DE WAAL, F.B.M. (2009). Putting altruism back into altruism: The evolution of empathy. *Annual Review of Psychology*, 59, 279–300.
- DODICH, A., CERAMI, C., CANESSA, N., CRESPI, C., IANNACCONE, S., MARCONE, A., REALMUTO, S., LETTIERI, G., PERANI, D. & CAPPA, S.F. (2015). A novel task assessing intention and emotion attribution: Italian standardization and normative data of the Story-based Empathy Task. *Neurological Science*, 36, 1907–1912.
- DZIOBEK, I. & HEEKEREN, H.R. (2008). Dissociation of cognitive and emotional empathy in adults with Asperger syndrome using the Multifaceted Empathy Test (MET). *Journal of Autism and Developmental Disorders*, 38 (3), 464–473.
- EKMAN, P. & FRIESEN, W.V. (1975). *Unmasking the Face*. Englewood Cliffs, New Jersey: Spectrum-Prentice Hall.
- GOLEMAN, D. (2007). *Three kinds of empathy: Cognitive, Emotional, Compassionate*. <http://www.danielgoleman.info/three-kinds-of-empathy-cognitive-emotional-compassionate/>
- GRUHN, D., REBUCAL, K., DIEHL, M., LUMLEY, M. & LABOUVIE-VIEF, G. (2008). Empathy across the adult lifespan: Longitudinal and experience-sampling findings. *Emotion*, 8, 753–765.
- HOJAT, M., GONNELLA, J.S., MANGIONE, S., NASCA, T.J., VELOSKI, J.J., ERDMANN, B., CALLAHAN, C.A. & MAGEE, M. (2002). Empathy in medical students as related to academic performance, clinical competence and gender. *Medical Education*, 36, 522–527.
- HOJAT, M., MANGIONE, S., GREGORY, C., KANE, G.C. & GONNELLA, J.S. (2005). Relationships between scores of the Jefferson Scale of Physician Empathy (JSPE) and the Interpersonal Reactivity Index (IRI). *Medical Teacher*, 27, 625–628.
- HOJAT, M. & LANOUE, M. (2014). Exploration and confirmation of the latent variable structure of the Jefferson scale of empathy. *International Journal of Medical Education*, 5, 73–81.
- JUNTUNEN, A. & SAARTI, J. (2000). Library as the Student’s Cornerstone or Obstacle: Evaluating the Method of Empathy-Based Stories. *Libri*, 50, 235–240.
- KANE, G.C., GOTTO, J.L., MANGIONE, S., WEST, S. & HOJAT, M. (2007). Jefferson Scale of Patient’s Perceptions of Physician Empathy: Preliminary Psychometric Data. *Croat Med J*, 48, 81–86.
- KENNY, D.A., MOHR, C.D. & LEVESQUE, M.J. (2001). A Social relations variance partitioning of dyadic behavior. *Psychological Bulletin*, 127 (1), 128–141.
- KIM, Y.T., LEE, J.J., SONG, H.J., KIM, J.H., KWON, D.H., KIM, M.N., YOO, D.S., LEE, H.J. & CHANG, Y. (2010). Alterations in cortical activity of male methamphetamine abusers performing empathy task: fMRI study. *Human Psychopharmacology*, 25, 63–70.
- LAWRENCE, E.J., SHAW, P., BAKER, D., BARON-COHEN, S. & DAVID, A.S. (2004). Measuring empathy: Reliability and validity of the empathy quotient. *Psychological Medicine: A Journal of Research in Psychiatry and the Allied Sciences*, 34 (5), 911–919.
- LEONG, L.E.M., CANO, A., WURM, L.H., LUMLEY, M.A. & CORLEY, A.M. (2015). A perspective-taking manipulation leads to greater empathy and less pain during the cold pressor task. *The*

- Journal of Pain*, 16 (11), 1176–1185.
- LYONS, C. & HAZLER, R.J. (2002). The influence of student development level on improving counselor student empathy. *Counselor education and supervision*, 42.
- MARTIN, G.B. & CLARK, R.D. (1982). Distress crying in neonates: Species and peer specificity. *Developmental Psychology*, 18, 3–9.
- MEHRABIAN, A. & EPSTEIN, N. (1972). A measure of emotional empathy. *Journal of Personality*, 40 (4), 525–543.
- MELCHERS, M., MONTAG, C., MARKETT, S. & REUTER, M. (2015). Assessment of empathy via self-report and behavioural paradigms: Data on convergent and discriminant validity. *Cognitive Neuropsychiatry*, 20 (2), 157–71.
- MERCER, S.W., MAXWELL, M., HEANEYD, D. & WATT, G. (2004a). The consultation and relational empathy (CARE) measure: Development and preliminary validation and reliability of an empathy-based consultation process measure. *Oxford Journals Medicine & Health Family Practice*, 21 (6), 699–705.
- MERCER, S.W., McCONNACHIE, A., MAXWELL, M., HEANEYD, D. & WATT, G. (2004b). Relevance and practical use of the Consultation and Relational Empathy (CARE) Measure in general practice. *Oxford Journals Medicine & Health Family Practice*, 22 (3), 328–334.
- MICHAELS, T.M., HORAN, W.P., GINGERA, E.J., MARTINOVICHA, Z., PINKHAMD, A.E. & SMITH, M.J. (2014). Cognitive empathy contributes to poor social functioning in schizophrenia: Evidence from a new self-report measure of cognitive and affective empathy. *Psychiatry Research*, 220 (3), 803–810.
- MONTANARI, P., PETRUCCI, C., RUSSO, S., MURRAY, I., DIMONTE, V. & LANCIA, L. (2015). Psychometric properties of Jefferson Scale of Empathy–Health Professional Student’s version: An Italian validation study with nursing students. *Nursing and health sciences*, 17, 483–491.
- O’BRIEN, E., KONRATH, S.H., GRUHN, D. & HAGEN, A.L. (2012). Empathic concern and perspective taking: Linear and quadratic effects of age across the adult life span. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*. doi:10.1093/geronb/gbs055.
- RENIERS, R., CORCORAN, R., DRAKE, J., SHRYANE, N. & VOLLM, B. (2011). The QCAE: A questionnaire of cognitive and affective empathy. *Journal of Personality Assessment*, 93 (1), 84–95.
- ROGERS, K., DZIOBEK, I., HASSENSTAB, J., WOLF, O.T. & CONVIT, A. (2007). Who cares? Revisiting empathy in Asperger syndrome. *Journal of Autism and Developmental Disorders*, 37 (4), 709–715.
- SONNBY–BORGSTROM, M., JONSSON, P. & SVENSSON, O. (2003). Emotional empathy as related to mimicry reactions at different levels of information processing. *Journal of Nonverbal Behavior*, 27 (1), 3–23.
- TOMKINS, S. (1991). *Affect, imagery and consciousness*. New York: Springer Publishing Company.
- WELLMAN, H.M., CROSS, D. & WATSON, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 72, 655–684.
- WIRTZ, M., BOECKERB, M., FORKMANN, T. & NEUMANN, M. (2011). Methodology in health communication research evaluation of the “Consultation and Relational Empathy” (CARE) measure by means of Rasch-analysis at the example of cancer patients. *Patient Education and Counseling*, 82 (3), 298–306.
- YANG, K.T. & YANG, J.H. (2013). A study of the effect of a visual arts-based program on the scores of Jefferson scale for physician empathy. *BMC Medical Education*, 13, 142.
- ZAHN–WAXLER, C., ROBINSON, J.L. & EMDE, R.N. (1992). The development of empathy in twins. *Developmental Psychology*, 28, 1038–1047.
- ZILLMANN, D. (1991). Empathy: Affect from bearing witness to the emotions of others. In J. Bryant & D. Zillmann (Eds.), *Communication. Responding to the screen: Reception and reaction processes*. Hillsdale, NJ: Lawrence Erlbaum Associates.

APPENDIX I

The list of key-words that were used to look for measurements of empathy

A

Amount

Assess

Assessing

Assessment

D

Degree

Detect

Detecting

Detection

E

Evaluate

Evaluating

Evaluation

Exam

Examination

Examining

F

Function

Functioning

I

Identify

Identifying

Index

Inspect

Inspecting

Inspection

Instrument

Interview

L

Level

M

Marker

Measure

Measuring

P

Performance

Q

Questionnaire

Quotient

R

Rating

S

Scale

Score

Scoring

Self-report

Survey

T

Task

Test

Testing

Tool

APPENDIX II

The list of studies that were used to find instruments that measure empathy

- AUSTIN, E.J., EVANS, P., MAGNUS, B. & O'HANLON, K. (2007). A preliminary study of empathy, emotional intelligence and examination performance in MBChB students. *Medical Education*.
- ABRAMOWITZ, A.C., GINGER, E.J., GOLLAN, J.K. & SMITH, M.J. (2014). Empathy, depressive symptoms, and social functioning among individuals with schizophrenia. *Psychiatry Research*.
- ACHIM, A.M., OUELLET, R., ROY, M.A. & JACKSON, P.L. (2011). Assessment of empathy in first-episode psychosis and meta-analytic comparison with previous studies in schizophrenia. *Psychiatry Research*.
- ADRIAANSEN, M., VAN ACHTERBERG, T. & BORM, G. (2008). The usefulness of the staff-patient interaction response scale for palliative care nursing for measuring the empathetic capacity of nursing students. *Journal of Professional Nursing*.
- AHRWEILER, F., SCHEFFER, C., ROLING, G., GOLDBLATT, H., HAHN, E.G. & NEUMANN, M. (2014). Clinical practice and self-awareness as determinants of empathy in undergraduate education: A qualitative short survey at three medical schools in Germany. *Klinische Praxis Und Selbstwahrnehmung Als Einflussfaktoren von Empathie Im Medizinstudium: Eine Qualitative Kurzumfrage an Drei Deutschen Universitäten*.
- AINLEY, V., MAISTER, L. & TSAKIRIS, M. (2015). Heartfelt empathy? No association between interoceptive awareness, questionnaire measures of empathy, reading the mind in the eyes task or the director task. *Frontiers in Psychology*.
- ARONSON, L. (2014). Examining empathy. *The Lancet*.
- AUSTIN, E.J., EVANS, P., MAGNUS, B. & O'HANLON, K. (2007). A preliminary study of empathy, emotional intelligence and examination performance in MBChB students. *Medical Education*.
- BAILEY, P.E., HENRY, J.D. & VON HIPPEL, W. (2008). Empathy and social functioning in late adulthood. *Aging & Mental Health*.
- BALCONI, M. & BORTOLOTTI, A. (2012). Detection of the facial expression of emotion and self-report measures in empathic situations are influenced by sensorimotor circuit inhibition by low-frequency rTMS. *Brain stimulus*.
- BALCONI, M. & BORTOLOTTI, A. (2012). Empathy in cooperative versus non-cooperative situations: The contribution of self-report measures and autonomic responses. *Applied Psychophysiology and Biofeedback*.
- BALCONI, M. & BORTOLOTTI, A. (2014). Self-report, personality and autonomic system modulation in response to empathic conflictual versus non conflictual situation. *Cognitive Emotions*.
- BALCONI, M. & BORTOLOTTI, A. (2014). Self-report measures, facial feedback, and personality differences (BEES) in cooperative vs. noncooperative situations: Contribution of the mimic system to the sense of empathy. *Cognitive Emotions*.
- BALCONI, M. & CANAVESIO, Y. (2013). Emotional contagion and trait empathy in prosocial behavior in young people: The contribution of autonomic (facial feedback) and balanced emotional empathy scale (BEES) measures. *Journal of Clinical Neuropsychology*.
- BALCONI, M. & CANAVESIO, Y. (2013). High-frequency rTMS improves facial mimicry and detection responses in an empathic emotional task. *Neuroscience*.
- BALTEŞ, F.R. & MIU, A.C. (2014). Emotions during live music performance: Links with individual differences in empathy, visual imagery, and mood. *Psychomusicology: Music, Mind and Brain*.
- BARON-COHEN, S. & WHEELWRIGHT, S. (2004). The empathy quotient: An investigation of adults with asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*.
- BASS, E.C., STEDNITZ, S.J., SIMONSON, K., SHEN, T. & GAHTAN, E. (2014). Physiological stress reactivity and empathy following social exclusion: A test of the defensive emotional analgesia hypothesis. *Social Neuroscience*.
- BERG, K., BLATT, B., LOPREIATO, J., JUNG, J., SCHAEFFER, A., HEIL, D., OWENS, T., CARTER-NOLAN, P.L., BERG, D., VELOSKI, J., DARBY, E. & HOJAT, M. (2015). Standardized patient assessment of medical student empathy: Ethnicity and gender effects in a multi-institutional study. *Academical medicine*.

- BERTHOZ, S., WESSA, M., GAYANNEE, K., WICKER, B. & GRÈZES, J. (2008). Cross-cultural validation of the empathy quotient in a french-speaking sample. *Canadian Journal of Psychiatry*.
- BIKKER, A.P., FITZPATRICK, B., MURPHY, D. & MERCER, S.W. (2015). Measuring empathic, person-centred communication in primary care nurses: Validity and reliability of the Consultation and Relational Empathy (CARE) Measure. *BMC Fam Pract*.
- BIKKER, A.P., MERCER, S.W. & COTTON, P. (2012). Connecting, Assessing, Responding and Empowering (CARE): A universal approach to person-centred, empathic healthcare encounters. *Educ Prim Care*.
- BILLIMEK, J., GUZMAN, H. & ANGULO, M.A. (2015). Effectiveness and feasibility of a software tool to help patients communicate with doctors about problems they face with their medication regimen (EMPATHy): Study protocol for a randomized controlled trial. *Trials*.
- BYLUND, C.L. & MAKOUL, G. (2005). Examining empathy in medical encounters: An observational study using the empathic communication coding system. *Health Commun*.
- BOUDJEMAD, V. & GANA, K. (2009). Ageism: Adaptation of the Fraboni of Ageism Scale-Revised to the French language and testing the effects of empathy, social dominance orientation and dogmatism on ageism. *Can J Aging*.
- BRAUN, S., ROSSEEL, Y., KEMPENAERS, C., LOAS, G. & LINKOWSKI, P. (2015). Self-report of empathy: A shortened french adaptation of the Interpersonal Reactivity Index (IRI) using two large Belgian samples. *Psychol Rep*.
- BROOK, M. & KOSSON, D. (2013). Impaired cognitive empathy in criminal psychopathy: Evidence from a laboratory measure of empathic accuracy. *Journal of Abnormal Psychology*.
- BUSBY, D. & GARDNER, C. (2008). How do I analyze thee? Let me count the ways: Considering empathy in couple relationships using self and partner ratings. *Family Process*.
- CACI, H., NADALET, L., BAYLÉ, F.J., ROBERT, P. & BOYER, P. (2003). Cross-cultural study of the Impulsiveness-Venturesomeness-Empathy questionnaire (IVE-7). *Comprehensive Psychiatry Empathy theory and practice: A survey of psychologists and counselors*.
- CARLOZZI, A., LAWRENCE, B., RAY, K. & BARNES, L. (2002). Empathy theory and practice: A survey of psychologists and counselors. *The Journal of Psychology*.
- CARRÉ, A. (2013). The Basic Empathy Scale in Adults (BES-A): Factor structure of a revised form. *Psychological Assessment*.
- CHEN, D.C., PAHILAN, M.E. & ORLANDER, J.D. (2010). Comparing a self-administered measure of empathy with observed behavior among medical students. *J Gen Intern Med*.
- CHRYSIKOU, E.G. & THOMPSON, W.J. (2015). Assessing medical student empathy in a family medicine clinical test: Validity of the CARE measure. *Assessment*.
- CHRYSIKOU, E.G. & THOMPSON, W.J. (2016). Assessing Cognitive and Affective Empathy Through the Interpersonal Reactivity Index: An Argument Against a Two-Factor Model. *Assessment*.
- CHU, C.I. & TSENG, C.C. (2013). A survey of how patient-perceived empathy affects the relationship between health literacy and the understanding of information by orthopedic patients? *BMC Public Health*.
- CICCHETTI, D. & MISAKI, N. (2014). Empathy as a “risky strength”: A multilevel examination of empathy and risk for internalizing disorders. *Development and Psychopathology*.
- CLIFFORDSON, C. (2002). The hierarchical structure of empathy: dimensional organization and relations to social functioning. *Scand J Psychol*.
- COLLINS, K., GRATTON, C., HENEAGE, C. & DAGNAN, D. (2015). Employed carers’ empathy towards people with intellectual disabilities: The development of a new measure and some initial theory. *J Appl Res Intellect Disabil*.
- CORRADINI, A. & ANTONIETTI, A. (2013). Mirror neurons and their function in cognitively understood empathy. *Conscious Cogn*.
- CONRAD, B. & MCGINLEY, J. (2014). Correlational and Exploratory Factor Analyses (EFA) of commonly used empathy questionnaires: New insights. *Motivation and Emotion*.
- CRANDALL, S.J. & MARION, G.S. (2009). Commentary: Identifying attitudes towards empathy. An essential feature of professionalism. *Acad Med*.
- CROMIE, T.R. (2015). Jane Addams and the “devil baby tales”: The usefulness of perplexity in “sympathetic understanding,” a tool in learning empathy. *J Am Psychoanal Assoc*.
- CURWEN, T. (2003). The importance of offense characteristics, victimization history, hostility, and social desirability in assessing empathy of male adolescent sex offenders. *Sex Abuse*.
- CUSI, A., MacQUEEN, G.M., MCKINNON, M.C. (2010). Altered self-report of empathic responding in patients with bipolar disorder. *Psychiatry Res*.
- DAVIDSON, J.H. (2014). Empathy and integrative thinking: Talmudic paradigms for the essentials of a medical interview. *Med Humanit*.
- de BARROS, D.M., DIAS, A.M., SERAFIM, A.P., CASTELLANA, G.B., ACHÁ, M.F. & BUSATTO, G.F. (2013). Dimensional assessment of psychopathy and its relationship with physiological responses to empathic images in juvenile offenders. *Front Psychiatry*.

- DECKER, S.E., NICH, C., CARROLL, K.M. & MARTINO, S. (2014). Development of the Therapist Empathy Scale. *Behavioural and Cognitive Psychotherapy*.
- DEMIDOVA, K.Y., DUBOVIK, V., KRAVCHENKO, I. & MAKARCHOUK, N. (2014). EEG activity during viewing of neutral and emotionally colored pictures by men and women with different levels of empathy. *Neurophysiology*.
- DENCKLA, C.A., FIORI, K.L. & VINGERHOETS, A.J. (2014). Development of the Crying Proneness Scale: Associations among crying proneness, empathy, attachment, and age. *J Pers Assess*.
- DERNTL, B., FINKELMEYER, A., EICKHOFF, S., KELLERMANN, T., FALKENBERG, D.I., SCHNEIDER, F. & HABEL, U. (2010). Multidimensional assessment of empathic abilities: Neural correlates and gender differences. *Psychoneuroendocrinology*.
- DEVLIN, H.C., ZAKI, J., ONG, D.C. & GRUBER, J. (2014). Not as good as you think? Trait positive emotion is associated with increased self-reported empathy but decreased empathic performance. *PLoS One*.
- DI GIUNTA, L. & KUPFER, A. (2010). Assessing perceived empathic and social self-efficacy across countries. *European Journal of Psychological Assessment*.
- DI LILLO, M., CICCETTI, A., LO SCALZO, A., TARONI, F. & HOJAT, M. (2009). The Jefferson Scale of Physician Empathy: Preliminary psychometrics and group comparisons in Italian physicians. *Acad Med*.
- DI ROSA, E., SORARÙ, G., KLEINBUB, J.R., CALVO, V., VALLESI, A., QUERIN, G., MARCATO, S., GRASSO, I. & PALMIERI, A. (2015). Theory of mind, empathy and neuropsychological functioning in X-linked spinal and bulbar muscular atrophy: A controlled study of 20 patients. *J Neurol*.
- DODICH, A., CERAMI, C., CANESSA, N., CRESPI, C., IANNACCONE, S., MARCONE, A., REALMUTO, S., LETTIERI, G., PERANI, D. & CAPPÀ, S.F. (2015). A novel task assessing intention and emotion attribution: Italian standardization and normative data of the Story-based Empathy Task. *Neurol Sci*.
- DUAN, C. & KIVLIGHAN, D.M. (2002). Relationships among therapist pre-session mood, therapist empathy, and session evaluation. *Psychother Res*.
- DZIOBEK, I., ROGERS, K., FLECK, S., BAHNEMANN, M., HEEKEREN, H., WOLF, O. & CONVIT, A. (2008). Dissociation of cognitive and emotional empathy in adults with Asperger syndrome using the Multifaceted Empathy Test (MET). *Journal of Autism and Developmental Disorders*.
- EASTER, D.W. & BEACH, W. (2004). Competent patient care is dependent upon attending to empathic opportunities presented during interview sessions. *Curr Surg*.
- EIDE, H., FRANKEL, R., HAAVERSEN, A.C., VAUPEL, K.A., GRAUGAARD, P.K. & FINSET, A. (2004). Listening for feelings: Identifying and coding empathic and potential empathic opportunities in medical dialogues. *Patient Educ Couns*.
- EIDE, H., SIBBERN, T. & JOHANNESSEN, T. (2011). Empathic accuracy of nurses' immediate responses to fibromyalgia patients' expressions of negative emotions: an evaluation using interaction analysis. *J Adv Nurs*.
- ESLINGER, P.J., MOORE, P., ANDERSON, C. & MURRAY, G. (2011). Social cognition, executive functioning, and neuroimaging correlates of empathic deficits in frontotemporal dementia. *The Journal of Neuropsychiatry and Clinical Neurosciences*.
- EVERSON, N., LEVETT-JONES, T., LAPKIN, S., PITT, V., van der RIET, P., ROSSITER, R., JONES, D., GILLIGAN, C., COURTNEY-PRATT, H. (2015). Measuring the impact of a 3D simulation experience on nursing students' cultural empathy using a modified version of the Kiersma-Chen Empathy Scale. *J Clin Nurs*.
- FERNÁNDEZ, A. & SANTIAGO, A. (2011). Testing the psychometric properties of the Interpersonal Reactivity Index (IRI) in Chile: Empathy in a different cultural context. *European Journal of Psychological Assessment*.
- FERRARI, P.F. (2014). The neuroscience of social relations. A comparative-based approach to empathy and to the capacity of evaluating others' action value. *Behaviour*.
- FIELDS, S.K., HOJAT, M., GONNELLA, J.S., MANGIONE, S., KANE, G. & MAGEE, M. (2004). Comparisons of nurses and physicians on an operational measure of empathy. *Eval Health Prof*.
- FIELDS, S.K., MAHAN, P., TILLMAN, P., HARRIS, J., MAXWELL, K. & HOJAT, M. (2011). Measuring empathy in healthcare profession students using the Jefferson Scale of Physician Empathy: Health provider-student version. *J Interprof Care*.
- FINN, S.E. (2009). The many faces of empathy in experiential, person-centered, collaborative assessment. *Journal of Personal Assessment*.
- FJORTOFT, N., VAN WINKLE, L.J. & HOJAT, M. (2011). Measuring empathy in pharmacy students. *Am J Pharm Educ*.
- FORD, R.M. & ABERDEIN, B. (2015). Exploring social influences on the joint Simon task: Empathy and friendship. *Front Psychol*.
- FREEDBERG, S. (2007). Re-examining empathy: A relational-feminist point of view. *Social Work*.
- FROMAN, R.D. & PELOQUIN, S.M. (2001). Rethinking the use of the Hogan Empathy Scale: A critical psychometric analysis. *Am J Occup Ther*.
- GARCÍA-MORILLO, S., BERNABEU-WITTEL, M., CASSANI, M., RINCÓN, M., YERRO PÁEZ, V. & OLLERO BATURONE, M.

- (2007). Influence of biopsychosocial assessment on degree of doctor-patient empathy in a cohort of patients with multiple diseases. *Rev Clin Esp*.
- GERDES, K.E., LIETZ, C.A. & SEGAL, E.A. (2011). Measuring empathy in the 21st century: Development of an empathy index rooted in social cognitive neuroscience and social justice. *Social Work Research*.
- GILET, A.L., STUDER, J., MELLA, N., GRÜHN, D. & LABOUVIE-VIEF, G. (2013). Assessing dispositional empathy in adults: A French validation of the Interpersonal Reactivity Index (IRI). *Canadian Journal of Behavioural Science*.
- GIUMMARRA, M.J., FITZGIBBON, B.M., GEORGIOU-KARISTIANIS, N., BEUKELMAN, M., VERDEJO-GARCIA, A., BLUMBERG, Z., CHOU, M. & GIBSON, S.J. (2015). Affective, sensory and empathic sharing of another's pain: The Empathy for Pain Scale. *Eur J Pain*.
- GLASER, K.M., MARKHAM, F.W., ADLER, H.M., McMANUS, P.R. & HOJAT, M. (2007). Relationships between scores on the Jefferson Scale of physician empathy, patient perceptions of physician empathy, and humanistic approaches to patient care: A validity study. *Med Sci Monit*.
- GLEICHGERRCHT, E., TORRALVA, T., RATTAZZI, A., MARENCO, V., ROCA, M. & MANES, F. (2013). Selective impairment of cognitive empathy for moral judgment in adults with high functioning autism. *Soc Cogn Affect Neurosci*.
- GOLAN, O. & BARON-COHEN, S. (2006). Systemizing empathy: Teaching adults with Asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia. *Development and Psychopathology*.
- GORDON-GILES, N. & WAFI, Z. (2009). Assessing the beyond words educational model for empowering women, decreasing prejudice and enhancing empathy. *American Journal of Dance Therapy*.
- GOSSELIN, É., PAUL-SAVOIE, É., LAVOIE, S. & BOURGAULT, P. (2015). Validity of the French version of the Reynolds Empathy Scale among Intensive Care Nurses. *Journal of Nursing Measurement*.
- GOUVEIA, V.V., MILFONT, T.L., GOUVEIA, R.S.V., NETO, J.R. & GALVÃO, L. (2012). Brazilian-Portuguese empathy quotient: Evidences of its construct validity and reliability. *The Spanish Journal of Psychology*.
- GRADY, M.D. & ROSE, R.A. (2011). The Empathy Index: An evaluation of the psychometric properties of a new empathy measure for sex offenders. *Journal of Interpersonal Violence*.
- GRYNBERG, D. & MAURAGE, P. (2014). Pain and empathy: The effect of self-oriented feelings on the detection of painful facial expressions. *PLoS One*.
- GROEN, Y., FUERMAIER, A.B., DEN HEIJER, A.E., TUCHA, O. & ALTHAUS, M. (2015). The empathy and systemizing quotient: The psychometric properties of the Dutch version and a review of the cross-cultural stability. *J Autism Dev Disord*.
- GROVE, R., BAILLIE, A., ALLISON, C., BARON-COHEN, S. & HOEKSTRA, R.A. (2014). The latent structure of cognitive and emotional empathy in individuals with autism, first-degree relatives and typical individuals. *Mol Autism*.
- GU, J., LAU, J.T., WANG, Z., WU, A.M., TAN, X. (2015). Perceived empathy of service providers mediates the association between perceived discrimination and behavioral intention to take up HIV antibody testing again among men who have sex with men. *PLoS One*.
- HAGENMULLER, F., RÖSSLER, W., WITTWER, A. & HAKER, H. (2014). Juicy lemons for measuring basic empathic resonance. *Psychiatry Res*.
- HALL, A.E. & BRACKEN, C.C. (2011). "I really liked that movie": Testing the relationship between trait empathy, transportation, perceived realism, and movie enjoyment. *Journal of Media Psychology: Theories, Methods, and Applications*.
- HANDFORD, C., LEMON, J., GRIMM, M.C. & VOLLMER-CONNA, U. (2013). Empathy as a function of clinical exposure--reading emotion in the eyes. *PLoS One*.
- HANŽEVČAKI, M., JAKOVINA, T., BAJIĆ, Ž., TOMAC, A. & MERCER, S. (2015). Reliability and validity of the Croatian version of Consultation and Relational Empathy (CARE) Measure in primary care setting. *Croat Med J*.
- HASAN, S., AL-SHARQAWI, N., DASHTI, F., ABDULAZIZ, M., ABDULLAH, A., SHUKKUR, M., BOUHAIMED, M. & THALIB, L. (2013). Level of empathy among medical students in Kuwait University, Kuwait. *Med Princ Pract*.
- HEIDENREICH, K.S. (2001). Empathy in the physician-patient relation-tool or ethics? *Tidsskr Nor Laegeforen*.
- HENGARTNER, M.P., AJDACIC-GROSS, V., RODGERS, S., MÜLLER, M., HAKER, H. & RÖSSLER, W. (2014). Fluid intelligence and empathy in association with personality disorder trait-scores: Exploring the link. *European Archives of Psychiatry and Clinical Neuroscience*.
- HENNESSY, M., WALTER, J.S. & VESS, J. (2002). An evaluation of the Empat as a measure of victim empathy with civilly committed sexual offenders. *Sex Abuse*.
- HENRY, J.D., BAILEY, P.E. & RENDELL, P.G. (2008). Empathy, social functioning and schizotypy. *Psychiatry Res*.
- HOJAT, M., GONNELLA, J.S., MANGIONE, S., NASCA, T.J., VELOSKI, J.J., ERDMANN, J.B., CALLAHAN, C.A. & MAGEE,

- M. (2002). Empathy in medical students as related to academic performance, clinical competence and gender. *Med Educ*.
- HOJAT, M., GONNELLA, J.S., NASCA, T.J., MANGIONE, S., VELOSKI, J.J. & MAGEE, M. (2002). The Jefferson Scale of Physician Empathy: Further psychometric data and differences by gender and specialty at item level. *Acad Med*.
- HOJAT, M. & GONNELLA, J.S. (2015). Eleven years of data on the Jefferson Scale of Empathy-Medical Student version (JSE-S): Proxy norm data and tentative cutoff scores. *Med Princ Pract*.
- HOJAT, M. & LANOUE, M. (2014). Exploration and confirmation of the latent variable structure of the Jefferson scale of empathy. *Int J Med Educ*.
- HOJAT, M., MANGIONE, S., KANE, G.C. & GONNELLA, J.S. (2005). Relationships between scores of the Jefferson Scale of Physician Empathy (JSPE) and the Interpersonal Reactivity Index (IRI). *Med Teach*.
- HOJAT, M., MICHALEC, B., VELOSKI, J.J. & TYKOCINSKI, M.L. (2015). Can empathy, other personality attributes, and level of positive social influence in medical school identify potential leaders in medicine? *Acad Med*.
- HOJAT, M., MANGIONE, S., NASCA, T.J., GONNELLA, J.S. & MAGEE, M. (2005). Empathy scores in medical school and ratings of empathic behavior in residency training 3 years later. *The Journal of Social Psychology*.
- HOWLAND, M. & RAFAELI, E. (2010). Bringing everyday mind reading into everyday life: Assessing empathic accuracy with daily diary data. *Journal of Personality*.
- HSIAO, C.Y., TSAI, Y.F. & KAO, Y.C. (2013). Psychometric properties of a Chinese version of the Jefferson Scale of Empathy-Health Profession Students. *J Psychiatr Ment Health Nurs*.
- IACONO, V., ELLENBOGEN, M.A., WILSON, A.L., DESORMEAU, P. & NIJJAR, R. (2015). Inhibition of personally-relevant angry faces moderates the effect of empathy on interpersonal functioning. *PLoS One*.
- IMRAN, N., AFTAB, A.M., HAIDER, I. & FARHAT, A. (2013). Educating tomorrow's doctors: A cross sectional survey of emotional intelligence and empathy in medical students of Lahore. *Pak J Med Sci*.
- JAMIESON, M., KRUPA, T., O'RIORDAN, A., O'CONNOR, D. et al. (2006). Developing empathy as a foundation of client-centred practice: Evaluation of a university curriculum initiative. *The Canadian Journal of Occupational Therapy*.
- JANG, S.S., SEO, J.H., CHO, G.J., HONG, S.C. & WOO, H.O. (2010). Correlation of communication skills for emotional empathy and academic achievement on clinical performance examinations. *Korean J Med Educ*.
- JANISSE, T. (2006). Empathy: In a moment, a powerful therapeutic tool. *Perm J*.
- JEON, S. & CHO, E. (2015). Assessment of Korean pharmacy students' empathy using the Jefferson Scale of Empathy. *Am J Pharm Educ*.
- JEUNG, H. & HERPERTZ, S.C. (2014). Impairments of interpersonal functioning: Empathy and intimacy in borderline personality disorder. *Psychopathology*.
- JOHNSTONE, B., COHEN, D., KIRK, K.B., GLASS, B. & SHAWN, E.C. (2015). Functional and structural indices of empathy: Evidence for self-orientation as a neuropsychological foundation of empathy. *Neuropsychology*.
- JUNTUNEN, A. & JARMO, S., (2000). Library as the student's Cornerstone or obstacle: Evaluating the method of empathy-based stories. *Libri*.
- KANE, G.C., GOTTO, J.L., MANGIONE, S., WEST, S. & HOJAT, M. (2007). Jefferson scale of patient's perceptions of physician empathy: Preliminary psychometric data. *Croat Med J*.
- KARAFANTIS, D.M. (2011). The effects of ethnocultural empathy on level of stereotyping toward college athletes. *J Evid Based Soc Work*.
- KATAOKA, H.U., KOIDE, N., OCHI, K., HOJAT, M. & GONNELLA, J.S. (2009). Measurement of empathy among Japanese medical students: Psychometrics and score differences by gender and level of medical education. *Acad Med*.
- KERR, J.L., STAHNKE, A.M. & BEHNEN, E.M. (2015). Assessing empathy and self-efficacy levels of pharmacy students in an elective diabetes management course. *Am J Pharm Educ*.
- KERSTEN, P., WHITE, P.J. & TENNANT, A. (2012). The consultation and relational empathy measure: An investigation of its scaling structure. *Disabil Rehabil*.
- KHADEMALHOSSEINI, M., KHADEMALHOSSEINI, Z. & MAHMOODIAN, F. (2014). Comparison of empathy score among medical students in both basic and clinical levels. *J Adv Med Educ Prof*.
- KIERSMA, M.E., CHEN, A.M., YEHLE, K.S. & PLAKE, K.S. (2013). Validation of an empathy scale in pharmacy and nursing students. *Am J Pharm Educ*.
- KILIÇ, S.Ç., DOĞAN, E., SEVİMLİGÜL, G., YÜCEL, B., BOLAT, F., KAVAKÇI, O. & SENCAN, M. (2013). Assessing anxiety levels and empathic tendency in blood and platelet donors. *Transfus Apher Sci*.
- KILPATRICK, K.L. & ARMIDALE, K.L. (2005). The parental empathy measure: A new approach to assessing child maltreatment risk. *American Journal of Orthopsychiatry*.
- KIM, Y.T., LEE, J.J., SONG, H.J., KIM, J.H., KWON, D.H., KIM,

- M.N., YOO, D.S., LEE, H.J., KIM, H.J. & CHANG, Y. (2010). Alterations in cortical activity of male methamphetamine abusers performing an empathy task: fMRI study. *Hum Psychopharmacol.*
- KIM, J. & LEE, S.J. (2010). Reliability and validity of the Korean version of the empathy quotient scale. *Psychiatry Investig.*
- KLEIH, S.C. & KÜBLER, A. (2013). Empathy, motivation, and P300 BCI performance. *Front Hum Neurosci.*
- KOLLER, I. & LAMM, C. (2015). Item response model investigation of the (German) Interpersonal Reactivity Index empathy questionnaire: Implications for analyses of group differences. *European Journal of Psychological Assessment.*
- KONSTANTAKOPOULOS, G., OULIS, P., PLOUMPIDIS, D., PATRIKELIS, P., NIKITOPOULOU, S., PAPADIMITRIOU, G.N. & DAVID, A.S. (2014). Self-rated and performance-based empathy in schizophrenia: The impact of cognitive deficits. *Soc Neurosci.*
- KOSONOGOV, V. (2014). The psychometric properties of the Russian version of the Empathy Quotient. *Psychology in Russia.*
- KOZENÝ, J., TISANSKÁ, L. & HÖSCHL, C. (2013). Assessing empathy among Czech medical students: A cross-sectional study. *Ceskoslovenska Psychologie.*
- LAMBERTON, C.M., LEANA, C.R. & WILLIAMS, J.M. (2015). Measuring empathetic care: Development and validation of a self-report scale. *J Appl Gerontol.*
- LAMM, C., NUSBAUM, H.C., MELTZOFF, A.N. & DECETY, J. (2007). What are you feeling? Using functional magnetic resonance imaging to assess the modulation of sensory and affective responses during empathy for pain. *PLoS One.*
- LASES, S.S., ARAH, O.A., PIERIK, E.G., HEINEMAN, E. & LOMBARTS, M.J. (2014). Residents' engagement and empathy associated with their perception of faculty's teaching performance. *World J Surg.*
- LATIMER, M., JACKSON, P., JOHNSTON, C. & VINE, J. (2011). Examining nurse empathy for infant procedural pain: Testing a new video measure. *Pain Research & Management: The Journal of the Canadian Pain Society.*
- LATOUR, J.M., VAN GOUDOEVER, J.B., DUIVENVOORDEN, H.J., ALBER, S.M.J., VAN DAM, N.A., DULLAART, E., VAN HEERDE, M., DE NEEF, M., VERLAAT, C.W., VAN VUGHT, E.M. & HAZELZET, J.A. (2011). Construction and psychometric testing of the EMPATHIC questionnaire measuring parent satisfaction in the pediatric intensive care unit. *Intensive Care Med.*
- LAWRENCE, E.J., SHAW, P., BAKER, D., BARON-COHEN, S. & DAVID, A.S. (2004). Measuring empathy: Reliability and validity of the empathy quotient. *Psychological Medicine.*
- LEE, S.J., KANG DO, H., KIM, C.W., GU, B.M., PARK, J.Y., CHOI, C.H., SHIN, N.Y., LEE, J.M. & KWON, J.S. (2010). Multi-level comparison of empathy in schizophrenia: An fMRI study of a cartoon task. *Psychiatry Res.*
- LEIBERG, S. & ANDERS, S. (2006). The multiple facets of empathy: A survey of theory and evidence. *Prog Brain Res.*
- LENOFF, L. (2003). Consequences of "Empathy": Rereading Kohut's (1959) "Examination of the Relationship Between Mode of Observation and Theory". *Progress in Self Psychology.*
- LEOMBRUNI, P., DI LILLO, M., MINIOTTI, M., PICARDI, A., ALESSANDRI, G., SICA, C., ZIZZI, F., CASTELLI, L. & TORTA, R. (2014). Measurement properties and confirmatory factor analysis of the Jefferson Scale of Empathy in Italian medical students. *Perspect Med Educ.*
- LEONG, L.E., CANO, A., WURM, L.H., LUMLEY, M.A., CORLEY, A.M. (2015). A perspective-taking manipulation leads to greater empathy and less pain during the cold pressor task. *Pain.*
- LEPAGE, J.F., LORTIE, M., TASCHEREAU-DUMOUCHEL, V. & THÉORET, H. (2009). Validation of French-Canadian Versions of the Empathy Quotient and Autism Spectrum Quotient. *Canadian Journal of Behavioural Science.*
- LEWIS, M.A. (2011). Gain of function: Empathy for the uncertain patient with cancer. *J Clin Oncol.*
- LISHNER, D.A., VITACCO, M.J., HONG, P.Y., MOSLEY, J., MISKA, K. & STOCKS, E.L. (2012). Evaluating the relation between psychopathy and affective empathy: Two preliminary studies. *Int J Offender Ther Comp Criminol.*
- LYONS, C. & HAZLER, R.J. (2002). The influence of student development level on improving counselor student empathy. *Counselor Education and Supervision.*
- LYSAKER, P.H., HASSON-OHAYON, I., KRAVETZ, S., KENT, J.S. & ROE, D. (2013). Self perception of empathy in schizophrenia: Emotion recognition, insight, and symptoms predict degree of self and interviewer agreement. *Psychiatry Res.*
- LOUREIRO, J., GONÇALVES-PEREIRA, M., TRANCAS, B., CALDAS-DE-ALMEIDA, J.M. & CASTRO-CALDAS, A. (2011). Empathy in the doctor-patient relationship as viewed by first-year medical students: Data on validity and sensibility to change of the Jefferson Measure in Portugal. *Acta Med Port.*
- LOW, M. & LaSCALA, S. (2015). Medical memoir: A tool to teach empathy to nursing students. *Nurse Educ Today.*
- LÜDTKE, J., BURKHARD, M.S. & JACOBS, A.M. (2014). Immersing in the stillness of an early morning: Testing the mood empathy hypothesis of poetry reception. *Psychology of Aesthetics, Creativity, and the Arts.*
- MANGIONE, S., KANE, G.C., CARUSO, J.W., GONNELLA, J.S.,

- NASCA, T.J. & HOJAT, M. (2002). Assessment of empathy in different years of internal medicine training. *Med Teach*.
- MARKIN, R. (2011). Does my group know how I feel? Measuring intellectual empathy in groups using the social relations model. *Group Dynamics: Theory, Research, and Practice*.
- MARCI, C.D. (2008). Measuring empathy during psychotherapy. *Harvard Health Publications. The Harvard Mental Health Letter*.
- MARTINEZ, A.G., STUEWIG, J. & TANGNEY, J.P. (2014). Can perspective-taking reduce crime? Examining a pathway through empathic-concern and guilt-proneness. *Pers Soc Psychol Bull*.
- MATHERSUL, D., McDONALD, S. & RUSHBY, J.A. (2013). Understanding advanced theory of mind and empathy in high-functioning adults with autism spectrum disorder. *J Clin Exp Neuropsychol*.
- MCCORMICK, L.M., BRUMM, M.C., BEADLE, J.N., PARADISO, S., YAMADA, T. & ANDREASEN, N. (2012). Mirror neuron function, psychosis, and empathy in schizophrenia. *Psychiatry Res*.
- MCDONALD, P., ASHTON, K., BARRATT, R., DOYLE, S., IMESON, D., MEIR, A. & RISSER, G. (2015). Clinical realism: A new literary genre and a potential tool for encouraging empathy in medical students. *BMC Med Educ*.
- McKEEVER, R. (2015). Vicarious experience: Experimentally testing the effects of empathy for media characters with severe depression and the intervening role of perceived similarity. *Health Commun*.
- McLELLAN, T. & MCKINLAY, A. (2013). Sensitivity to emotion, empathy and theory of mind: Adult performance following childhood TBI. *Brain Inj*.
- MELCHERS, M., MONTAG, C., MARKETT, S. & REUTER, M. (2015). Assessment of empathy via self-report and behavioural paradigms: Data on convergent and discriminant validity. *Cogn Neuropsychiatry*.
- MERCER, S.W., MAXWELL, M., HEANEY, D. & WATT, G.C. (2004). The consultation and relational empathy (CARE) measure: Development and preliminary validation and reliability of an empathy-based consultation process measure. *Fam Pract*.
- MERCER, S.W., McCONNACHIE, A., MAXWELL, M., HEANEY, D. & WATT, G.C. (2005). Relevance and practical use of the Consultation and Relational Empathy (CARE) Measure in general practice. *Fam Pract*.
- MICHAELS, T.M., HORAN, W.P., GINGER, E.J., MARTINOVICH, Z., PINKHAM, A.E. & SMITH, M.J. (2014). Cognitive empathy contributes to poor social functioning in schizophrenia: Evidence from a new self-report measure of cognitive and affective empathy. *Psychiatry Res*.
- MICHALEC, B. (2010). An assessment of medical school stressors on preclinical students' levels of clinical empathy. *Current Psychology*.
- MOMMAERTS, J.L., GOUBERT, L. & DEVROEY, D. (2012). Empathy beyond the conceptual level: Core nonspecific factors of psychotherapy. *Perspect Biol Med*.
- MONTANARI, P., PETRUCCI, C., RUSSO, S., MURRAY, I., DIMONTE, V. & LANCIA, L. (2015). Psychometric properties of the Jefferson Scale of Empathy-Health Professional Student's version: An Italian validation study with nursing students. *Nurs Health Sci*.
- MORRIS, R., BRAMHAM, J., SMITH, E. & TCHANTURIA, K. (2014). Empathy and social functioning in anorexia nervosa before and after recovery. *Cogn Neuropsychiatry*.
- MORSANYI, K., PRIMI, C., HANDLEY, S.J., CHIESI, F. & GALLI, S. (2012). Are systemizing and autistic traits related to talent and interest in mathematics and engineering? Testing some of the central claims of the empathizing-systemizing theory. *Br J Psychol*.
- MUELLER, M.A. & WAAS, G.A. (2002). College students' perceptions of suicide: the role of empathy on attitudes, evaluation, and responsiveness. *Death Stud*.
- MUSSO, C.G. & ENZ, P.A. (2015). Art as an instrument to develop empathy. *Arch Argent Pediatr*.
- NEUMANN, M., SCHEFFER, C., LÄNGLER, A., TAUSCHEL, D., JOOS, S., LUTZ, G. & EDELHÄUSER, F. (2010). Relevance and barriers of physician empathy in daily practice - current state of research and qualitative survey of physicians. *Rehabilitation (Stuttg)*.
- NEUMANN, M., WIRTZ, M., BOLLSCHWEILER, E., WARM, M., WOLF, J. & PFAFF, H. (2008). Psychometric evaluation of the German version of the "Consultation and Relational Empathy" (CARE) measure at the example of cancer patients. *Psychotherapy Psychosomatic Medical Psychology*.
- NEW, A.S., ROT, M., RIPOLL, L.H., PEREZ-RODRIGUEZ, M., LAZARUS, S., ZIPURSKY, E., WEINSTEIN, S.R., KOENIGSBERG, H.W., HAZLETT, E.A., GOODMAN, M. & SIEVER L.J. (2012). Empathy and alexithymia in borderline personality disorder: clinical and laboratory measures. *Journal of Personality Disorders*.
- NOMURA, K. & AKAI, S. (2012). Empathy with fictional stories: Reconsideration of the fantasy scale of the interpersonal reactivity index. *Psychol Rep*.
- NORMAN, M., AMBROSE, S.A. & HUSTON, T.A. (2006). Assessing and addressing faculty morale: Cultivating consciousness, empathy, and empowerment. *Review of Higher Education*.
- NORRING, M., WIKMAN, I., HOKKANEN, A.H., KUJALA, M.V. &

- HÄNNINEN, L. (2014). Empathic veterinarians score cattle pain higher. *Vet J*.
- O'SHEA, E.C., PAVIA, S., DYER, M., CRADDOCK, G. & MURPHY, N. (2014). Measuring the design of empathetic buildings: A review of universal design evaluation methods. *Disabil Rehabil Assist Technol*.
- O'CONNOR, K., KING, R., MALONE, K.M. & GUERANDEL, A. (2014). Clinical examiners, simulated patients, and student self-assessed empathy in medical students during a psychiatry objective structured clinical examination. *Acad Psychiatry*.
- OLDERBAK, S., SASSENATH, C., KELLER, J. & WILHELM, O. (2014). An emotion-differentiated perspective on empathy with the emotion specific empathy questionnaire. *Front Psychol*
- PARK, K.H., ROH, H., SUH, D.H. & HOJAT, M. (2015). Empathy in Korean medical students: Findings from a nationwide survey. *Med Teach*.
- PARO, H.B., DAUD-GALLOTTI, R.M., TIBÉRIO, I.C., PINTO, R.M. & MARTINS, M.A. (2012). Brazilian version of the Jefferson Scale of Empathy: Psychometric properties and factor analysis. *BMC Med Educ*.
- PEDERSEN, A.F., CARLSEN, A.H. & VEDSTED, P. (2015). Association of GPs' risk attitudes, level of empathy, and burnout status with PSA testing in primary care. *Br J Gen Pract*.
- PÉLOQUIN, K. & LAFONTAINE, M.F. (2010). Measuring empathy in couples: Validity and reliability of the Interpersonal Reactivity Index for couples. *J Pers Assess*.
- PFALLER, M., VOLBERG, G. & FISCHER, P. (2012). Letter on Yamada and Decety (2009): Unconscious affective processing and empathy: an investigation of subliminal priming on the detection of painful facial expressions. *Pain*.
- PINEDA, D.A., AGUIRRE-ACEVEDO, D.C., TRUJILLO, N., VALENCIA, A.M., PAREJA, Á., TOBÓN, C., VELILLA, L. & IBÁÑEZ, A. (2013). Dimensions of empathy in ex-combatants of the Colombian armed conflict using a standardized scale. *Rev Colomb Psiquiatr*.
- PRETI, A., VELLANTE, M., BARON-COHEN, S., ZUCCA, G., PETRETTO, D.R. & MASALA, C. (2011). The Empathy Quotient: A cross-cultural comparison of the Italian version. *Cogn Neuropsychiatry*.
- PREUSCHE, I. & WAGNER-MENGHIN, M. (2013). Rising to the challenge: Cross-cultural adaptation and psychometric evaluation of the adapted German version of the Jefferson Scale of Physician Empathy for Students (JSPE-S). *Adv Health Sci Educ Theory Pract*.
- RALFS, S. & BEAIL, N. (2012). Assessing components of empathy in sex-offenders with intellectual disabilities. *J Appl Res Intellect Disabil*.
- RASCH, V., MUHAMMAD, H., URASSA, E. & BERGSTROM, S. (2000). Self-reports of induced abortion: An empathetic setting can improve the quality of data. *American Journal of Public Health*.
- RASOAL, C., JUNGERT, T., HAU, S. & ANDERSSON, G. (2011). Development of a Swedish version of the scale of ethnocultural empathy. *Psychology*.
- RENIERS, R.L., CORCORAN, R., DRAKE, R., SHRYANE, N.M. & VÖLLM, B.A. (2011). The QCAE: A questionnaire of cognitive and affective empathy. *J Pers Assess*.
- RICHTER, D. & KUNZMANN, U. (2011). Age differences in three facets of empathy: Performance-based evidence. *Psychology and Aging*.
- RYSEWYK, SV. (2009). Comment on: Unconscious affective processing and empathy: An investigation of subliminal priming on the detection of painful facial expressions. *Pain*.
- RODRIGUEZ, C.M., COOK, A.E. & JEDRZIEWSKI, C.T. (2012). Reading between the lines: Implicit assessment of the association of parental attributions and empathy with abuse risk. *Child Abuse Negl*.
- ROFF, S. (2015). Reconsidering the "decline" of medical student empathy as reported in studies using the Jefferson Scale of Physician Empathy-Student version (JSPE-S). *Med Teach*.
- ROH, M.S., HAHM, B.J., LEE, D.H. & SUH, D.H. (2010). Evaluation of empathy among Korean medical students: a cross-sectional study using the Korean version of the Jefferson Scale of Physician Empathy. *Teach Learn Med*.
- SAMSON, A.C., STANFORD, A.C. & HUBER, O.W. (2010). Short German versions of empathizing and systemizing self-assessment scales. *Swiss Journal of Psychology/Schweizerische Zeitschrift für Psychologie/Revue Suisse de Psychologie*.
- SCHNEIDER, K., REGENBOGEN, C., PAULY, K.D., GOSSSEN, A., SCHNEIDER, D.A., MEVISSSEN, L., MICHEL, T.M., GUR, R.C., HABEL, U. & SCHNEIDER, F. (2013). Evidence for gender-specific endophenotypes in high-functioning autism spectrum disorder during empathy. *Autism Res*.
- SEIDEL, E.M., HABEL, U., FINKELMEYER, A., HASMANN, A., DOBMEIER, M. & DERNTL, B. (2012). Risk or resilience? Empathic abilities in patients with bipolar disorders and their first-degree relatives. *Journal of Psychiatry Responses*.
- SHARIAT, S.V., ESHTAD, E. & ANSARI, S. (2010). Empathy and its correlates in Iranian physicians: A preliminary psychometric study of the Jefferson Scale of Physician Empathy. *Med Teach*.
- SHARIAT, S.V. & HABIBI, M. (2013). Empathy in Iranian medical students: measurement model of the Jefferson scale of empathy. *Med Teach*.

- SHASHIKUMAR, R., CHAUDHARY, R., RYALI, V.S., BHAT, P.S., SRIVASTAVA, K., PRAKASH, J. & BASANNAR, D. (2014). Cross sectional assessment of empathy among undergraduates from a medical college. *Med J Armed Forces India*.
- SHEA, S.C. & BARNEY, C. (2015). Teaching clinical interviewing skills using role-playing: Conveying empathy to performing a suicide assessment: A primer for individual role-playing and scripted group role-playing. *Psychiatr Clin North Am*.
- SIMPSON, J.A., ORIÑA, M.M. & ICKES, W. (2003). When accuracy hurts, and when it helps: A test of the empathic accuracy model in marital interactions. *Journal of Personality and Social Psychology*.
- SMITH, M.J., HORAN, W.P., COBIA, D.J., KARPOUZIAN, T.M., FOX, J.M., REILLY, J.L. & BREITER, H.C. (2014). Performance-based empathy mediates the influence of working memory on social competence in schizophrenia. *Schizophr Bull*.
- SMITH, M.J., HORAN, W.P., KARPOUZIAN, T.M., ABRAM, S.V., COBIA, D.J. & CSERNANSKY, J.G. (2012). Self-reported empathy deficits are uniquely associated with poor functioning in schizophrenia. *Schizophr Res*.
- SMITH, M.J., SCHROEDER, M.P., ABRAM, S.V., GOLDMAN, M.B., PARRISH, T.B., WANG, X., DERNTL, B., HABEL, U., DECETY, J., REILLY, J.L., CSERNANSKY, J.G. & BREITER, H.C. (2015). Alterations in brain activation during cognitive empathy are related to social functioning in schizophrenia. *Schizophr Bull*.
- SONNBY-BORGSTRÖM, M., JÖNSSON, P. & SVENSSON, O. (2003). Emotional empathy as related to mimicry reactions at different levels of information processing. *Journal of Nonverbal Behavior*.
- SPRENG, R.N., MCKINNON, M.C., MAR, R.A. & LEVINE, B. (2009). The Toronto Empathy Questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures. *J Pers Assess*.
- STANSFIELD, R.B., SCHWARTZ, A., O'BRIEN, C.L., DEKHTYAR, M., DUNHAM, L. & QUIRK, M. (2015). Development of a metacognitive effort construct of empathy during clinical training: a longitudinal study of the factor structure of the Jefferson Scale of Empathy. *Adv Health Sci Educ Theory Pract*.
- STEELE, D.J. & HULSMAN, R.L. (2008). Empathy, authenticity, assessment and simulation: A conundrum in search of a solution. *Patient Educ Couns*.
- STEINHAUSEN, S., OMMEN, O., THÜM, S., LEFERING, R., KOEHLER, T., NEUGEBAUER, E. & PFAFF, H. (2014). Physician empathy and subjective evaluation of medical treatment outcome in trauma surgery patients. *Patient Education Counselling*.
- SUCKSMITH, E., ALLISON, C., BARON-COHEN, S., CHAKRABARTI, B. & HOEKSTRA, R.A. (2013). Empathy and emotion recognition in people with autism, first-degree relatives, and controls. *Neuropsychologia*.
- SUH, D.H., HONG, J.S., LEE, D.H., GONNELLA, J.S. & HOJAT, M. (2012). The Jefferson Scale of Physician Empathy: A preliminary psychometric study and group comparisons in Korean physicians. *Med Teach*.
- SUZUKI, Y., GALLI, L., IKEDA, A., ITAKURA, S. & KITAZAKI, M. (2015). Measuring empathy for human and robot hand pain using electroencephalography. *Sci Rep*.
- TAVAKOL, S., DENNICK, R. & TAVAKOL, M. (2011). Psychometric properties and confirmatory factor analysis of the Jefferson Scale of Physician Empathy. *BMC Med Educ*.
- TEHERANI, A., HAUER, K.E. & O'SULLIVAN, P. (2008). Can simulations measure empathy? Considerations on how to assess behavioral empathy via simulations. *Patient Educ Couns*.
- TELLES, S., SINGH, N. & BALKRISHNA, A. (2015). Augmenting brain function with meditation: Can detachment coincide with empathy? *Front Syst Neurosci*.
- TENO, J.M. & CONNOR, S.R. (2009). Referring a patient and family to high-quality palliative care at the close of life: "We met a new personality... with this level of compassion and empathy". *JAMA*.
- THOMA, P. & BELLEBAUM, C. (2012). Your Error's Got me Feeling - How Empathy Relates to the Electrophysiological Correlates of Performance Monitoring. *Front Hum Neurosci*.
- THOMA, P., NORRA, C., JUCKEL, G., SUCHAN, B. & BELLEBAUM, C. (2015). Performance monitoring and empathy during active and observational learning in patients with major depression. *Biol Psychol*.
- THORSTEN, M.E. & TOPOLINSKI, S. (2015). Spatial and empathic perspective-taking correlate on a dispositional level. *Social Cognition*.
- TIERNEY, D.W. & McCABE, M.P. (2001). An evaluation of self-report measures of cognitive distortions and empathy among Australian sex offenders. *Archives of Sexual Behavior*.
- TRAN, U.S., LAIREITER, A.R., SCHMITT, D.P., NEUNER, C., LEIBETSEDER, M., SZENTE-VORACEK, S.L. & VORACEK, M. (2013). Factorial structure and convergent and discriminant validity of the E (Empathy) scale. *Psychol Rep*.
- TRUMPETER, N.N., WATSON, P.J., O'LEARY, B.J. & WEATHINGTON, B.L. (2008). Self-functioning and perceived parenting: Relations of parental empathy and love inconsistency with narcissism, depression, and self-esteem. *The Journal of Genetic Psychology*.
- TSIANTOU, D., LAZARIDOU, D., COOLIDGE, T., ARAPOSTATHIS, K.N. & KOTSANOS, N. (2013). Psychometric properties of the

- Greek version of the Toronto Composite Empathy Scale in Greek dental students. *Eur J Dent Educ*.
- ULLRICH, P., WOLLBRÜCK, D., DANKER, H. & SINGER, S. (2011). Evaluation of psycho-social training for speech therapists in oncology. Impact on general communication skills and empathy. A qualitative pilot study. *Journal of Cancer Education*.
- VACHON, D.D. & LYNAM, D.R. (2015). Fixing the problem with empathy: Development and validation of the affective and cognitive measure of empathy. *Assessment*.
- VAN VUGT, E., ASSCHER, J., HENDRIKS, J., STAMS, G.J., BIJLEVELD, C. & VAN DER LAAN, P. (2012). Assessment of moral judgment and empathy in young sex offenders: A comparison of clinical judgment and test results. *Int J Offender Ther Comp Criminol*.
- VEMURI, K. & SURAMPUDI, B.R. (2015). an exploratory investigation of functional network connectivity of empathy and default mode networks in a free-viewing task. *Brain Connect*.
- VITAGLIONE, G.D. & BARNETT, M.A. (2003). Assessing a new dimension of empathy: Empathic anger as a predictor of helping and punishing desires. *Motivation and Emotion*.
- VÖLLM, B.A., TAYLOR, A.N.W., RICHARDSON, P., CORCORAN, R., STIRLING, J., MCKIE, S., DEAKIN, J.F.W. & ELLIOTT, R. (2006). Neuronal correlates of theory of mind and empathy: A functional magnetic resonance imaging study in a nonverbal task. *NeuroImage*.
- WALOCHA, E., TOMASZEWSKA, I.M. & MIZIA, E. (2013). Empathy level differences between Polish surgeons and physicians. *Folia Med Cracov*.
- WANG, Y., YUAN, B., ROBERTS, K., WANG, Y., LIN, C. & SIMONS, R.F. (2014). How friendly is a little friendly competition? Evidence of self-interest and empathy during outcome evaluation. *Int Journal Psychophysiology*.
- WANG, Y., NEUMANN, D.L., SHUM, D.H., LIU, W.H., SHI, H.S., YAN, C., LUI, S.S., ZHANG, Q., LI, Z., CHEUNG, E.F. & CHAN, R.C. (2013). Cognitive empathy partially mediates the association between negative schizotypy traits and social functioning. *Psychiatry Res*.
- WANG, X., GU, X., FAN, J., WANG, S., ZHAO, F., HOF, P.R., LIU, P. & GAO, Z. (2014). Recovery of empathetic function following resection of insular gliomas. *Journal of Neurooncology*.
- WANG, Y.W., DAVIDSON, M.M., YAKUSHKO, O.F., SAVOY, H.B., TAN, J.A. & BLEIER, J.K. (2003). The Scale of Ethnocultural Empathy: Development, validation, and reliability. *Journal of Counseling Psychology*.
- WARD, J., SCHAAL, M., SULLIVAN, J., BOWEN, M.E., ERDMANN, J.B. & HOJAT, M. (2009). Reliability and validity of the Jefferson Scale of Empathy in undergraduate nursing students. *Journal of Nursing Measurement*.
- WEAR, D. & VARLEY, J.D. (2008). Rituals of verification: The role of simulation in developing and evaluating empathic communication. *Patient Educ Couns*.
- WEBSTER, S.D. (2002). Assessing victim empathy in sexual offenders using the victim letter task. *Sex Abuse*.
- WEN, D., MA, X., LI, H. & XIAN, B. (2013). Empathy in Chinese physicians: Preliminary psychometrics of the Jefferson Scale of Physician Empathy (JSPE). *Med Teach*.
- WHEELWRIGHT, S., BARON-COHEN, S., GOLDENFELD, N., DELANEY, J., FINE, D., SMITH, R., WEIL, L. & WAKABAYASHI, A. (2006). Predicting Autism Spectrum Quotient (AQ) from the Systemizing Quotient-Revised (SQ-R) and Empathy Quotient (EQ). *Brain Res*.
- WILLIAMS, B., BROWN, T., BOYLE, M. & DOUSEK, S. (2013). Psychometric testing of the Jefferson Scale of Empathy Health Profession Students' version with Australian paramedic students. *Nurs Health Sci*.
- WILLIAMS, J.H., CAMERON, I.M., ROSS, E., BRAADBAART, L. & WAITER, G.D. (2015). Perceiving and expressing feelings through actions in relation to individual differences in empathic traits: The Action and Feelings Questionnaire (AFQ). *Cogn Affect Behav Neurosci*.
- WILLIAMS, J.H., NICOLSON, A.T., CLEPHAN, K.J., DE GRAUW, H. & PERRETT, D.I. (2013). A novel method testing the ability to imitate composite emotional expressions reveals an association with empathy. *PLoS One*.
- WINEFIELD, H.R. & CHUR-HANSEN, A. (2000). Evaluating the outcome of communication skill teaching for entry-level medical students: Does knowledge of empathy increase? *Medical Education*.
- WIRTZ, M., BOECKER, M., FORKMANN, T. & NEUMANN, M. (2011). Evaluation of the "Consultation and Relational Empathy" (CARE) measure by means of Rasch-analysis at the example of cancer patients. *Patient Education Counselling*.
- WRIGHT, B., MCKENDREE, J., MORGAN, L., ALLGAR, V.L. & BROWN, A. (2014). Examiner and simulated patient ratings of empathy in medical student final year clinical examination: Are they useful? *BMC Med Educ*.
- WRIGHT, D.B. & SKAGERBERG, E.M. (2012). Measuring empathizing and systemizing with a large US sample. *PLoS One*.
- YAMADA, M. & DECETY, J. (2009). Unconscious affective processing and empathy: An investigation of subliminal priming on the detection of painful facial expressions. *Pain*.
- YANG, K.T. & YANG, J.H. (2013). A study of the effect of a visual

- arts-based program on the scores of Jefferson Scale for Physician Empathy. *BMC Med Educ*.
- YOUNG, S., SEDGWICK, O., PERKINS, D., LISTER, H., SOUTHGATE, K., DAS, M., KUMARI, V., BISHOPP, D. & GUDJONSSO, G.H. (2015). Measuring victim empathy among mentally disordered offenders: Validating VERA-2. *J Psychiatr Res*.
- YU, J. & KIRK, M. (2009). Evaluation of empathy measurement tools in nursing: Systematic review. *Journal Advanced Nurse*.
- ZELINKOVÁ, J., SHAW, D.J., MARECEK, R., MIKL, M., URBÁNEK, T., HAVLÍCKOVÁ, D., ZÁMECNIK, P., HAITOVÁ, P. & BRÁZDIL, M. (2014). An evaluation of traffic-awareness campaign videos: Empathy induction is associated with brain function within superior temporal sulcus. *Behavioral and Brain Functions*.

Can video games be an innovative tool to assess personality traits of the Millennial generation? An exploratory research

Elena Formica, Elena Gaiffi, Mario Magnani, Andrea Mancini, Ezio Scatolini, Manuele Ulivieri

Department of Education and Psychology, University of Florence

✎ **ABSTRACT.** Lo scopo della presente ricerca è quello di esplorare l'esistenza di una possibile relazione tra l'utilizzo dei videogiochi su dispositivi mobili (utilizzo dei videogiochi, frequenza di gioco e preferenze riportate rispetto alle diverse categorie e meccaniche di gioco) e i tratti di personalità, utilizzando il modello dei Big Five. I dati sono stati raccolti su un campione di 981 soggetti omogenei per genere e con un'età media di 23 anni; i risultati mostrano correlazioni positive e potrebbero gettare le basi per un utilizzo innovativo dei videogiochi come strumenti di selezione e valutazione delle risorse umane nelle organizzazioni.

✎ **SUMMARY.** *The main purpose of this paper is to explore the possible relationship between video games' use on mobile devices and personality traits. Play's developmental impact on learning has been long established, but little has been said about the possible different utilization of games, e.g. as a tool for skills, performance and personality traits assessment in HR and recruitment context. The research questions aimed to verify existing connections between one of the most well-known personality theory (Big Five model), video game utilization (gamers vs. non-gamers), gaming frequency (casual vs. hardcore gamers) and reported preferences to different video games categories and mechanics. Data from 981 subjects was analyzed by descriptive statistics, t-test, Effect Size and correlation analysis. Results showed that gamers differ from non-gamers on Neuroticism and its relative sub-dimension, Impulse and Emotion control; casual gamers (who play monthly or weekly) tend to prefer routine tasks, while hardcore gamers (who play every day or more than once in a day) tend to like unusual ideas, adventure and creative tasks. Players of Role Playing games seems to be more scrupulous and more open, in particular to experience, than those who do not play with games of this category. Players of Puzzle category seem to be more cooperative, friendly, scrupulous and perseverant than those who do not play to this game category, as well as logical, rational, and capable of impulse control. Simulation and strategy category share significant results in Openness to culture dimension. No statistically significant results were found for Action and Adventure categories. Correlations found between BFA dimensions and game mechanics could allow to imagine a new video games' taxonomy that transcend both academic and industrial definitions toward a nomenclature substantiated on psychological basis. This kind of redefinition could help to lay the groundwork to use video games as an assessment tool in personnel selection and evaluation.*

Keywords: Videogames, Assessment, Personality traits, Millennials generation, BFA, HR

INTRODUCTION

In recent years, due to the widespread availability of affordable video games on desktop, laptop computers and smartphones, is common experience to see children and adults, boys and girls, spending their spare time playing to the latest immersive game app, challenging themselves or their friends online. Why are videogames and mobile gaming applications so popular? How do they manage to engage people? What mechanisms underlie their success? In a psychological research perspective, given their wide dissemination, can they be used as a tool for learning and personality traits assessment?

Over the past century, a considerable amount of literature has been published on gaming and its developmental impact on learning, especially in early childhood. Traditionally, it has been argued that play is a complex and very important activity in evolutionary term, that helps preparing for adult life (Bateson & Bateson, 1987; Callois, 1981; Groos, 1898, Huizinga, 1939). Several studies have reported gaming pivotal role as a mean through which children can develop their physical, emotional, social and moral capacities; moreover, they can learn in a controlled environment, where risks related to rules infringement are minimized and where is possible to test different behaviors and problem solving (Bruner, Jolly & Sylva, 1981; Piaget, 1959; Winnicott, 1974).

Despite its great importance in growth and human evolution, play manifests itself primarily as a spontaneous activity: it acts as a perfect medium that allows the expression of children's natural curiosity and their motivation to explore the world. Researchers agree that play provides a state of mind that is uniquely suited for high-level reasoning, problem solving and creative and imaginative acting: through play, children actively make sense of the world around them, building critical basic skills for cognitive and relational achievement that includes verbalization, language comprehension, vocabulary, imagination, questioning, problem solving, observation, empathy, co-operation skills, and taking the perspective of others. According to Gray (2008), play can be defined as a confluence of several characteristics that can be narrowed down to the following five: (a) play is self-chosen and self-directed; (b) is an activity in which means are more valued than results; (c) it has a structure, or rules, which are not dictated by physical necessity but emanate from players' mind; (d) is imaginative, non-literal, in some way mentally removed from "real" or "serious" life;

(e) involves an active, alert, but non-stressed frame of mind. As highlighted by the author, play can be considered, first and foremost, an expression of freedom: players not only choose to play, but they also direct their own actions during play. Play always involves some sort of rules, but all players must freely accept them and, in case of their modification, then all players must agree to this change: that is why playing is one of the most democratic activities. Regarding players' age, Gray emphasizes that what is true for children is also true for adults' sense of play: research studies have shown that adults who have a great deal of freedom about how and when to do their work, often experience work as a game, even (in fact, especially) in case of difficult tasks; in contrast, people who work in an environment where they must do just what others tell them to do, rarely experience this kind of feeling.

However, in business organizational environment, this view is seldom supported: modern society tends to dismiss play for adults, because it is perceived as unproductive, petty or even a "guilty pleasure", and if it does, the only kind of honored play is a competitive one. The belief that seems to underlie here is that reaching adulthood only means acting serious: between personal and professional responsibilities, there seems to be no time to play. But, as suggested by Eberle (2014), adults don't lose the need for novelty and pleasure as they grow up: play continues to be interactive, satisfying, highly involving, vital for problem solving, creativity and relationship; moreover, adults do not cease to learn just because they finished their schooling time.

For these reasons, over the past decades, there has been an increasing interest in using playful tools in business and organizational contexts to achieve "continuous learning", which refers to the ability to continually develop and improve skills and knowledge to perform effectively and adapt to changes in the workplace. Experience, involvement, and attribution of meaning can be seen as key factors in continuous learning, and in the learning process in general (Kolb, 1984; Lewin, 1951). In addition, recent evidence suggest that learning is most effective when it is active, problem-based, experiential, and providing immediate feedback (Connolly, Boyle, MacArthur, Hainey & Boyle, 2012): serious games and business games seem to fully meet this need because, through simulation and direct involvement, are able to convey a message, teach a lesson, provide experience; therefore, they can be used in managerial contexts to promote forms of individual and organizational learning, training soft skills and supporting collaboration, motivation and teamwork

abilities. Interactivity, high involvement, and the possibility to have instant feedback are central aspects of videogames too: these characteristics give them great opportunities to succeed as a means of communication and learning in very different types of environments.

Nowadays, when talking about games, it is impossible to ignore video games' topic: a video game can be defined as an electronic game that involves human interaction with a user interface generating visual feedback on a video device (i.e. TV screen or computer monitor, but in the 2000s, any display device that can produce two- or three-dimensional images). The electronic systems used to play video games are known as platforms, that can range from large mainframe computers to small handheld computing devices, like smartphones; the input device used for playing, the game controller, varies across platforms and include gamepads, joysticks, mice, keyboards, touchscreens of mobile devices and buttons. Players typically view the game on a video screen and game sounds from loudspeakers are often provided; touchscreen's introduction on smartphones has allowed to include haptic, vibration-creating effects, force feedback peripherals and virtual reality headsets, which brought players to a more immersive game experience.

Due to their constant innovation and transformation in graphics quality, artificial intelligence, avatar representation, and story line, video games' classification can result as a difficult task to manage. According to Zammitto (2010), grouping games by genre do provide a quite efficient framework toward a clear classification but there have been many different approaches, both from academia and game industry, trying to reach an agreement on vocabulary and definitions that has yet not been achieved. On one hand, the academic perspective leans toward building up a common vocabulary to discuss video games, but does not offer a suitable approach for generalization without falling too short or employing too much overlapping; on the other hand, the industrial perspective offers multiple classification, but does not define which game types should be included in a specific genre, and does not seem to be completely agnostic from companies' interest.

Given this scenario, Rolling and Adams' work (2003) stands out as a thorough and systematic analysis for game genres that allows a consistent interpretation of games for classification: they identified ten game genres (action, strategy, role-playing, sports, vehicle simulation, construction and management simulation, adventure, artificial life, puzzle and

games for girls) and recognized that there are some games that fall within more than one genre. Moreover, they introduced the concept of "key elements of games" to indicate that games are composed by certain elements (equivalent to the concept of "atoms", the smallest parts that games can have) which are: 1) rules, 2) types of challenges to overcome, 3) victory conditions, 4) world settings, 5) level of abstraction-realism, 6) interaction mode, 7) player roles, 8) structures and 9) narrative. In their theorization, game's genres are clusters of a particular array of elements: this perspective made possible to describe the relation between games genres and game elements, allowing an identification of the core aspects that make games cohesive as an instrument to play.

Using this work as a starting point, Zammitto (2010) revised their genres' categorization and then created and validated a gaming preference questionnaire, with a series of implications in games' designing that are out of the purpose of this paper: in this study, Zammitto's gaming preference questionnaire was taken as a fundamental reference questionnaire to develop a game mechanics instrument for data collection, to investigate the possible relation between gaming preferences, game mechanics and personality traits.

A list of game's genres for the Italian context is shown in Table 1.

In 2010, the video game industry increased its commercial importance, with growth driven particularly by emerging Asian markets and mobile games. In 2015, video games generated sales of USD 74 billion annually worldwide: with regards to the Italian context, AESVI (the association that represents video game industry in Italy) stated that video games' market ended 2015 with a turnover of nearly EUR 1 billion and a growth trend of 6.9% compared to 2014. In 2015, there are more than 25 million video gamers in Italy (49.7% of Italian population aging more than 14), equally distributed by gender. The distribution by age groups shows a widespread of gamers up to 54 years, with a significant concentration in the 14-24 age range (19.2% of gamers, compared to 12.4% of the Italian population), in the 25-34 age range (18.1% of gamers, compared to 13.3% of the Italian population) and in 35-44 age range (24.3% of gamers, compared to 17.7% of the Italian population). Deeper analysis of socio-demographic variables delineated a gamer profile with a medium-high level of education (49.8% of gamers are holding a high school diploma or a degree, 7.7% more than the national average). Life goals declared by gamers also outline a balance between professional and family needs (achievement of success

Table 1 – Comparison between Rolling and Adams' genres, Zammitto's Gaming Preference Questionnaire genres and list of game's genres for Italian context

| Rolling and Adam's genres | | Zammitto's Gaming Preference Questionnaire genres | | List of game's genres for Italian Context |
|--------------------------------------|-------------|---------------------------------------------------|-------------------------|-------------------------------------------|
| Action | Shooting | Action | Shooting | Action |
| | No shooting | | No shooting Fighting | |
| Strategy | Turn based | Strategy | Turn based | Strategy |
| | Real Time | | Real time | |
| Role playing | | Role playing | | Role playing |
| Sports | | Sports | | (none) |
| Vehicle simulation | | Simulations | Vehicle | Simulation |
| Construction & management simulation | | | Construction | |
| Artificial Life | | | Artificial Intelligence | |
| Adventure | | Adventure | | Adventure |
| Puzzle | | Puzzle | | Puzzle |
| Games for Girls | | (none) | | (none) |

in the profession and/or in the study, obtaining a secure job and a decent salary, and/or have children) and play or entertainment needs (fun and pleasures of life, holidays and sports). The wide diffusion of devices, the development of mobile network's infrastructure, and the most assiduous use of technology by the Italian consumers, led gaming operators to increase their investment in applications and mobile sites in the last two years, contributing to develop more complex, immersive, engaging and challenging mobile games.

We usually consider video games as a mere upgrade of traditional analog games but, as explained by Johnson (2005), video games demand far more from a player than traditional games: the process of learning boundaries, goals, and control of a video game is often highly challenging, and calls on many different areas of cognitive function, as well as a great amount of patience and focus from the player. This means that, contrary to the popular perception that games provide instant gratification, video games delay gratification far longer

than other forms of entertainment: moreover, some research suggests that video games may even increase players' attention capacities, in addition to increase hand-eye coordination and visual-motor skills, sensitivity to information in the peripheral vision and the ability to count briefly presented object (Green & Bavelier, 2003). Learning principles found in video games have been identified as possible techniques with which seems possible to reform education system: Gee (2003; 2007) noticed that gamers adopt an attitude that is of such high concentration, they do not realize they are learning; it seems that, while playing video games, they "learn by doing" and this seems to also foster creative thinking (Glazer, 2006).

Findings from these studies suggest that play games and video games is closely connected with learning, but there is little published data on how games and videogames can be used as assessing tool to evaluate skills, performances and personality traits.

As stated by Zammitto (2010), most of the work on the

relationship between personality and video games has been focused on susceptibility to aggression and violence, while a considerably smaller number of studies explored personality aspects to better understand gamers and their preferences. Personality is defined as the organized totality that makes a person unique: this combination of traits, needs and motivations influences the way of behaving, thinking and approaching internal and external situations.

One of the most well-known factor theories is the Five Factor Model developed by Costa & McCrae in 1992. The model, one of the most used in work assessment context, defines personality as a combination of attitudes, motivations, interpersonal skills, emotional and experiential styles. This combination is composed of five factors: Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism. These factors are continuous variables and one's personality can be described as the likeliness that those trends will appear. Tools developed following this model, like NEO-PI-3 and NEO-FFI-3 (Costa & McCrae, 2010) and BFQ-2 (Caprara, Barbaranelli, Borgogni & Vecchione, 2007), are widely used in recruitment and assessment contexts. As the current organizational context is characterized by a frame of increasing complexity, involving continuous changes and facing the unexpected, is necessary to find new ways of working that require the development of several abilities (as symbolization, anticipation, self-regulation, cognitive flexibility, criticism, synthesis skills to distinguish important information from the continuous flow, creativity, and innovation); for this reason, is no longer enough just to analyze the job, the task, the knowledge and techniques owned by a person: personality, defined as a dimension that persists over time, is increasingly a privileged place of analysis to study, though evaluation plans, the ongoing expertise, dispositions and traits which, in a certain context, allow workers to express themselves competently.

To imagine games and videogames as an assessment tool for skills, performance and personality traits could allow researchers and HR manager to get in touch with the Millennials, which includes individuals who are the current new generation of workers or are next to enter the work market and present personality traits that seem to be fully satisfied through videogames. Millennials are a demographic cohort between Generation X and Generation Z: the name was coined by Strauss & Howe to refer to those born in a range from the early 1980s to the early 2000s. Regarding their personality traits, Millennials are represented as civic-minded, with a

strong sense of community (both local and global) (Strauss & Howe, 2000); as confident and tolerant, but also as narcissistic and with a sense of entitlement (Twenge, 2006); as "trophy kids", a term that reflects a trend in competitive sports where mere participation is frequently enough for a reward (Alsop, 2008); as constantly looking for versatility, flexibility and innovation in the workplace (Kunreuther, Kim & Rodriguez, 2008); as optimistic, engaged and team players (Furlong, 2012). Given this brief yet complex description, it should be unsurprising to acknowledge the significant and steady increase of videogame players all over the world, including Italian context: videogames characteristics of engagement, versatility, innovation, competition and instant feedback (but, as reported above, not instant gratification) seem to combine very well with all that Millennials are searching for in real life.

Zammitto (2010) suggested that people who has determined personality traits would prefer certain type of videogames: the aim of her investigation was to contribute to demographic game design by identifying gamers' personality profile to better satisfy their needs and enjoyment.

The aim of this paper is to explore the possible existing relationship between video game utilization and personality traits and to understand the possible video games utilization as an innovative, interactive, and effective assessing tool for HR and personnel selection context.

Given the lack of data about video games utilization in assessment activities, the research questions that led our efforts were three:

- Is there a relationship between one of the most well-known personality theory (Big Five factor model), video game utilization and gaming frequency?
- Is there a relationship between Big Five factor model and reported preferences to different video games genres' categories?
- Is there a relation between the Big Five factor model and game mechanics?

METHOD

Participants

981 subjects participated to this study. The average age was 23 ($SD = 6$); minimum age is 18 and maximum age is 61. Participants' gender distribution shows that 60% of subjects

were female; 58% of the participants were gamers (571 out of 981); within the group of gamers, 62% self-denominated as casual gamers, while 38% self-denominated as hardcore gamers. To better explain this division in two subgroups, is important to underline that there were no “casual” or “hardcore” definitions given to participants: they were asked about the frequency of their gaming sessions (monthly, weekly, daily, several times a day); those who answered “monthly” or “weekly” were categorized as casual (low frequency of gaming sessions), while those who answered “daily” and “several times a day” were categorized as hardcore (high frequency of gaming sessions).

Procedure

The research method used for this study was survey. Subjects recruited for the research were voluntary participants to informative and selective activities promoted within a job fair, where people can meet and get in contact with business companies, gather information about them, submit their CVs and be interviewed by HR managers to access any open positions in their business company. Each business company has its stand and participants can freely move around to interact and access to several proposed services. The organization staff of the job fair gave to the authors of this paper an independent stand to conduct the survey, to explain participants the aim of the research, to give them space and time to complete all the instruments and to ask for further information if they needed to.

As an incentive to participation, those who entered the research sample were also informed about the opportunity to receive an individualized report of the Barbaranelli, Caprara & Steca's (2002) Big Five Adjective (BFA) questionnaire, to be added to their résumés; this opportunity could be pursued by leaving their e-mail contact in the demographical questionnaire: since the main goal of attending the job fair was to be selected by featured companies and get a chance to find a job, this possibility has been very favorably welcomed by participants to the research. In fact, in this way, they had the chance to expand their self-awareness and make their CVs more complete and captivating. We are aware that giving this kind of opportunity to candidates could have influenced their participation in the study. Nevertheless, our research aims to build a shared culture with respect to the awareness of their abilities, skills and potential in the workplace. A high

level of awareness of their current capabilities and potential could enable people not only to have a higher success rate during job interviews and assessments, but also to reach those work positions that allow them to experience a high level of organizational well-being. Job fairs represent in the Italian context a real chance for personnel assessment and selection, with candidates competing for open positions in companies: for this reason, we used regulatory tables for selection and evaluation context for scoring BFA's data.

The time required to complete the entire battery of tests was about 20 minutes.

Measures

Participants were invited to complete three self-administered instruments during a single session: demographic questionnaire; a list of game mechanics, inspired by the “Gaming preference questionnaire” (Zammitto, 2010); BFA personality test (Barbaranelli et al., 2002).

Demographic questionnaire

The first questionnaire collected data on demographic variables (e-mail contact, age, gender, geographic location) and gathered information on gaming habits, such as gaming utilization (that allowed the distinction between gamers and non-gamers), gamer self-denomination (as seen above, this allowed the categorization between casual vs hardcore gamers), self-reported preference to different categories of games, and favorite game titles.

Game mechanics checklist

The second instrument consist of a list of game mechanics, designed with the aim to be as comprehensive as possible. To do so, Zammitto's (2010) “Gaming preference questionnaire” was considered, since it represents the most recent and updated thorough list of game elements which assesses whether players enjoy such game characteristics. This tool was not considered as a scientific questionnaire with defined psychometric properties, but as a checklist of game mechanics, to be analyzed individually. The final list of mechanics consists in 50 items (e.g., “I prefer

games where I can shoot”; “I prefer to control one avatar at the time”), answers ranging on a 4 point Likert scale (1 = Strongly agree, 2 = Quite agree, 3 = Quite disagree, 4 = Strongly disagree).

Big Five Adjectives Personality Test

The third part of the survey was BFA personality test (Big Five Adjectives; Barbaranelli et al., 2002). This inventory was created within the Big Five Factor theory framework, to assess people’s personality; results yield their scoring in *Neuroticism*, *Extraversion*, *Openness*, *Agreeableness* and *Conscientiousness*. For the Italian adaptation, authors defined ten sub-dimensions, two for each dimension: *Dynamism* and *Dominance* for *Extraversion* dimension; *Cooperativeness* and *Friendliness* for *Agreeableness* dimension; *Scrupulousness* and *Perseverance* for *Conscientiousness* dimension; *Emotion Control* and *Impulse Control* for *Neuroticism* dimension; *Openness to culture* and *Openness to experience* for *Openness* dimension.

In this study, an adjective-based inventory had been used for different reasons: (a) the set of possible descriptors is finished, being represented by all the adjectives contained in the vocabulary; (b) adjectives are related directly to the behavior through the lexical hypothesis; (c) adjectives are an easy and rapid assessment method (a list of adjectives can be easily completed in 10–15 minutes); (d) adjectives allow a personality assessment not anchored to a specific situation or a specific behavior, for this can be used in very differentiated situations (from a self-assessment led by the respondent to an *assessment center* finalized to obtain ratings provided by judges).

Data analysis

Descriptive, *t-test*, Effect Size and correlation analysis were conducted using SPSS statistical software (ver. 20). For each variable, skewness and kurtosis was analyzed, with values ranging from -1 to +1. Effect Size (ES) is a name given to a family of indices that measure the magnitude of the difference between two means. Unlike significance tests, these indices are independent of sample size: therefore, we have chosen to implement this kind of data analysis. In this study, we chose to use Cohen’s among the possible ES’ index and refer to Cohen’s

benchmarks to interpret resulting data (Cohen’s $d > |.20|$ small effect size; Cohen’s $d > |.50|$ medium effect size; Cohen’s $d > |.80|$ big effect size; Cohen’s $d > |1|$ huge effect size).

RESULTS

Relationship between BFA dimensions, video game utilization and gaming frequency

As the first research question aimed to investigate the relationship between BFA dimensions, video game utilization, and gaming frequency, *t*-tests were run to find significant differences between non-gamers and gamers and, within these last ones, significant differences between casual and hardcore gamers in personality traits following BFA dimensions. Results are shown in Table 2.

Comparing non-gamers (N = 571) with gamers (N = 410), data analysis show significant differences between the two subgroups for the sub-dimension of *Agreeableness*, *Friendliness* (gamers: $M = 40.6$, $SD = 13.2$; non-gamers: $M = 34.6$, $SD = 15.3$; $t = 1.08$, $p < .05$) and the dimension of *Neuroticism* (gamers: $M = 39.5$, $SD = 10.9$; non-gamers: $M = 27.7$, $SD = 7.9$; $t = 2.63$, $p < .01$), with its relative sub-dimensions of *Emotional control* (gamers: $M = 38$, $SD = 11.6$; non-gamers: $M = 29.6$, $SD = 10.2$; $t = 1.74$, $p < .05$) and *Impulse Control* (gamers: $M = 44$, $SD = 11.9$; non-gamers: $M = 32.6$, $SD = 6.9$; $t = 2.32$, $p < .05$). To figure out how gamers and non-gamers differ on Big Five factors, Effect Size analysis, using Cohen’s *d*, was run: data show a small effect size difference (Cohen’s $d = -.41$) on *Friendliness* sub-dimension; a huge effect size is reported on *Neuroticism* (Cohen’s $d = -1.23$) and its sub-dimension of *Impulse Control* (Cohen’s $d = -1.17$), while a medium effect size can be noted on *Emotion Control* sub-dimension (Cohen’s $d = -.77$). These data suggest a substantial difference between gamers and non-gamers for this factor and its relative sub-dimension.

Regarding gaming session’s frequency, casual (N = 608) and hardcore (N = 373) gamers differ for the sub-dimension of *Extraversion*, *Dynamism* (casual: $M = 38.4$, $SD = 12.4$; hardcore: $M = 35.5$, $SD = 13.1$; $t = 1.80$, $p = .05$), for the sub-dimension of *Conscientiousness*, *Perseverance* (casual: $M = 45.4$, $SD = 13.2$; hardcore: $M = 34$, $SD = 14.5$; $t = 1.76$, $p < .05$) and for the *Openness* dimension (casual: $M = 42.6$, $SD = 10.3$; hardcore: $M = 45.2$, $SD = 9.2$; $t = -2.09$,

Table 2 – Relationship between Big Five factor model, video game utilization and gaming frequency

| Big Five traits | Gamers | | | | | | | | | | | | | |
|--------------------------|-------------------------|-----------|---------------------|-----------|----------|------------|----------|---------------------|-----------|-----------------------|-----------|----------|------------|----------|
| | Non-gamers ^a | | Gamers ^b | | | | | Casual ^c | | Hardcore ^d | | | | |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>t</i> | <i>p</i> | <i>d</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>t</i> | <i>p</i> | <i>d</i> |
| Extraversion | 34.00 | 11.10 | 36.90 | 12.20 | .59 | .55 | -.25 | 37.90 | 11.80 | 35.50 | 12.60 | 1.560 | .11 | .19 |
| Dynamism | 35.00 | 12.60 | 37.20 | 12.70 | .41 | .67 | -.21 | 38.40 | 12.40 | 35.50 | 13.10 | 1.80 | .05 | .22 |
| Dominance | 37.10 | 12.60 | 38.40 | 12.10 | .25 | .80 | -.11 | 39.20 | 12.50 | 37.40 | 12.70 | 1.11 | .26 | .14 |
| Agreeableness | 39.30 | 16.60 | 40.10 | 12.60 | .15 | .87 | -.05 | 40.30 | 12.90 | 39.90 | 12.30 | .46 | .64 | .03 |
| Cooperativeness | 45.60 | 15.50 | 41.90 | 12.40 | -.72 | .46 | .26 | 42.30 | 12.80 | 41.30 | 11.80 | .63 | .52 | .08 |
| Friendliness | 34.60 | 15.30 | 40.60 | 13.20 | 1.08 | .04 | -.41 | 40.90 | 13.50 | 40.10 | 12.90 | .46 | .64 | .06 |
| Conscientiousness | 39.60 | 13.60 | 39.10 | 13.10 | -.09 | .92 | .03 | 39.50 | 12.50 | 38.60 | 13.90 | .50 | .61 | .06 |
| Scrupulousness | 45.10 | 15.30 | 43.80 | 12.40 | -.26 | .79 | .09 | 44.10 | 11.50 | 43.30 | 13.40 | .50 | .61 | .06 |
| Perseverance | 33.50 | 10.30 | 34.80 | 13.70 | .23 | .81 | -.10 | 45.40 | 13.20 | 34.00 | 14.50 | 1.76 | .04 | .82 |
| Neuroticism | 27.70 | 7.90 | 39.50 | 10.90 | 2.63 | .01 | -1.23 | 38.80 | 11.10 | 40.50 | 10.60 | .87 | .24 | -.15 |
| Emotion Control | 29.60 | 10.20 | 38.00 | 11.60 | 1.74 | .04 | -.77 | 37.50 | 11.90 | 38.70 | 11.10 | -.80 | .42 | .10 |
| Impulse Control | 32.60 | 6.90 | 44.00 | 11.90 | 2.32 | .02 | -1.17 | 42.90 | 12.00 | 45.40 | 11.50 | -1.63 | .10 | -.21 |
| Openness | 45.00 | 4.60 | 43.70 | 9.70 | -.31 | .75 | .17 | 42.60 | 10.30 | 45.20 | 9.20 | -2.09 | .03 | -.26 |
| Openness to culture | 43.10 | 9.40 | 41.30 | 12.10 | -.35 | .72 | .16 | 39.80 | 12.30 | 43.50 | 11.40 | -2.39 | .01 | -.31 |
| Openness to experience | 50.80 | 2.40 | 47.50 | 8.80 | -.90 | .36 | .51 | 46.80 | 9.20 | 48.50 | 8.10 | -1.50 | .13 | -.19 |

Note. ^a *n* = 412; ^b *n* = 569; ^c *n* = 353; ^d *n* = 216

$p < .05$) and its sub-dimension *Openness to culture* (casual: $M = 42.6$; $SD = 12.3$; hardcore: $M = 43.5$, $SD = 11.4$; $t = -2.39$, $p = .01$). Effect Size analysis shows that there is a big effect size on *Perseverance* (Cohen's $d = .82$), while a small effect size has been registered on *Dynamism* (Cohen's $d = .22$), *Openness* (Cohen's $d = -.26$) and on *Openness to culture* (Cohen's $d = -.31$). These data testify a substantial difference between casual and hardcore gamers on *Perseverance*, while the difference on *Openness* and *Openness to culture* is not so relevant as it could appear.

Relationship between BFA dimensions and reported preference to different video games' categories

T-test and Effect Size analysis were run on gamers and non-gamers of different categories of video games (adventure, action, role playing, puzzle, simulation, strategy). No statistical significant results were found between participants who reported to play Adventure and Action categories and those who reported not to play to these categories on BFA

dimensions (see Table 3 and Table 4).

Regarding Role Playing category, significant results were reported for those who play to this category on *Scrupulousness* sub-dimensions, *Openness* factor and *Openness to experience* sub-dimension (see Table 5). Effect size analysis for *Scrupulousness* sub-dimensions showed a small effect size (Cohen's $d = .40$), as well as for *Openness* (Cohen's $d = -.27$) and *Openness to experience* (Cohen's $d = -.47$).

Puzzle category was the one with more statistical significant results. Except from *Extraversion* dimension, *Emotional control* sub-dimension, *Openness* and *Openness to experience*, those who reported to play to this kind of games showed higher mean values than non-gamers (see Table 6). Effect Size analysis showed small effect size values (Cohen's d) for all statistical significant dimensions.

For *Simulation category* (see Table 7), only for *Openness to Culture* dimension results were statistically significant. Gamers showed a higher mean value ($M = 43.4$, $SD = 11.3$) than non-gamers ($M = 40.9$, $SD = 12.2$; $t = -1.41$, $p < .05$). Effect Size analysis has found a small effect size (Cohen's $d = -.21$).

Gamers who plays to *Strategy category* showed higher mean values than non-gamers on *Openness dimension* ($t = -.36$, $p < .001$) and both its relative sub-dimensions, *Openness to culture* ($t = -2.61$, $p < .001$) and *Openness to experience* ($t = -2.87$, $p < .001$) (see Table 8).

Effect Size analysis showed a small effect size for all three dimensions (Cohen's $d = -.42$ for *Openness dimension*; Cohen's $d = -.34$ for *Openness to culture*; Cohen's $d = -.37$ for *Openness to experience*).

Analyzing the correlation between BFA dimension and the number of video games' categories played by the respondent, the only small positive correlation can be retrieved between *Openness* (Pearson's $r = .24$, $p < .01$), *Openness to culture* (Pearson's $r = .19$, $p < .01$), and *Openness to experience* (Pearson's $r = .15$, $p < .05$).

Relationship between BFA dimension and game mechanics

Results of the correlational analysis (Pearson's r) between BFA factors and game mechanics are shown in Table 9. Pearson's r is a measure of the linear dependence (correlation) between two variables (in this case, BFA and game mechanics). It has a value between +1 and -1 inclusive, where +1 represent a total positive linear correlation, 0 is no

linear correlation, and -1 is total negative linear correlation. Results are shown in Table 9.

Pearson's r values are modest, ranging from $r = -.21$ to $r = .20$; nevertheless, all result showed a statistical significance ($p < .001$). For a better legibility, results will be reported referring to Big Five dimensions.

Data showed a small positive correlation between *Extraversion* dimension and the item "I enjoy controlling multiple units": the same pattern has been found on the *Dominance* sub-dimension; a small negative correlation has been found on game mechanic of dealing with big and complex world, both on *Extraversion* dimension and *Dynamism* sub-dimension. *Dominance* sub-dimension showed a small positive correlation with game mechanics relative to the presence of music and rhythm as an important part of the gameplay and where is necessary only to resolve puzzles.

Agreeableness dimension showed a small positive correlation with the item "I prefer games where my character can learn abilities" and a small negative correlation with the item "I prefer games where I have the chance of controlling several avatars at a time". *Cooperativeness* sub-dimension result positively correlated with mechanics referring to the evolution and the level growth of the game's character ("I prefer games where I can decide evolution paths for my units", "I prefer games where my character can learn abilities", and "I enjoy levelling my character"), and with mechanics related to intellectual challenges and quests. A small negative correlation is found between this sub-dimension and mechanics related to control of multiple units (e.g., "I prefer games that I have the chance of controlling several avatars at a time"). Regarding *Friendliness* sub-dimension, small positive correlations has been found on items of receiving hints for play optimization and of preferring intellectual challenges; a small negative correlation is retrieved on item "I prefer games where I have to mainly kick and punch enemies".

About *Conscientiousness* dimension, small positive correlations are found on mechanics related to building and pulling on structures, resolving puzzles and dealing with challenges that require eye-hand coordination; small negative correlations with this dimension are found with items that underlines the importance of gun's usage in the gameplay, the wander without clear finality and objectives, and playing online (with or without others). Same correlation patterns are retrieved on the *Scrupulousness* sub-dimension, except for item "I enjoy fooling around the game world without any main reason or objective"; in

Table 3 – Relationship between BFA dimensions and reported preference to Adventure video game category

| Big Five traits | Adventure | | | | <i>t</i> | <i>p</i> | <i>d</i> |
|--------------------------|-------------------------|-----------|---------------------|-----------|----------|----------|----------|
| | Non-gamers ^a | | Gamers ^b | | | | |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | | | |
| Extraversion | 36.60 | 12.10 | 37.90 | 12.40 | -.74 | .46 | -.10 |
| Dynamism | 36.90 | 12.90 | 37.90 | 12.40 | -.54 | .59 | -.07 |
| Dominance | 38.30 | 12.20 | 39.10 | 14.00 | -.42 | .67 | -.06 |
| Agreeableness | 39.90 | 12.10 | 40.60 | 14.80 | -.36 | .72 | -.05 |
| Cooperativeness | 42.10 | 11.90 | 41.60 | 14.50 | .25 | .80 | .03 |
| Friendliness | 40.30 | 12.70 | 41.10 | 15.30 | -.41 | .68 | -.05 |
| Conscientiousness | 39.00 | 12.10 | 39.50 | 16.00 | -.28 | .78 | -.03 |
| Scrupulousness | 44.10 | 11.20 | 42.90 | 15.80 | .67 | .50 | .08 |
| Perseverance | 34.30 | 13.10 | 36.60 | 15.60 | -1.17 | .24 | -.16 |
| Neuroticism | 38.90 | 10.40 | 40.70 | 13.00 | -1.16 | .25 | -.15 |
| Emotional control | 37.50 | 11.30 | 39.20 | 12.80 | -.99 | .32 | -.14 |
| Impulse control | 43.60 | 11.30 | 44.20 | 13.80 | -.35 | .73 | -.05 |
| Openness | 43.60 | 8.30 | 44.50 | 13.40 | -6.78 | .50 | -.08 |
| Openness to culture | 41.50 | 11.00 | 41.10 | 15.00 | .27 | .79 | .03 |
| Openness to experience | 47.40 | 7.40 | 48.40 | 12.30 | -.74 | .46 | -.09 |

Note. ^a *n* = 755; ^b *n* = 266

addition, a small positive correlation is found between this sub-dimension and the presence of intellectual challenges' mechanic, while a small negative correlation is established with moving around own avatar fast in the gameplay. *Perseverance* sub-dimension showed small positive correlations with getting high scores and gameplays that have a story that unfolds while playing; small negative correlations are found between this sub-dimension and guns' using, engaging only sometimes with characters stronger than the average, and fooling around the game world without any main reason or objectives.

Small positive correlations are found between *Neuroticism*

dimension and items "I prefer games where my character can learn abilities", "I prefer games that are an intellectual challenge", and "I enjoy resolving puzzles for their own sake": the same patterns are showed on *Impulse control* sub-dimension; in addition, this sub-dimension has a small positive correlation with managing resources' mechanic. A small negative correlation is found for *Neuroticism* dimension, and its relative sub-dimensions of *Emotion* and *Impulse control*, and weapon using mechanic.

Finally, *Openness* dimension only showed a small positive correlation with item "I prefer games with intelligent life"; unexpectedly, this item do not correlate with *Openness* sub-

Table 4 – Relationship between BFA dimensions and reported preference to Action video game category

| Big Five traits | Action | | | | <i>t</i> | <i>p</i> | <i>d</i> |
|--------------------------|-------------------------|-----------|---------------------|-----------|----------|----------|----------|
| | Non-gamers ^a | | Gamers ^b | | | | |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | | | |
| Extraversion | 36.00 | 12.10 | 38.20 | 12.20 | −1.44 | .15 | −.18 |
| Dynamism | 36.40 | 12.70 | 38.30 | 12.80 | −1.16 | .25 | −.15 |
| Dominance | 37.70 | 12.40 | 39.50 | 12.90 | −1.12 | .26 | −.14 |
| Agreeableness | 40.10 | 12.50 | 40.10 | 13.10 | .03 | .97 | .00 |
| Cooperativeness | 42.20 | 11.80 | 41.60 | 13.60 | .36 | .72 | .05 |
| Friendliness | 40.60 | 13.50 | 40.30 | 13.10 | .20 | .84 | .02 |
| Conscientiousness | 39.80 | 12.30 | 38.20 | 14.20 | .93 | .35 | .12 |
| Scrupulousness | 44.90 | 11.70 | 42.10 | 13.40 | 1.8 | .08 | .02 |
| Perseverance | 34.70 | 12.80 | 35.00 | 15.10 | −.20 | .84 | −.02 |
| Neuroticism | 39.70 | 10.90 | 38.70 | 11.40 | .72 | .47 | .09 |
| Emotional control | 38.10 | 11.60 | 37.50 | 11.80 | .42 | .68 | .05 |
| Impulse control | 44.40 | 11.30 | 42.80 | 12.90 | 1.01 | .31 | .13 |
| Openness | 43.20 | 8.90 | 44.70 | 10.80 | −1.22 | .22 | −.15 |
| Openness to culture | 40.90 | 11.80 | 42.20 | 12.30 | −.79 | .43 | −.11 |
| Openness to experience | 47.20 | 7.60 | 48.20 | 10.30 | −.89 | .37 | −.11 |

Note. ^a *n* = 592; ^b *n* = 389

dimensions. *Openness to culture* sub-dimension showed small (only) positive correlations with “making building and structure”, “resolving puzzles for own sake”, being challenged with eye-hand coordination tasks, and “dealing with a story that unfolds while playing” mechanics. Regarding *Openness to experience* sub-dimension, small (only) positive correlations are found between this sub-dimension and items “I prefer games that I can decide evolution paths for my units”, “I enjoy that rarely I have to engage with a character stronger than the average”, and “I prefer games where my character’s stats have a key role in hitting and resisting while fighting”.

DISCUSSION

As for the first research question, gamers and non-gamers had been compared on the BFA dimensions. Gamers show a higher mean scores on *Friendliness* sub-dimension, compared with non-gamers: this data can be interpret as that gamers are more friendly than non-gamers; yet, different sample size could have affected this kind of result; for this reason, running Effect Size analysis has been an important step to interpret correctly this difference: since there is a small, yet significant, Effect Size difference, the gap between

Table 5 – Relationship between BFA dimensions and reported preference to Role Playing video game category

| Big Five traits | Role Playing | | | | <i>t</i> | <i>p</i> | <i>d</i> |
|--------------------------|-------------------------|-----------|---------------------|-----------|----------|-----------------|----------|
| | Non-gamers ^a | | Gamers ^b | | | | |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | | | |
| Extraversion | 37.00 | 12.20 | 36.30 | 12.10 | .31 | .75 | .06 |
| Dynamism | 37.50 | 12.80 | 34.70 | 12.20 | 1.19 | .23 | .22 |
| Dominance | 38.20 | 12.50 | 39.80 | 13.20 | -.69 | .49 | -.12 |
| Agreeableness | 40.30 | 12.50 | 39.10 | 14.70 | .50 | .64 | .09 |
| Cooperativeness | 42.10 | 12.20 | 41.40 | 14.40 | .29 | .77 | .05 |
| Friendliness | 40.80 | 13.00 | 38.60 | 15.30 | .87 | .38 | .15 |
| Conscientiousness | 39.50 | 12.70 | 37.00 | 15.30 | 1.03 | .30 | .18 |
| Scrupulousness | 44.50 | 12.10 | 39.30 | 13.90 | 2.27 | .02 | .40 |
| Perseverance | 34.60 | 13.40 | 36.00 | 15.50 | -.55 | .58 | -.09 |
| Neuroticism | 39.50 | 10.70 | 38.00 | 13.50 | .75 | .45 | .12 |
| Emotional control | 37.90 | 11.50 | 37.80 | 13.00 | .02 | .98 | .01 |
| Impulse control | 44.10 | 11.50 | 41.30 | 14.60 | 1.31 | .19 | .21 |
| Openness | 42.60 | 10.30 | 45.20 | 9.20 | -2.52 | .01 | -.27 |
| Openness to culture | 41.20 | 11.70 | 43.20 | 14.10 | -.92 | .36 | -.15 |
| Openness to experience | 47.00 | 8.10 | 51.70 | 11.50 | -2.97 | <.001 | -.47 |

Note. ^a *n* = 853; ^b *n* = 128

gamers and non-gamers is not so deep as it could appear from mean scores comparison, suggesting that gamers and non-gamers do not differ substantially in the *Friendliness* dimension.

Furthermore, gamers showed higher mean scores on the *Neuroticism* dimension, and the same pattern can be noted on the sub-dimension of *Emotion* and **Impulse control**: in this case, Effect Size analysis (medium for *Emotion control*, and huge for *Neuroticism* and *Impulse control*) allows to say that gamers seem to be more capable of emotion and impulse control. This can be probably due to the training that playing

video games provides: as stated by Zillmann & Bryant (1994), people use video games to better manage emotional states, and this can happen in a more or less conscious way. For example, through video games people can manage their emotions, searching for relax (e.g. playing to “disconnect” from a hard day by soaking up into game play and relieving stress) or, on the contrary, looking for specific emotions (e.g., adrenaline in a survival horror game).

Video games structure seems to be suited to ensure that people get into a psychological flow state (Csikszentmihalyi & Csikszentmihalyi, 1998; Massimini & Carli, 1998) that is

Table 6 – Relationship between BFA dimensions and reported preference to Puzzle video game category

| Big Five traits | Puzzle | | | | <i>t</i> | <i>p</i> | <i>d</i> |
|--------------------------|-------------------------|-----------|---------------------|-----------|----------|----------|----------|
| | Non-gamers ^a | | Gamers ^b | | | | |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | | | |
| Extraversion | 36.80 | 12.70 | 37.00 | 11.40 | -.12 | .91 | -.02 |
| Dynamism | 37.10 | 13.00 | 37.20 | 12.50 | -.11 | .91 | -.01 |
| Dominance | 38.40 | 13.20 | 38.50 | 11.80 | -.02 | .98 | -.01 |
| Agreeableness | 38.00 | 13.30 | 43.20 | 11.20 | -3.30 | <.001 | -.42 |
| Cooperativeness | 40.10 | 12.80 | 44.70 | 11.60 | -3.0 | <.001 | -.38 |
| Friendliness | 38.10 | 14.10 | 43.90 | 11.30 | -3.49 | <.001 | -.45 |
| Conscientiousness | 37.10 | 12.70 | 42.10 | 13.20 | -3.05 | <.001 | -.39 |
| Scrupulousness | 41.70 | 12.30 | 47.00 | 12.00 | -3.45 | <.001 | -.44 |
| Perseverance | 33.60 | 13.20 | 36.60 | 14.30 | -1.78 | .05 | -.22 |
| Neuroticism | 38.20 | 11.50 | 40.90 | 10.30 | -1.96 | .05 | -.25 |
| Emotional control | 37.50 | 11.90 | 38.40 | 11.40 | -.60 | .55 | -.08 |
| Impulse control | 41.90 | 12.60 | 46.50 | 10.30 | -3.16 | <.001 | -.40 |
| Openness | 43.30 | 9.90 | 44.50 | 9.40 | -.94 | .35 | -.12 |
| Openness to culture | 40.10 | 11.70 | 43.40 | 12.20 | -2.22 | .03 | -.31 |
| Openness to experience | 48.20 | 9.00 | 46.80 | 8.30 | 1.30 | .19 | -.19 |

Note. ^a *n* = 581; ^b *n* = 400

characterized by a) an intense concentration on what is being done; b) the merging of action and awareness; c) the loss of self-perception as social actors; d) the feeling of being able to effectively handle the situation because the necessary skills to face the challenge presented by the context (in this case, video game) have been developed; e) the feeling that time passes fastest than normal; f) the feeling that the activity that is taking place is satisfactory.

Psychological flow, however, is not always a present condition during gaming sessions; it is, indeed, an emotional state reached only when there is a good balance between

player's skills and game's difficulty. If the game is too simple, for example, the player could be bored, while if it is too complex, it could evocate anxiety and frustration. For this reason, game's difficulty will increase progressively, so that the player will develop new skills and strengthen those that he/she already possess. When a high skills level is reached in adventure or action games, a player can no longer experience strong emotions yet, in a paradoxical way, can get to relax (Keller & Bless, 2008).

Comparing casual and hardcore gamers on Big Five dimensions, data showed that, on *Dynamism* sub-

Table 7 – Relationship between BFA dimensions and reported preference to Simulation video game category

| Big Five traits | Simulation | | | | <i>t</i> | <i>p</i> | <i>d</i> |
|--------------------------|-------------------------|-----------|---------------------|-----------|----------|------------|----------|
| | Non-gamers ^a | | Gamers ^b | | | | |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | | | |
| Extraversion | 36.30 | 12.00 | 38.80 | 12.80 | −1.37 | .17 | −.16 |
| Dynamism | 36.80 | 12.70 | 38.40 | 13.10 | −.85 | .40 | −.12 |
| Dominance | 37.90 | 12.40 | 40.30 | 13.10 | −1.27 | .21 | −.19 |
| Agreeableness | 40.20 | 13.10 | 39.70 | 11.40 | .29 | .77 | .04 |
| Cooperativeness | 42.20 | 12.70 | 41.20 | 11.60 | .53 | .59 | .08 |
| Friendliness | 40.80 | 13.60 | 39.30 | 12.30 | .72 | .47 | .11 |
| Conscientiousness | 39.20 | 13.10 | 39.20 | 13.20 | −.00 | .99 | .00 |
| Scrupulousness | 43.80 | 12.50 | 44.00 | 12.10 | −.10 | .92 | −.02 |
| Perseverance | 35.10 | 13.70 | 34.00 | 13.70 | .52 | .60 | .08 |
| Neuroticism | 39.20 | 11.60 | 39.50 | 9.10 | −.13 | .90 | −.03 |
| Emotional control | 37.60 | 11.90 | 38.50 | 10.90 | −.74 | .46 | −.08 |
| Impulse control | 44.10 | 12.30 | 42.70 | 10.50 | .80 | .42 | .12 |
| Openness | 43.30 | 9.60 | 45.40 | 9.90 | −1.42 | .16 | −.21 |
| Openness to culture | 40.90 | 12.20 | 43.40 | 11.30 | −1.41 | .04 | −.21 |
| Openness to experience | 47.40 | 8.90 | 48.40 | 8.40 | −.79 | .43 | −.19 |

Note. ^a *n* = 758; ^b *n* = 223

dimension, casual gamers report a higher mean score than hardcore gamers; the same pattern is verifiable on the *Conscientiousness* sub-dimension, *Perseverance*. From these results, it appears that those who self-refer to play video games monthly or weekly describe themselves as more dynamic, brisk and active, as well as responsible, liable, self-disciplined, and striving for achievements in activities in which they are involved.

Running Effect Size analysis, Cohen's *d* showed different values for these two sub-dimensions: for *Dynamism* sub-dimension, a small Effect Size is retrieved from analysis, while

a big Effect Size is reported for *Perseverance*; these results suggest that those who self-refer to play video games daily or several times a day do not substantially differ on the dynamic sub-factor from those who report to play monthly or weekly, while casual gamers seem to show more persistent attitudes and behaviors than hardcore gamers. In other words, less time passed playing video games has not a substantial impact on energy or active behavior, but could tell something about showed preferences for planned behavior, being prepared and paying attention for details, liking order and schedules, and acting dutifully.

Table 8 – Relationship between BFA dimensions and reported preference to Strategy video game category

| Big Five traits | Strategy | | | | <i>t</i> | <i>p</i> | <i>d</i> |
|--------------------------|-------------------------|-----------|---------------------|-----------|----------|-----------------|----------|
| | Non-gamers ^a | | Gamers ^b | | | | |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | | | |
| Extraversion | 37.10 | 12.00 | 36.50 | 12.40 | .36 | .72 | .05 |
| Dynamism | 37.60 | 12.90 | 36.40 | 12.60 | .68 | .49 | .09 |
| Dominance | 38.00 | 12.80 | 39.20 | 12.40 | -.71 | .47 | -.09 |
| Agreeableness | 39.70 | 13.00 | 40.80 | 12.40 | v-.65 | .52 | -.09 |
| Cooperativeness | 41.00 | 13.00 | 42.10 | 11.70 | -.14 | .89 | -.09 |
| Friendliness | 39.90 | 13.30 | 41.50 | 13.40 | -.94 | .35 | -.12 |
| Conscientiousness | 38.60 | 13.10 | 40.00 | 13.20 | -.82 | .41 | -.11 |
| Scrupulousness | 43.60 | 12.70 | 44.20 | 12.10 | -.36 | .72 | -.05 |
| Perseverance | 33.90 | 13.30 | 36.40 | 14.30 | -1.41 | .16 | -.18 |
| Neuroticism | 38.70 | 11.10 | 40.30 | 10.90 | -1.17 | .24 | -.14 |
| Emotional control | 37.70 | 12.10 | 38.20 | 11.00 | -.34 | .74 | -.04 |
| Impulse control | 42.90 | 12.10 | 45.20 | 11.50 | -1.54 | .12 | -.19 |
| Openness | 42.30 | 9.50 | 46.30 | 9.50 | -3.36 | <.001 | -.42 |
| Openness to culture | 39.90 | 12.20 | 43.90 | 11.40 | -2.61 | <.001 | -.34 |
| Openness to experience | 46.40 | 8.90 | 49.60 | 8.20 | -2.87 | <.001 | -.37 |

Note. ^a *n* = 615; ^b *n* = 366

Regarding the *Openness* dimension, and its sub-dimension *Openness to Culture*, hardcore gamers showed significant higher mean scores, compared with casual gamers. For both, a small Effect Size is reported, meaning that the difference found in results is not indicative of a profound gap between the two sub-groups: a high frequency of gaming sessions seems to have little impact on creativity, intellectual curiosity, preference for novelty and variety, and on the extent to which a person is imaginative or independent. Therefore, is not possible to state that hardcore gamers are certainly more creative and

intellectually curious than casual gamers; yet, a general tendency in this direction was recorded. This tendency seems supportive and consistent with previous results on *Persistence* sub-dimension: casual gamers tend to prefer routine tasks, while hardcore gamers tend to like unusual ideas, adventure and creative tasks.

Although interpreting these results in a clear and straightforward way is not always easy, it is interesting to note that a relationship between these variables exists indeed; therefore, knowing preferences and playing habits seems to be a useful and innovative way to get to know important

Table 9 – Correlation (Pearson's r) between Big Five factor model and game mechanics

| Games mechanics | Extroversion | Dynamism | Dominance | Agreeableness | Cooperativeness | Friendliness | Conscientiousness | Scrupulousness | Perseverance | Neuroticism | Emotion control | Impulse control | Openness | Openness to culture | Openness to experience |
|-----------------|--------------|----------|-----------|---------------|-----------------|--------------|-------------------|----------------|--------------|-------------|-----------------|-----------------|----------|---------------------|------------------------|
| Item 5 | | | | | | | | -.15* | | | | | | | |
| Item 7 | | | | | | | -.19* | -.21* | -.12* | -.18* | -.15* | -.16* | | | |
| Item 9 | | | | | | .14* | | | | | | | | | |
| Item 10 | | | | | | | | .15* | | | | | | | |
| Item 13 | -.14* | -.15* | | | | | | | | | | | | | |
| Item 14 | | | .13* | | | | | | | | | | | | |
| Item 16 | | | | | | -.14* | | | | | | | | | |
| Item 18 | | | | | | | -.16* | -.18* | | | | | | | |
| Item 19 | | | | | .15* | | | | | | | | | | .13* |
| Item 20 | | | | | | | -.16* | -.17* | | | | | | | |
| Item 22 | | | | .13* | .14* | | | | | .13* | .14* | | | | |
| Item 23 | | | | | .12* | .13* | | .14* | | .15* | .20* | | | | |
| Item 25 | | | | | | | .14* | .14* | | | | | | .14* | |
| Item 29 | | | | | | | .14* | .14* | | .14* | .16* | | | .16* | |
| Item 31 | | | | | | | | | -.14* | | | | | | .13* |
| Item 32 | | | | | | | | | | | | | .14* | | |
| Item 35 | | | | | | | -.13* | -.18* | | | | | | | |
| Item 36 | | | | | .12* | | | | | | | | | | |
| Item 37 | | | | | | | | | | | .12* | | | | |
| Item 39 | | | | -.12* | -.14* | | | | | | | | | | |
| Item 40 | | | | | | | .14* | .13* | | | | | | .20* | |
| Item 41 | | | | | | | | | .14* | | | | | | |
| Item 42 | | | | | | | .16* | .15* | | | | | | .15* | |
| Item 43 | | | | | | | | | | | | | | | .14* |
| Item 44 | .13* | | .18* | | -.13* | | | | | | | | | | |
| Item 47 | | | | | .14* | | | | | | | | | | |
| Item 50 | | | .13* | | | | | | | | | | | | |

Note. * $p < .001$

aspects (such as personality traits) of Millennials generation, serial players who are next to enter the work market.

The second research question was about relationships between Big Five factor model and reported preferences to games' categories. Regarding scores interpretation process, Barbaranelli et al. (2002) suggest that it can be referred both to the individual scales and to the overall profile resulting from all dimensions. In fact, ratings in a single scale assume significance especially in relation to scores reported in other scales: this observation invites to maintain a cautious attitude in data interpretation. While no statistical significant results were found for Action and Adventure categories, data showed that those who play to Role Playing games seems to be more scrupulous, more open (and in particular open to experience) than non-gamers of this category. These findings may be due to the fact that, in this kind of games, individuals play the role of one or more characters and, through conversation and dialectic exchange, create an imaginary space where fictitious, adventurous facts happen in a neat narrative; each character is defined by a variety of features (e.g., strength, dexterity, intelligence, charisma and so on), generally testified through scores that describe their capabilities; actions taken in the game succeed or fail according to a formal system of rules or guidelines. For these characteristics, it is no surprising that gamers that reported preference for role playing games need to implement scrupulous behaviors, as well as to be open to new experience, to be successful; following specific rules and being open to sudden changes of the gameplay scenario are key skills in managing this kind of games.

Puzzle category players seem to be more cooperative and friendly, scrupulous and perseverant than those who do not play to this category, as well as capable of impulse control and logical and rational. Puzzle games (such as Candy Crush Saga) can be defined as a non-competitive and cognitive games, in which the player should think about his/her moves in advance and can help other players by giving lives or accesses to next levels. To be precise, methodic, systematic, determined, attentive, tenacious and efficient represent a set of skills necessary to be successful in this category, as well as generous, helpful, patient and analytical.

Simulation and Strategy categories share significant results in *Openness to culture* dimension. Simulation games try to reproduce an actual aspect of reality and put the player in a position that demand to act like he/she was actually in the situation presented: generally, it require a mix of skill, luck and strategy, and for this reason can be defined as strategy games'

subcategory; both categories are influenced by the capability of the player to make effective decisions and require to be receptive to changes, analytical, logic and rational, as well as intuitive, because strategic moves are generally contrasted by fate and fortune. Moreover, those who play Strategy games seems to be more open mentally, and, in particular, open to experience: this means that, to play strategy games, there is a need of being imaginative, original, creative and unpredictable that seems not to be as fundamental as for simulation games: maybe, these skills are not that useful in strictly reproducing various aspects of real or fantasy life, while seem important to individuate new path to solve incoming problems and situations.

Regarding findings of positive correlation between the number of games' category played by participants and BFA *Openness* dimension and its relative sub-dimension, these results seems consistent with openness to different values and lifestyles: trying different games' categories, an individual can get involved in different situation and test different skills, achieving a global skills training. This could be an important aspect in candidate's personality evaluation, that could tell something about his/her behavior not only in games' world, but also in the real work environment.

The third research question aimed to explore existing correlation between BFA dimensions and game mechanics. Correlations found between BFA dimensions and game mechanics could allow to imagine a new video games' taxonomy that transcend both academic and industrial definitions toward a nomenclature substantiated on psychological basis. This kind of redefinition could help to lay the groundwork for using video games as an assessment tool in personnel selection and evaluation: in a future perspective, this could allow business companies and HR managers to use video games as suitable instrument for selecting the right candidate that applies for a specific job profile or open position. For instance, if a business company is looking for a candidate with a personality profile that demands particularly high levels of conscientiousness, it may use a video game that asks the player to solve puzzles or tasks that require a high degree of eye-hand coordination effort.

Given the knowledge and ease of use of technological devices that characterizes the Millennials' generation, this kind of assessment, on one hand, might be more engaging and less anxiety-provoking for candidates, allowing them to fully express their true potential; on the other hand, it might give business companies an effective and realistic assessment of the candidate who will cover the open position.

Conclusion, limitation and inputs for future research

The intention of this discussion is not to support inconsistent theories like that whoever plays is necessarily better than those who do not play, but to underline that inferring personality traits from video gaming behavioral habits and video games' preference could be possible. Using video games as an assessment tool is a new field of research in Italian context, that surely needs several other investigations with different instruments and type of measurement: one of the aims of this paper was to support already existent theories and intuitions with empirical data, hoping to awaken interest in this subject and add potentially relevant information for further research.

Each scientific study has its own strengths and its weaknesses; for example, we have previously discussed the hypothetical beneficial impact that giving the BFA profile might have had on research participants. We do not know if this element has been crucial to participation or whether participants have been moved by interest in the research topic. We hypothesized that the participants, mainly belonging to the Millennials generation, were moved by a strong interest and curiosity towards the research topic (that is the connection between job search, personality and video

games): in any case, future research can better clarify this point by eliminating this kind of incentive or replicating the study in a different setting.

In our perspective, this research work arises in scientific literature as an innovative contribution that tries to take into account recent macroscopic changes both in technology, such as the widespread and pervasive use of technological devices, both in socio-demographic aspects, such as the presence in the work market of a new generation, Millennials, that has its own personality characteristics and peculiar ways to see and interact with external reality.

As an input for future exploration and improvements, research could investigate some other aspects that have remained outside of the present study, as for instance the possible relationship between games' scores and achievement. In fact, if gamers' reported preferences of different categories of games (and underlying elements that give structure to the game, the so-called "game mechanics") can be considered as an effective path to collect information on their personality traits, games' scores and achievements could be considered as a mean to collect information on gamers' performance: this could be a way to measure them more precisely and to possibly correlate players' achievements and results to performance tests (such as, for instance, DAT and Raven Matrices), avoiding the exclusive use of self-report instruments.

References

- AESVI'S PRESS RELEASE (2016, April 5). Retrieved February 27, 2017, from http://www.aesvi.it/cms/view.php?cms_pk=2617&dir_pk=902
- ALSOB, R. (2008). *The trophy kids grow up: How the millennial generation is shaking up the workplace*. Hoboken, NJ: John Wiley & Sons.
- BARBARANELLI, C., CAPRARA, G.V. & STECA, P. (2002). *BFA: Big Five Adjective. Manuale*. Firenze: Giunti OS Organizzazioni Speciali.
- BATESON, G. & BATESON, M.C. (1987). Innocence and Experience. *Angels Fear: Towards an Epistemology of the Sacred*, 167–182.
- BRUNER, J.S., JOLLY, A. & SYLVA, K. (1981). *Il gioco*. Roma: Armando.
- CAILLOIS, R. (1981). *I giochi e gli uomini. La maschera e la vertigine*. Milano: Bompiani.
- CAPRARA, G.V., BARBARANELLI, C., BORGOGNI, L. & VECCHIONE, M. (2007). *BFQ-2: Manuale*. Firenze: Giunti OS Organizzazioni Speciali.
- CONNOLLY, T.M., BOYLE, E.A., MacARTHUR, E., HAINEY, T. & BOYLE, J.M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, 59 (2), 661–686.
- COSTA Jr, P.T. & McCRAE, R. (2010). *NEO™ Inventories for the NEO™ Personality Inventory-3 (NEO™-PI-3) NEO™ Five-Factor Inventory-3 (NEO™-FFI-3) NEO™ Personality Inventory-Revised (NEO-PI-R™)*. Professional manual. Florida: PAR.
- CSIKSZENTMIHALYI, M. & CSIKSZENTMIHALYI, I.S. (Eds.). (1998). *Optimal experience: Psychological studies of flow in consciousness*. New York: Cambridge University Press.
- EBERLE, S.G. (2014). The elements of play: Toward a philosophy and a definition of play. *American Journal of Play*, 6 (2), 214.
- FURLONG, A. (2012). *Youth studies: An introduction*. England: Routledge.
- GEE, J.P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE)*, 1 (1), 20.
- GEE, J.P. (2007). *Good video games plus good learning* (Vol. 27). New York: Peter Lang.
- GLAZER, S. (2006). "Video Games: Do they have educational value?". *Congressional Quarterly Researcher*, 16 (40), 937–960.
- GREEN, C.S. & BAVELIER, D. (2003). Action video game modifies visual selective attention. *Nature*, 423 (6939), 534–537.
- GRAY, P. (2008). The value of play I: The definition of play provides clues to its purposes. *Psychology Today*, 1–5.
- GROOS, K. (1898). *The play of animals*. New York: D. Appleton and Co.
- HUIZINGA, J. (2014). *Homo Ludens* IIs 86. London, England: Routledge.
- JOHNSON, S. (2005). *Everything Bad is Good for You: How Popular Culture is Making Us Smarter*. London, England: Penguin.
- KELLER, J. & BLESS, H. (2008). Flow and regulatory compatibility: An experimental approach to the flow model of intrinsic motivation. *Personality and Social Psychology Bulletin*, 34, 196–209.
- KOLB, D. (1984). *Experiential learning as the science of learning and development*. New Jersey: Englewood Cliffs NPH.
- KUNREUTHER, F., KIM, H. & RODRIGUEZ, R. (2008). *Working across generations: Defining the future of nonprofit leadership* (Vol. 32). Hoboken, NJ: John Wiley & Sons.
- LEWIN, K. (1951). *Field theory in social science: Selected theoretical papers*. New York: Harper & Brothers.
- MASSIMINI, F. & CARLI, M. (1998). The systematic assessment of flow in daily experiences. In M. Csikszentmihalyi & I. Csikszentmihalyi (Eds.), *Optimal experience: Psychological studies of flow in consciousness* (pp. 226–287). New York: Cambridge University Press.
- PIAGET, J. (1959). *La formation du symbole chez l'enfant: Imitation, jeu et rêve, image et représentation*. Paris, France: Delachaux et Niestlé.
- ROLLINGS, A. & ADAMS, E. (2003). *Andrew Rollings and Ernest Adams on Game Design*. Indianapolis, Ind.: New Riders Games.
- STRAUSS, W. & HOWE, N. (2000). *Millennials rising: The next great generation*. New York: Vintage.
- TWENGE, J.M. (2006). *Generation Me-Revised and Updated: Why Today's Young Americans Are More Confident, Assertive, Entitled—and More Miserable Than Ever Before*. New York: Simon and Schuster.
- WINNICOTT, D.W. (1974). *Gioco e realtà*. Roma: Armando.
- ZAMMITTO, V.L. (2010). *Gamers' personality and their gaming preferences*. Thesis for the degree of Master of Science. Canada: Simon Fraser University.
- ZILLMANN, D. & BRYANT, J. (1994). Entertainment as media effect. In J.B. Zillmann, *Media effects: Advances in theory and research* (pp. 437–461). Hillsdale (NJ): Erlbaum.

Validation and psychometric properties of the Italian Copenhagen Psychosocial Questionnaire II - short version

Ilaria Setti¹, Angelo d'Errico², Daniela Di Cuonzo^{3,4}, Elena Fiabane^{1,5}, Piergiorgio Argentero¹

¹ Unit of Applied Psychology, Department of Brain and Behavioral Sciences, University of Pavia

² Unit of Epidemiology, Piedmont Region, ASL TO3, Grugliasco

³ Department of Psychology, University of Turin

⁴ Unit of Cancer Epidemiology, "Città della Salute e della Scienza" Hospital and CPO Piemonte, Turin

⁵ Department of Physical and Rehabilitation Medicine, Salvatore Maugeri Foundation, Work and Rehabilitation, IRCCS, Scientific Institute of Genoa

● **ABSTRACT.** Il contributo è finalizzato a fornire una validazione italiana del COPSOQ II - versione breve (*Copenhagen Psychosocial Questionnaire II - short version*; Pejtersen, Kristensen, Borg, & Bjorner, 2010), questionario volto ad indagare alcuni dei principali rischi psico-sociali in ambito lavorativo. L'interesse verso questo strumento deriva dalla numerosità degli aspetti critici considerati e dal suo vasto utilizzo in molti contesti internazionali. Le analisi statistiche, condotte su un vasto campione di lavoratori appartenenti al settore industriale - ed ottenuti tramite EFA, CFA, validità concorrente e discriminante - mostrano buone proprietà psicometriche del questionario. I 4 fattori estratti sono: 1) relazioni con il management, 2) supporto da parte dei superiori, 3) controllo, 4) relazioni tra richieste lavorative e salute. I risultati ottenuti suggeriscono la possibilità di utilizzare il questionario anche nel contesto lavorativo italiano.

● **SUMMARY.** *Introduction: In the current workplace it is important to consider psychosocial risks, as they can lead to negative consequences. The aim of this study is to analyse the psychometric properties of the Italian COPSOQ II - short version, a questionnaire which covers a broad range of psychosocial risks. Methods: The questionnaire was administered to 1,845 industry workers. Information about occupational hazards and health conditions was collected. The psychometric properties of the questionnaire were assessed by means of EFA, CFA, discriminant and concurrent validity. Results: The statistical analyses gave support to the validity of the Italian COPSOQ II. The factorial analyses demonstrated that the 4-factors model had the most reasonable good of fit to the data. Conclusions: The results provide evidence of the validity of the Italian COPSOQ II, that can be used to assess psychosocial risks in the Italian work environment across different economic sectors.*

Keywords: COPSOQ II, Psychosocial factors, Validation

INTRODUCTION

The assessment of psychosocial risk factors is actually a relevant subject since many studies have established that the psychosocial working environment can negatively influence workers' health and organizational outcomes (Backé, Seidler, Latza, Rossnagel & Schumann, 2012; Bernal et al., 2015; Bianchi, Schonfeld & Laurent, 2014; Eurofound, 2016; EU OSHA, 2014; Stansfeld, Shipley, Head & Fuhrer, 2012).

In recent decades, important changes have affected the modern workplaces, such as the increasing globalization and significant demographic changes, such as population ageing (Setti, Dordoni, Piccoli, Bellotto & Argentero 2015). These changes are associated with contractual arrangements, temporary work and, more in general, changes in the workforce (EU OSHA, 2014; Eurofound, 2016). In this framework, both governments and researchers have paid an increasing concern about which effects the new forms of work may have on workers' health. In particular, the interest in studying industry employees' well-being is due to the several risks for their health, especially for those who belong to dangerous sectors, mainly caused by the exposure to physical hazards.

Psycho-social risk factors

Beyond the conventional risk factors, it is also important to consider the risks associated with work organization and management, i.e. psychosocial risks, which can occur in every workplace and negatively affect health and business outcomes (Dollard, Skinner, Tuckey, & Bailey, 2007; EU OSHA, 2014). Furthermore, there is a lack of studies on the health of women workers belonging to the industry sector, even if it has been found that they are likely to develop psychosomatic symptoms and work-life imbalance.

Psychosocial risks at work are defined as aspects of the work design and the organization and management of work, and their social contexts (EU OSHA, 2014). Psychosocial risks are usually assessed through subjective methodologies, since they are mostly determined by the way in which people perceive them. For the individual, the main negative effects of psychosocial risks concern mental disturbances - such as depressive symptoms - and physical illness - such as cardiovascular diseases. Specifically, several studies have previously demonstrated that work stress is a significant risk factor for depression (i.e., Blackmore et al., 2007; Bonde,

2008; Stansfeld et al., 2012; Wang et al., 2015). The main negative organizational effects include, for example, reduced performance, and increased absenteeism and injury rates (Backé et al., 2012; Bianchi et al., 2014).

The assessment of psycho-social risk factors: the Copenhagen Psychosocial Questionnaire (COPSOQ)

A number of self-report questionnaires has been developed in order to assess workplace psychosocial risks, but only some of them have been validated in Italian. A comprehensive questionnaire which covers a broad range of psychosocial work environment stressors and resources, and which is applicable to all types of occupations, is the Copenhagen Psychosocial Questionnaire (COPSOQ; Kristensen, Hannerz, Høgh & Borg, 2005). More recently, a second version of the questionnaire (COPSOQ II; Pejtersen et al., 2010) has been developed. The COPSOQ II includes questions on some new factors, such as Reward, Justice, Trust and Discrimination. Since its validation in Denmark, it has been translated into several languages.¹

The purpose of the present study was to give a first contribution to the validation of the Italian COPSOQ II short version, and to examine its psychometric properties.

METHODS

Study sample and procedure

A survey was targeted to Italian workers employed in 25 companies belonging to the following sectors: mechanics, food production, cleaning, elder and child care, textiles and garment, and trading. All workers were invited to participate in the study and were informed on its goals. Through a structured self-report questionnaire, information about socio-demographics, physical and psychosocial occupational hazards, and health conditions was collected. 1,874 questionnaires were filled in and returned to the research team, together with an informed

¹ Documentation on the different COPSOQ questionnaires and on the construction of the psychosocial scales can be found at: <http://www.arbejdsmiljoforskning.dk/en/publikationer/spoergeskemaer/psykisk-arbejdsmiljoe>

consent form signed by each participant. Participation rate was approximately 50%: 1,845 questionnaires have been used for the analysis (completion rate: 98.5%).

Structure of the COPSOQ II - short version

The English version of the short version of the COPSOQ II questionnaire (Pejtersen et al., 2010) was translated into Italian language by one of the authors and subsequently back-translated in English by another author.

The Italian COPSOQ II consisted of 34 items, 28 of which were combined into 14 scales made of two items, and 2 items were used as single variables; the remaining 4 items were dichotomous and were not considered for the analysis. Items were assessed through Likert scales, ranging from 0 (never/disagree) to 3 or 4 (always/agree). As for other national adaptations of the COPSOQ II (see for example the French version by Dupret, Bocéréan, Teherani, Feltrin & Pejtersen, 2012 and the Spanish version by Moncada et al., 2014), not all the original scales were included in the Italian version. As suggested by Persson and Kristiansen (2012), three scales strongly influenced by current mental health conditions, namely the *Meaning of work*, the *Stress* and the *Commitment to the workplace* scales, were not included in this Italian version. The *Meaning of work* scale was excluded also because of the low test-retest reliability shown in the validation study by Thorsen and Bjorner (2010), as well as its uncertain predictive validity reported by a prospective study on burnout among hospital workers, according to which - and in contrast with expectations - higher perceived meaning of work corresponded to higher levels of burnout (Borritz et al., 2005). The *Stress* scale was not included also because it was quite similar to questions of the General Health Questionnaire-12 (GHQ-12; Ware et al., 1996), which was already present in the main questionnaire. Last, the *Commitment to the workplace* scale - which in the original short version of the COPSOQ questionnaire is intended to capture mainly the affective dimension of the organizational commitment - was not included, since in one study it was largely explained by other psychosocial exposures in the workplace, such as influence at work, role ambiguity and quality of leadership (Clausen & Borg, 2010).

The Italian COPSOQ II - short version, included also 4 dichotomous variables on offensive behaviours (bullying,

sexual harassment, physical violence, threat of violence), which were excluded from the factor analysis after verifying that no model, constructed using such variables, was found that displayed an acceptable fit.

Data analysis

Psychometric properties of the COPSOQ II - short version questionnaire were assessed by the following statistical analyses: exploratory and confirmatory factor analysis, discriminant and concurrent validity.

All the dimensions considered, in the COPSOQ II - short version questionnaire, were assessed through two items, except for job satisfaction and general health (1 item) (Kristensen et al., 2005). For the scales made of two items, a composite scale score was computed as the sum of the two items and factorial analyses were conducted using such composite scale scores. If an item of a scale was missing, the scale score was not computed. To make scale scores comparable, they were normalized to 0-100 scales points. Cronbach's alphas and Pearson's correlation coefficients were computed for each scale to evaluate its internal consistency and correlation of the items within each scale. The factor structure of the Italian COPSOQ II - short version was analysed through SEM (Structural Equation Modelling), using the MPlus statistical software, version 7, in two steps:

- an Exploratory Factor Analysis (EFA) was conducted on a 50% random sample of the study population, stratified by sex, age class and economic sector. EFA was performed with oblique GEOMIN rotation and maximum likelihood extraction. Extending previous work done on COPSOQ-II by Bjorner and Pejtersen (2010) and Dupret et al. (2012), the fit of the data was tested for models from 2 to 6 factors, choosing the most parsimonious model, i.e. that with the least number of factors, which displayed acceptable fit indices (Hu & Bentler, 1998); factor loadings below .32 were not considered and are not shown in the results (Tabachnick & Fidell, 2001);
- a Confirmatory Factor Analysis (CFA) was run on the other half of the sample.

Discriminant validity of the questionnaire was assessed comparing means of the factor scores obtained from the CFA by gender and economic sector, through ANOVA, on the ground that scores of the psychosocial factors obtained should show differences across these groups.

The concurrent validity was assessed through multiple linear regression, using as the dependent variable a depression score, obtained by administering the Personal Health Questionnaire Depression Scale, which is a validated instrument for assessing depressive symptoms (PHQ-9; Kroencke, Spitzer & Williams, 2001)². In these analyses, missing values for the categorical variables (age, gender and economic sector) were kept as a separate category, in order to avoid their exclusion from the analysis.

Both discriminant and concurrent validity were evaluated on the factors obtained from a new 4-factors CFA performed on the whole dataset.

RESULTS

The study population was mainly composed of females (71.2%), with more than half subjects in the mid-age (≥ 40 years). The great majority was blue collars (90%), mainly employed in food production (48%), followed by the textile industry (20%), mechanics (17%), elder and child care (6%), trading (5%) and cleaning (2.7%).

Table 1a reports scale names, number of items, Cronbach's alphas, correlation coefficients between items in each scale, mean scores (standard deviations) and number of missing values. Table 1b shows the frequency distribution of the COPSOQ II "General health" variable in the sample. Internal consistency of the scales was generally high, although six of them did not reach the threshold value of .70 and one - Skill Discretion - was even below .60.

Exploratory factor analysis

Results of the Exploratory Factor Analysis indicated that the goodness of fit started to be satisfying above 3 factors (see Table 2). The 4-factors solution showed satisfactory levels in chi-square test, RMSEA, CFI, TLI and SRMR indices, so it was chosen as the most parsimonious model, including the least number of factors, on which the CFA was conducted.

The analysis allowed identifying the following factors, together with their associated scales (see Table 3):

1. *Relations with Management*: Predictability, Reward, Role

Clarity, Trust, Justice;

2. *Supervisor Support*: Quality of Leadership, Supervisor Support, Job Satisfaction;
3. *Job Control*: Decision Authority, Skill Discretion, Predictability;
4. *Job Demands-Health Interface*: Work Pace, Emotional Demand, Quantitative Demand, Work-Family Conflict, Burnout, Perceived General Health.

Confirmatory factor analysis

In order to obtain a satisfactory model fit, and as indicated by modifications indices, in the CFA one variable, i.e. Predictability, was allowed to load on different factors (*Relations with Management* and *Job Control*).

The results of the CFA demonstrated that the 4-factors model had a reasonable good of fit to the data (see Table 4).

Discriminant validity

The results of the ANOVA indicate that all factors were sufficiently capable of discriminating among genders and different economic sectors (all $p < .01$), with the strongest differences found for *Job Control* and for the *Demand-Health* factor ($p < .0001$) (see Table 5).

Concurrent validity

A multiple regression analysis with PHQ-9 (Kroencke et al., 2001) depression score as the predicted variable and the Italian COPSOQ II - short version factors' scores as independent variables was setup. Only one factor was included in each regression model, as factors were strongly correlated, with the consequence of multicollinearity among them when put together in a single model (see Table 6). Therefore, Table 6 displays the regression coefficient of depression associated with each factor, together with the R squared, p -value and confidence limits of the four regression models, adjusted for age, gender, and economic sector. Depression was significantly associated with all factors in the expected direction (all $p < .001$), with the strongest association and the largest explained variability observed for the *Demand-Health* dimension ($\beta = .142$; $R^2 = .32$).

2 available at: www.sipc.eu/share/pagine/55/PHQ-9.pdf

Table 1a – Summary statistics of the COPSOQ II – short version - variables

| COPSOQ continuous scales | N° items (alpha) | Correlation between items (r) | Mean (SD) | Missing values (%) |
|--------------------------|------------------|-------------------------------|-------------|--------------------|
| Quantitative demand | 2 (.62) | .45 | 2.79 (1.98) | 120 (6.5) |
| Work pace | 2 (.89) | .81 | 5.60 (2.06) | 68 (3.7) |
| Emotional demand | 2 (.66) | .49 | 3.08 (2.44) | 96 (5.2) |
| Decision authority | 2 (.66) | .50 | 2.65 (2.36) | 94 (5.1) |
| Skill discretion | 2 (.55) | .38 | 2.76 (2.13) | 60 (3.3) |
| Predictability | 2 (.63) | .46 | 3.22 (2.14) | 62 (3.4) |
| Reward | 2 (.70) | .54 | 2.90 (1.92) | 153 (8.3) |
| Role clarity | 2 (.60) | .43 | 4.81 (1.99) | 116 (6.3) |
| Quality of leadership | 2 (.76) | .62 | 3.44 (2.15) | 132 (7.2) |
| Supervisor support | 2 (.83) | .70 | 3.96 (2.35) | 87 (4.7) |
| Job satisfaction | 1 | - | 1.67 (.78) | 70 (3.8) |
| Work-family conflict | 2 (.84) | .72 | 3.21 (1.87) | 66 (3.6) |
| Trust | 2 (.78) | .64 | 3.75 (1.73) | 163 (8.8) |
| Justice | 2 (.74) | .59 | 3.32 (1.98) | 152 (8.2) |
| Burnout | 2 (.81) | .67 | 4.42 (2.19) | 70 (3.8) |

Table 1b – Frequency distribution of the “General health” COPSOQ II – short version - variable

| General health | N° items | % | Missing values |
|----------------|----------|-------|----------------|
| Very poor | | 2.71 | |
| Poor | | 6.47 | |
| Fair | 1 | 55.84 | 38 (2.1) |
| Good | | 28.61 | |
| Very good | | 6.36 | |

DISCUSSION

The aim of this study was to provide evidence of the validity of the Italian COPSOQ II - short version with data obtained from a large sample of industry employees. In the last decades, the COPSOQ has become a popular and relevant tool for

both research and preventive practice in the workplace, it was translated in several languages and used in many international studies. As for other countries, also the Italian context may benefit from the diffusion of this instrument in order to measure the psychosocial work environment (Pejtersen et al., 2010). As the French validation (Dupret et al., 2012), we decided to use the

Table 2 – Goodness of fit indices for 2, 3, 4, 5 and 6-factor models from Exploratory Factor Analysis

| Model | Chi-square | df | Chi-square p-value | RMSEA | RMSEA 90% CI | CFI | TLI | SRMR |
|-----------|------------|----|--------------------|-------|--------------|------|------|------|
| 2 factors | 703.39 | 89 | <.0001 | .087 | .081 – .093 | .877 | .834 | .051 |
| 3 factors | 350.26 | 75 | <.0001 | .064 | .057 – .070 | .945 | .912 | .029 |
| 4 factors | 198.65 | 62 | <.0001 | .049 | .042 – .57 | .973 | .947 | .024 |
| 5 factors | 127.971 | 50 | <.0001 | .041 | .033 – .050 | .984 | .963 | .016 |
| 6 factors | 119.83 | 39 | <.0001 | .048 | .038 – .058 | .984 | .950 | .016 |

Table 3 – COPSOQ scales with loadings associated with the 4 factors identified by Exploratory Factor Analysis (loadings factors <.3 not shown)

| COPSOQ scales | 1-Relations with Management | 2-Job Control | 3-Supervisor Support | 4-Job Demands-Health Interface |
|--------------------------|-----------------------------|---------------|----------------------|--------------------------------|
| Quantitative demand | | | | .305 |
| Predictability | .469 | .362 | | |
| Reward | .626 | | | |
| Role clarity | .378 | | | |
| Trust | .598 | | | |
| Justice | .766 | | | |
| Job satisfaction | | | .335 | |
| Decision authority | | .657 | | |
| Skill discretion | | .785 | | |
| Emotional demand | | | | .478 |
| Quality of leadership | | | .945 | |
| Supervisor support | | | .574 | |
| Work pace | | | | .447 |
| Work-family conflict | | | | .632 |
| Burnout | | | | .837 |
| Perceived general health | | | | -.465 |

Table 4 – Goodness of fit indices for the 4-factor model from Confirmatory Factor Analysis

| Model | Chi-square | df | Chi-square p-value | RMSEA | RMSEA 90% CI | CFI | TLI | SRMR |
|-------|------------|----|--------------------|-------|--------------|------|------|------|
| CFA | 412.380 | 97 | <.0001 | .059 | .053–.065 | .937 | .922 | .046 |

Table 5 – Analysis of variance on factor scores by gender and economic sector

| Covariates | 1 - Relations with Management | 2 - Supervisor Support | 3 - Job Control | 4 - Job Demands-Health interface |
|------------------------|-------------------------------|------------------------|-----------------|----------------------------------|
| Gender | | | | |
| <i>p</i> -value | .0083 | .0114 | <.0001 | <.0001 |
| R-squared | .004 | .004 | .039 | .017 |
| Economic sector | | | | |
| <i>p</i> -value | .0011 | .0021 | <.0001 | <.0001 |
| R-squared | .011 | .01 | .06 | .046 |

Table 6 – Multiple linear regression models of depression with psychosocial factors, adjusted for age class, sex and economic sector ^a

| Factors | Beta | R ² | <i>p</i> -value | 95% CI |
|---------------------------------|-------|----------------|-----------------|---------------|
| 1. Relations with Management | -.090 | .15 | <.001 | -.102 - -.783 |
| 2. Supervisor Support | -.091 | .15 | <.001 | -.102 - -.079 |
| 3. Job Control | -.048 | .06 | <.001 | -.060 - -.037 |
| 4. Job Demands-Health Interface | .142 | .32 | <.001 | .132 - .153 |

Note. Coefficients corresponding to a 1% increase in the factor scales normalized to 0-100 points.

^a R² of the regression model with only age class, sex and economic sector: .022

short version, which is designed for use in firms and it is more likely to be accepted by Italian organizations which generally do not use long questionnaires. Whilst the original long version of the COPSOQ II is made of 128 items, and the medium one of 87 items, we propose the following model of the relationships among the factors: the 30 items of the questionnaire, excluding four dichotomous ones on offensive behaviors, are grouped in 16 scales (14 of them made of two items, and the remaining two

items used as single variables) that resulted in 4 factors. Some differences with the dimensions intended to be captured by the original questionnaire (Kristensen et al., 2005; Pejtersen et al., 2010) emerged. A main difference is that the *Interpersonal Relations* and *Leadership* factor was replaced in our analysis by two separate factors, one representing the quality of the relationship with the management and with the company in general terms, the other one that with the direct supervisor.

This finding suggests that it is possible to separate a dimension concerning the relationship of the workers with the company's organizational structure, its way of functioning and the social climate perceived at the workplace, from one specifically addressing the relation with a person who is in charge of supervising and support the worker. It has been commented that the dimension of interpersonal relationships at work is a construct partially overlapping with that of social support, but that they differ in that the latter refers more to social relations specifically oriented towards task performing, whereas the former to the opportunity for pleasant and meaningful contacts and for feeling part of a greater social system (Schabracq, 2003). The observation that job satisfaction loaded on the Supervisor Support factor, rather than on that of *Relations with Management*, suggests that job satisfaction is more strongly determined by a supportive work environment, where the relationship with supervisors is characterized by getting help in performing the job, than by good social relationships with the management. Another important difference is that the demand and health dimensions, conceptualized as distinct by the COPSOQ authors, in the present study were merged in a single factor together with work-family conflict, which is an established mediator of the relationship between demand and health (du Prel & Peter, 2015), indicating that the demand dimension is very intertwined with that of work-family conflict and with health. This observation could be explained either by the consideration that psychological demand is the dimension having more impact on health, as suggested by several reviews on mental health (Bonde, 2008; Netterstrom et al., 2008; Stansfeld & Candy, 2006), or that in cross-sectional studies - like the present one - exposure to demand, among workers with health problems or high levels of work-family conflict, is affected by overestimation more than the other psychosocial exposures examined, as observed by Tang (2014).

As for the French and the Danish versions of the questionnaire, some of the Cronbach's alphas values are relatively weak, as 6 out of 14 scales show an internal consistency lower than the threshold value of .7. As suggested by Dupret et al. (2012), the medium and short versions of the questionnaire have been developed by focusing mainly on the content validity of the instrument, and not only on high internal consistency. Furthermore, it is widely recognized that the internal consistency is affected by the number of items (Cortina, 1993): the low internal consistency of some scales of the COPSOQ II - short version, appears explained by the fact that the scales are based at most on only two items.

The ANOVA analysis allowed assessing the discriminant validity of the questionnaire, with results showing that all factors' means were significantly different across gender and economic sector, meaning that they were all able to discriminate among these groups. In general, the ability of the COPSOQ factors to discriminate among different economic sectors gives support to the opinion that this questionnaire is applicable to a wide range of occupations (Dupret et al., 2012).

The concurrent validity was assessed through multiple regression analyses with depression as the predicted variable: the focus on depression is due to its detrimental effects on individuals' quality of life, functioning and job performance. All the factors - especially the *Demand-Health* one - predict depressive symptoms. Our results confirm previous research on the direct association of well-being with job control, showing that the perception of low decision latitude is associated to depressive symptoms (Arcury, Grzywacz, Chen, Mora & Quandt, 2014), as well as with *Supervisor Support*, according to which supervisors' support would buffer the effects of job demands against negative effects (Arcury et al., 2014; Hall, Dollard, Winefield, Dormann & Bakker, 2013). The mechanisms of the development of depression in relation to these stressors are mostly unknown, with several neurotransmitters likely playing a role, but a sustained activation of the hypothalamus-pituitary-adrenal cortex (HPA) axis appears involved in chronic stress, as suggested by persistently high cortisol levels in both subjects exposed to different types of stressors and patients affected by depression (Chida and Steptoe, 2009). The relation between poor health and depression has been widely demonstrated, with particular attention for the role of burnout: recent research demonstrated that people with severe symptoms of burnout meet also diagnostic criteria for depression (Bianchi et al., 2014). Another critical psychosocial exposure, known to be associated with depression and loading on the *Demand-Health* factor, is work-family conflict, since the need to conciliate work and family demands may produce negative health effects (Garcia, Milkovits & Bordia, 2014). With regard to the factor *Relations with Management*, there is evidence of the link between job characteristics, management conditions and mental health problems. Because individuals spend large part of their time at work, it is conceivable that job relations play a key role in the aetiology of health complaints. That is, stressful work conditions - due, for example, to low predictability or unfair rewards - may predict poor mental health. Indeed, employees who report high levels of job

strain due to job characteristics are more likely to suffer from depression (Wang & Patten, 2001). In particular, it seems that some “job context” management practices (such as the level of role clarity), if not properly managed, are relevant in determining negative health effects: role ambiguity seems to be related to depression. More in general, and as confirmed by our results, management practices based on organizational justice, fair rewards, and role clarity can reduce health complaints, also in terms of depression (Mackie, Holahan & Gottlieb, 2001).

The results of this study should be interpreted in light of some limitations. First of all, even if our sample was drawn from several economic sectors, it was not representative of the general population, as it mainly included blue-collar workers. Furthermore, we are aware of the importance of being prudent in comparing our results with others, based on medium or long versions of the COPSOQ. Because the medium and short scales do not need to be made of the same number of items across different countries (Dupret et al., 2012), we chose to use not all the items of the short version of COPSOQ II, and this could represent a further limitation. However, the items selection was firstly due to parsimony reason and secondly because three scales may be strongly influenced by participants' current mental state.

This study has also a number of strengths, including the large sample size and the extensive collection of information,

which allowed verifying not only the psychometric properties of the questionnaire, but also relationships among variables. Furthermore, the ability of the COPSOQ factors to discriminate among different economic sectors confirms the applicability of this questionnaire to a wide range of occupations (Dupret et al., 2012). Finally, the prevalent presence of females should be considered a strength of this study because there is a dearth of research specifically focused on female industry workers (Arcury et al., 2014).

From an applicative perspective, the assessment of psychosocial risks may be considered as the first preventive step that represents a relevant topic in the current workplace, since psychosocial risk control can help reducing accidents and absenteeism. Accordingly, organizations engaged in managing psychosocial risks can be more competitive, also by improving employees' health at the organizational level. Indeed greater job satisfaction and productivity, and lower absenteeism, have been found as the main organizational outcomes resulting from an effective psychosocial risks management (EU OSHA, 2014). Management practices should be focused at improving the psychosocial safety climate, which represents the first and main step towards higher levels of employees' well-being.

In conclusion, our results support the validity of the Italian COPSOQ II - short version, but further studies might be carried out in order to confirm these findings.

References

- ARCURY, T.A., GRZYWACZ, J.G., CHEN, H., MORA, D.C. & QUANDT, S.A. (2014). Work organization and health among immigrant women: Latina manual workers in North Carolina. *American Journal of Public Health, 104*, 2445-2452.
- BACKÉ, E.M., SEIDLER, A., LATZA, U., ROSSNAGEL, K. & SCHUMANN, B. (2012). The role of psychosocial stress at work for the development of cardiovascular diseases: a systematic review. *International Archives of Occupational and Environmental Health, 85* (1), 67-79.
- BERNAL, D., CAMPOS-SERNA, J., TOBIAS, A., VARGAS-PRADA, S., BENAVIDES, F.G. & SERRA, C. (2015). Work-related psychosocial risk factors and musculoskeletal disorders in hospital nurses and nursing aides: A systematic review and meta-analysis. *International Journal of Nursing Studies, 52* (2), 635-648.
- BIANCHI, R., SCHONFELD, I.S. & LAURENT, E. (2014). Is burnout a depressive disorder? A reexamination with special focus on atypical depression. *International Journal of Stress Management, 21*, 307-324.
- BJORNER, J.B. & PEJTERSEN, J.H. (2010). Evaluating construct validity of the second version of the Copenhagen Psychosocial Questionnaire through analysis of differential item functioning and differential item effect. *Scandinavian Journal of Public Health, 38*, 90-105.
- BLACKMORE, E.R., STANSFELD, S.A., WELLER, I., MUNCE, S., ZAGORSKI, B.M., & STEWART, D.E. (2007). Major depressive episodes and work stress: results from a national population survey. *American Journal of Public Health, 97* (11), 2088-2093.
- BONDE, J.P.E. (2008). Psychosocial factors at work and risk of

- depression: A systematic review of the epidemiological evidence. *Occupational and Environmental Medicine*, 65 (7), 438-445.
- CHIDA, Y. & STEPTOE, A. (2009). Cortisol awakening response and psychosocial factors: a systematic review and meta-analysis. *Biological Psychology*, 80, 265-278.
- CORTINA, J.M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*, 78, 98.
- DOLLARD, M., SKINNER, N., TUCKEY, M.R. & BAILEY, T. (2007). National surveillance of psychosocial risk factors in the workplace: An international overview. *Work & Stress*, 21 (1), 1-29.
- DUPRET, E., BOCÉRÉAN, C., TEHERANI, M., FELTRIN, M. & PEJTERSEN, J.H. (2012). Psychosocial risk assessment: French validation of the Copenhagen Psychosocial Questionnaire (COPSOQ). *Scandinavian Journal of Public Health*, 40, 482-490.
- du PREL, J.B. & PETER, R. (2015). Work-family conflict as a mediator in the association between work stress and depressive symptoms: cross-sectional evidence from the German lidA-cohort study. *International Archives of Occupational and Environmental Health*, 88, 359-68.
- EUROFOUND (2016) *The sixth European Working Conditions Survey*, <https://www.eurofound.europa.eu/surveys/european-working-conditions-surveys/sixth-european-working-conditions-survey-2015>
- EUROPEAN AGENCY FOR SAFETY AND HEALTH AT WORK (2014). *Calculating the cost of work-related stress and psychosocial risks*. Luxembourg: Publications Office of the European Union.
- GARCIA, P.R.J. M., MILKOVITS, M. & BORDIA, P. (2014). The impact of work-family conflict on late-career workers' intentions to continue paid employment: A social cognitive career theory approach. *Journal of Career Assessment*, 22 (4), 682-699.
- HALL, G.B., DOLLARD, M.F., WINEFIELD, A.H., DORMANN, C. & BAKKER, A.B. (2013). Psychosocial safety climate buffers effects of job demands on depression and positive organizational behaviors. *Anxiety Stress et Coping*, 26 (4), 355-377.
- HU, L.T. & BENTLER, P.M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological methods*, 3 (4), 424.
- KRISTENSEN, T.S., HANNERZ, H., HØGH, A. & BORG, V. (2005). The Copenhagen Psychosocial Questionnaire-a tool for the assessment and improvement of the psychosocial work environment. *Scandinavian Journal of Work, Environment et Health*, 438-449.
- KROENCKE, K., SPITZER, R. & WILLIAMS, J. (2001). The phq-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16, 606-613.
- MACKIE, K.S., HOLAHAN, C.K. & GOTTLIEB, N.H. (2001). Employee involvement management practices, work stress, and depression in employees of a human services residential care facility. *Human Relations*, 54 (8), 1065-1092.
- MONCADA, S., UTZET, M., MOLINERO, E., LLORENS, C., MORENO, N., GALTÉS, A. & NAVARRO, A. (2014). The Copenhagen Psychosocial Questionnaire II (COPSOQ II) in Spain: A tool for psychosocial risk assessment at the workplace. *American Journal of Industrial Medicine*, 57, 97-107.
- NETTERSTROM, B., CONRAD, N., BECH, P. et al. (2008). The relation between work-related psychosocial factors and the development of depression. *Epidemiologic Reviews*, 30, 118-132
- PEJTERSEN, J.H., KRISTENSEN, T.S., BORG, V. & BJORNER, J.B. (2010). The second version of the Copenhagen Psychosocial Questionnaire. *Scandinavian Journal of Public Health*, 38, 8-24.
- PERSSON, R. & KRISTIANSEN, J. (2012). The challenge of assessing the psychosocial working environment: why some self-reports should not be interpreted as environmental exposures. *Occupational and Environmental Medicine*, 69, 932-933.
- SCHABRACQ, M.J. (2003). Organisational culture, stress and change. In: M.J. Schabracq, J.A.M. Winnubst & C.L. Cooper (Eds.), *The handbook of work and health psychology*. Chichester (UK): John Wiley.
- SETTI, I., DORDONI, P., PICCOLI, B., BELLOTTO, M. & ARGENTERO, P. (2015). Proactive personality and training motivation among older workers: A mediational model of goal orientation. *European Journal of Training and Development*, 39, 681-699.
- STANSFELD, S.A. & CANDY, B. (2006). Psychosocial work environment and mental health: A meta-analytic review. *Scandinavian Journal of Work Environment & Health*, 32, 443-462.
- STANSFELD, S.A., SHIPLEY, M.J., HEAD, J. & FUHRER, R. (2012). Repeated job strain and the risk of depression: Longitudinal analyses from the Whitehall II Study. *American Journal of Public Health*, 102 (12), 2360-2366.
- TABACHNICK, B.G. & FIDELL, L.S. (2001). *Using Multivariate Statistics*. Boston: Allyn and Bacon.
- WANG, J. & PATTEN, S.B. (2001). Perceived work stress and major depression in the Canadian employed population, 20-49 years old. *Journal of Occupational Health Psychology*, 6, 283.
- WANG, S.M., LAI, C.Y., CHANG, Y.Y., HUANG, C.Y., ZAUSZNIIEWSKI, J.A. & YU, C.Y. (2015). The relationships among work stress, resourcefulness, and depression level in psychiatric nurses. *Archives of Psychiatric Nursing*, 29 (1), 64-70.

The CBQ-p: A confirmatory study on factor structure and convergent validity with psychotic-like experiences and cognitions in adolescents and young adults

Andrea Pozza¹, Davide Dèttore²

¹ Department of Experimental and Clinical Medicine, University of Florence

² Department of Health Sciences, University of Florence

• **ABSTRACT.** Il Cognitive Biases Questionnaire for psychosis (CBQ-p), recentemente sviluppato, è un questionario che misura distorsioni cognitive considerate fattori di vulnerabilità e di mantenimento specifici per sintomi psicotici, in particolar modo le convinzioni deliranti. Nel contesto italiano sono assenti strumenti di misura di questi aspetti e, inoltre, nessuno degli studi internazionali ha indagato la validità convergente con esperienze psicotiche sotto-soglia. Il presente studio ha indagato la struttura fattoriale della versione italiana del CBQ-p con analisi confermate in un gruppo di adolescenti e giovani adulti tratti dalla popolazione generale. Un ulteriore obiettivo è stato indagare la sua validità convergente con misure di salienza aberrante, distorsioni cognitive specifiche per i sintomi psicotici, confusione inferenziale ed esperienze psicotiche sotto-soglia. Trecentottantotto adolescenti e giovani adulti tratti dalla popolazione generale (età media = 19.22, 55% femmine) hanno compilato il CBQ-p, misure di distorsioni cognitive specifiche per i sintomi psicotici, rimuginio ed esperienze psicotiche sotto-soglia. È stata scelta una soluzione bifattoriale, composta dal fattore Sovrastima del pericolo e da Percezioni anomale, sulla base dello studio originale di validazione e dei risultati sull'affidabilità. In conclusione, la versione italiana del CBQ-p ha dimostrato adeguate proprietà psicometriche e validità convergente con esperienze psicotiche sotto-soglia.

• **SUMMARY.** The Cognitive Biases Questionnaire for psychosis (CBQ-p) is a recently developed self-report measure assessing cognitive distortions relevant to psychotic symptoms and experiences, specifically for the onset and maintenance of delusional ideas. In Italy, there is a lack of assessment tools measuring these aspects. In addition, no international study investigated the relations of CBQ-p with subthreshold psychotic-like experiences. The current study assessed the factor structure of the Italian CBQ-p with confirmatory analyses in community adolescents and young adults. A further aim was to examine its convergent validity with measures of aberrant salience, cognitive biases specific to psychosis, inferential confusion, worry, and subthreshold psychotic-like experiences. Three hundred eighty-eight adolescents and young adults of the community (mean age= 19.22, 55% females) completed the CBQ-p, measures of cognitive distortions of psychosis, aberrant salience, inferential confusion, worry and subthreshold psychotic-like experiences. Confirmatory factor analysis, internal consistency and Pearson's correlations were computed. The Italian CBQ-p demonstrated good psychometric properties; the total scale and subscales reported convergent validity with subthreshold psychotic experiences.

Keywords: Cognitive biases, Distortions, Cognitive Biases Questionnaire for psychosis, Psychotic experiences, Psychotic-like features, Adolescents

INTRODUCTION

Research with population-based studies has shown that the dichotomous disease model of psychosis can be replaced with a more comprehensive model of psychosis as an extended phenotype across clinical and non-clinical manifestations, where at one end of the continuum lies schizophrenia, in the middle are non-psychotic psychological disorders with psychotic experiences (for example, panic disorders with derealisation or depression with psychotic features), and at the other extreme lie these experiences in subthreshold intensity among healthy, non-help-seeking individuals (van Os & Linscott, 2012). Schizophrenia only represents the poorest outcome segment of this wider spectrum of psychotic manifestations (van Os & Linscott, 2012). Indeed, subthreshold psychotic features are not uncommon in the general population: having one of psychotic symptoms was reported in about 25% ($n = 5877$) of the American population (Kendler, Gallagher, Abelson & Kessler, 1996) and 17.50% ($n = 2548$) of the German population (Spauwen, Krabbendam, Lieb, Wittchen & van Os, 2003). In an English-Italian cohort study (Ohayon, 2000), where hypnagogic and hypnopompic hallucinations were considered, the percentage increased to about 40% ($n = 13057$).

Subthreshold psychotic-like experience in the general population include a variety of subtypes, such as sensory experiences which are not shared by other present people, related to hearing sounds (voices, noises), unexplained visual experiences (visions, seeing ghosts), unusual bodily experiences (feeling touched), distorted self-experiences (a decreased ability to be affected by people and events, depersonalization experiences, feelings of derealisation), and perplexity (difficulty automatically grasping the meaning of the everyday situations) (Kelleher & Cannon, 2011).

In the last decade, there has been an increased attention by researchers and clinicians to cognitive factors potentially involved in psychotic symptoms and experiences (Garety, Bebbington, Fowler, Freeman & Kuipers, 2007). Contemporary cognitive models assume that biased reasoning processes contribute to the onset and maintenance of delusional ideas (Garety et al., 2005). Following the continuum of psychotic disorders, the knowledge of factors associated with psychotic experiences in the general population can inform prevention, suggesting the development of strategies, which could target these factors with the aim to prevent psychosis (van der Gaag, Nieman & van den Berg, 2013).

A variety of assessment tools has been designed to measure cognitive factors involved in psychotic symptoms in both clinical and non-clinical groups to inform clinical and prevention practice (Aardema, O'Connor, Emmelkamp, Marchand & Todorov, 2005; Cicero, Kerns & McCarthy, 2010; van der Gaag et al., 2013). The *Cognitive Biases Questionnaire for psychosis (CBQ-p)*; Peters et al., 2014) is a recently developed self-report measure, aimed to assess cognitive distortions considered relevant in psychotic symptoms and experiences, specifically for delusional ideas. It is based on the *Cognitive Style Test (CST)*; Blackburn, Jones & Lewin, 1986), which was designed to measure common thinking biases in depression. The original CST consisted of 30 vignettes describing everyday scenarios. Respondents are asked to select their own cognitive response to each scenario out of 4 possible reactions: a very negative one (scored 4), a somewhat negative (3), a somewhat positive (2) or a very positive (1). The CBQ-p is built on this format and is composed of 30 adapted vignettes relevant to psychotic symptoms/experiences. The scale measures five specific cognitive biases, that literature and clinical impressions of a large groups of researchers and therapists specialized in the field identified as maintenance factors of psychotic experiences/symptoms (Garety et al., 2007; Peters et al., 2014; van der Gaag et al., 2013): Jumping-to-conclusions (drawing firm interpretations based on scarce evidence), Intentionalising (interpreting events or behaviours as deliberate), Catastrophising (worst-case-scenario thinking), Emotional Reasoning (describing definite threatening meaning to one's feelings on a particular moment) and Dichotomous Thinking (i.e. "black or white" reasoning style). Fifteen of the 30 scenarios of the CBQ-p relate to "anomalous experiences", and the other 15 concern "threatening events". Respondents must select one out of three given statements as their most likely reaction to the presented scenario, with one possible choice identifying the presence of bias (scored 3), another the absence of bias (scored 1), and a third option the presence of bias with some doubt about it (scored 2).

To validate the CBQ-p, Peters and colleagues (2014) tested different models, including a single-factor, two-factor and a five-factor model. The model composed by five correlated factors comprised the five above-mentioned biases; the two-factor model instead consisted of the Anomalous experiences and Threatening events dimensions; finally, the model with a single factor represented a general interpretation bias. Findings showed that a 2-factor solution, with the two factors

correlated to each other, had the best fit to the data (Peters et al., 2014). Although the five-factor model demonstrated good fit as well, the different factors were highly correlated, and could hardly be differentiated empirically. In addition, good internal consistency was found for the first model (Cronbach's alpha coefficient was .89 for the total group). A point of criticism was, however, that scores on the CBQ-p were not associated with existing self-report measures and experimental tasks supposed to cover similar reasoning biases, such as the *Beads Task*, *Catastrophising Interview* (Startup, Freeman & Garety, 2007), *Ambiguous Intentions and Hostility Questionnaire* (AIHQ; Combs, Penn, Wicher & Waldheter, 2007), the *Dysfunctional Attitudes Scale* (DAS; Weissman & Beck, 1978). This evidence suggested that the CBQ-p maybe does not conceptualize reasoning, judgment or decision-making processes, but "rather taps into a different construct, perhaps a bias of interpretation" (Peters et al., 2014).

No study in the international literature investigated the relations of the CBQ-p with subthreshold psychotic-like experiences, despite the growing attention paid to these phenomena in the general population, not only in clinical groups. In addition, in the Italian context, a translated version of the CBQ-p does not still exist and there is a lack of assessment tools to measure cognitive biases specific to psychotic symptoms and experiences.

Starting from these points, the aim of the current study was to investigate the factor structure of the Italian version of the CBQ-p with confirmatory analyses in a group of adolescents and young adults of the community. Subsequently, reliability as internal consistency was assessed. A further aim was to examine its convergent validity with measures of aberrant salience, cognitive biases specific for psychosis, inferential confusion, worry, and subthreshold psychotic experiences. Both aberrant salience and inferential confusion have been shown to be associated specifically with psychotic and delusional symptoms and experiences in clinical and non-clinical groups (Aardema et al., 2005; Cicero et al., 2010).

METHOD

Participants and procedure

The total group consisted of 388 adolescents and young adults recruited from the Italian community. Mean age was 19.22 years ($SD = 4.55$, $range = 14-35$). Two hundred and

thirteen (55%) were females. An overview of demographics is presented in Table 1. Participants were recruited from high schools and universities. Data were collected from October 2015 to November 2016. All the participants completed the questionnaires individually or in groups in classrooms. In accordance with the Ethical Principles of Psychologists and Code of Conduct (American Psychological Association, 2002), all the participants, who were included, provided written informed consent to participate in the study after having received a detailed description of the aims. For participants aged under 18 years, written informed consent was requested from both parents. Individuals with certified learning disabilities and mental retardation problems were excluded.

Table 1 – Sociodemographics of the community group of adolescents and young adults ($n = 388$)

| | M (SD; range) | n (%) |
|--------------------------|------------------------|--------------|
| Age (years) | 19.22 (4.55; 19-35) | |
| Sex | | |
| Females | | 213 (55) |
| Education | | |
| Primary school license | | 0 |
| Secondary school license | | 311 (80.2) |
| High school license | | 31 (8) |
| Degree | | 41 (10.6) |
| Ph.D. | | 5 (1.3) |
| Work status | | |
| Student | | 339 (86.8) |
| Employed | | 42 (10.8) |
| Unemployed | | 7 (1.8) |
| Civil status | | |
| Single | | 384 (98) |
| Married/cohabitant | | 3 (.9) |
| Separated | | 1 (.1) |

Measures

A packet of the following self-report measures was administered.

– *Cognitive Biases Questionnaire for psychosis (CBQ-p)*

The CBQ-p (Peters et al., 2014) consists of 30 vignettes of everyday situations (half pleasant and half unpleasant). Respondents imagine themselves in each situation and choose 1 of 3 possible cognitive responses to the scenario. Each vignette requests a forced-choice response between 3 statements, illustrating absence of bias (score of 1), possible presence of bias (score of 2), and likely presence of bias (score of 3). The potential range of scores is 30-90. Before proceeding to the translation of the measure, permission was obtained by the author who developed the scale (Prof. Emanuelle Peters, Department of Psychology, King's College, Institute of Psychiatry, London, UK). The translation process was carried out using a protocol conforming to international standards (Behling & Law, 2000), which included a forward and a backward translation. The forward translation was made by a native Italian-speaking clinical psychologist with excellent fluency in English, then checked by two Italian professional translators. The forward translator discussed the translation in consultation meetings with the professional translators. Subsequently, this version was translated back to English by a bilingual professional translator, who was blind to the original version of the CBQ-p. The back-translation was then compared with the original version, and discussed by the forward-translator with the back-translator in a consensus meeting, which generated the final Italian version of the CBQ-p.

– *Aberrant Salience Inventory (ASI)*

The Aberrant Salience Inventory (ASI; Cicero et al., 2010) is a 29-item self-report questionnaire with a dichotomous response format (“Yes” = 1, “No” = 0), which has five subscales measuring different aspects of the experience of aberrant salience. This is a cognitive factor believed to be specific to the development of delusional ideas (Kapur, 2003). The questionnaire assesses feelings of increased significance (e.g., “Do certain trivial things suddenly seem especially important or significant to you?”), sharpening of senses (e.g., “Do your senses ever seem especially strong or clear?”), impending understanding (e.g., “Do you sometimes feel like you are on the verge of something really big or important but you aren't sure what it is?”),

heightened emotionality (e.g., “Do you go through periods in which you feel over-stimulated by things or experiences that are normally manageable?”), and heightened cognition (e.g., “Do you ever feel like the mysteries of the universe are revealing themselves to you?”). High scores indicate more intense aspects of aberrant salience. The ASI demonstrated good internal consistency and satisfactory convergent validity with measures of psychosis-proneness (Cicero et al., 2010), and the Italian version of the ASI had good internal consistency (Cronbach's alpha = .89). In the present study, internal consistency was good (Cronbach's alpha = .79).

– *Davos Assessment of Cognitive Biases Scale (DACOBS)*

The Davos Assessment of Cognitive Biases Scale (DACOBS; van der Gaag et al., 2013) was developed by the research group of Mark van der Gaag and colleagues (2013). It consists of 42 statements relating to seven subscales, constructed by means of exploratory factor analysis: (1) Jumping to conclusions bias, (2) Belief Inflexibility bias (i.e. confirmation bias), (3) Attention to threat bias, (4) External attribution bias, (5) Social cognition problems, (6) Subjective cognitive problems, and (7) Safety behaviours. Respondents score each statement using a 7-point rating scale, ranging from 1 (“Totally disagree”) to 7 (“Totally agree”) taking into account the past two weeks. Van der Gaag and colleagues (2013) found good reliability (Cronbach's alpha = .90), with the scale differentiating adequately between schizophrenia spectrum patients and healthy control individuals. High scores indicate more severe cognitive biases specific to psychotic symptoms. The Italian version (Pozza & Dèttore, submitted) has been translated according to forward- and backward-translation and showed a seven-factor solution with acceptable to good internal consistency (range of Cronbach's alpha = .75-.84). It showed satisfactory convergent validity with significant moderate correlations with the Paranoid Ideation and the Psychoticism subscales of the *Symptoms Checklist 90-Revised (SCL-90-R*; Derogatis, 1992). In the current study, internal consistency was good (Cronbach's alpha = .84).

– *Inferential Confusion Questionnaire-Extended Version (ICQ-EV)*

The ICQ-EV (Aardema al. 2010) is a 30-item questionnaire on a 6-point Likert scale (“Strongly disagree” = 1, “Strongly agree” = 6) to measure inferential confusion tendencies.

It showed excellent internal consistency (Cronbach's $\alpha = .96$) and strong correlations with OCD symptoms beyond the effects of negative affect and obsessive beliefs (Aardema, Trihey, Kleijer, O'connor & Emmelkamp, 2006). Inferential confusion represents a cognitive factor believed to be associated with delusional thinking (Aardema et al., 2005) as patients with delusional disorders reported higher scores on this measure compared with non-clinical controls. High scores on the ICQ-EV represent an overreliance on imagination, a distrust of the senses, and a tendency to confuse imagination with reality. The Italian version (Pozza, Torniai & Dèttore, submitted) showed excellent internal consistency (Cronbach's $\alpha = .97$). In the current study, internal consistency was excellent (Cronbach's $\alpha = .92$).

– *Penn State Worry Questionnaire (PSWQ)*

The PSWQ (Meyer, Miller, Metzger & Borkovec, 1990) is a self-report measure, designed to cover aspects of clinically significant worry, specifically the tendency, intensity, and uncontrollability of worry. It consists of 16 items rated on a 5-point Likert scale, with values ranging from 1 (“Not at all typical of me”) to 5 (“Very typical of me”). Meyer and colleagues (1990) conducted a series of studies evidencing very good or excellent internal consistency ($\alpha = .88-.95$), good test-retest reliability ($r = .74-.92$), and good convergent and divergent validity in clinical and non-clinical samples. High scores indicate more intense clinical worry. The Italian version (Morani, Pricci & Sanavio, 1999) had good internal consistency. In the current study, internal consistency was good (Cronbach's $\alpha = .86$).

– *Screening for Psychotic Experiences (SPE)*

The SPE (Magnani et al., 2010) is a self-report scale, composed by 20 items, which represents a shorter version of the *Prodromal Questionnaire (PQ)*; Loewy, Bearden, Johnson, Raine & Cannon, 2005), a 92-item self-report screening instrument, aimed to identify individuals needing for further diagnostic assessments of at-risk-mental states of psychotic disorders and symptoms. The SPE covers subthreshold experiences and feelings typical of psychotic symptoms, that need for further assessments, such as self-reference ideas, delusional perceptions, self-neglecting, depersonalization/derealization (eg, “I think that people look at me or talk about me”). All of these experiences have been found to be predictors of psychotic symptoms and have been defined as “early initial prodromal

states” (Yung et al., 1998). Respondents are asked to give an answer based on a true/false response format, where “True” is scored as 1 and “False” as 0. High scores suggest more intense subthreshold psychotic experiences. The SPE demonstrated good internal consistency ($\alpha = .79$) in a large group of Italian adolescents (Magnani et al., 2010). In the current study, internal consistency was acceptable ($\alpha = .77$).

Statistical analysis

The distributional properties of the CBQ-p items were assessed by inspecting the skewness and kurtosis indices of the items' distributions. Subsequently, to examine fit of the data to the factor structure, confirmatory factor analysis (CFA) was carried using a structural equations modelling (Bollen, 1989). As reported in Peters and colleagues (2014), three models were tested: a two-correlated, five-correlated, and a single-factor models, respectively.

To evaluate goodness of fit of the model to the data, the following indices recommended by Hu and Bentler (1999) were adopted: the Adjusted Goodness-of-Fit Index (AGFI), the Goodness of Fit Index (GFI), the Bentler-Bonett Normed Fit Index (NFI; Bentler & Bonett, 1980), the Bollen's Relative Fit Index (RFI; Bollen, 1986). For these indices, values ranging from .95 and 1 represent excellent fit, values ranging from .90 and .95 good fit. In addition, the Root Mean Square Residual (RMR) was considered; values less than .08 represent acceptable fit, and those less than .05 represent good fit.

Reliability was examined as internal consistency using Cronbach's α coefficients and assessed according to Nunnally and Bernstein's guidelines (1994) ($\alpha > .70$ = acceptable, $\alpha > .80$ = good, $\alpha > .90$ = excellent).

Independent sample *t*-tests were calculated, in order to compare means of male and female subgroups on the CBQ-p total and subscales. Effect sizes were computed as Cohen's *d* indices. Convergent validity was examined investigating the correlations between the CBQ-p scores and measures of cognitive biases involved in psychotic symptoms, aberrant salience, inferential confusion, subthreshold psychotic-like characteristics, and worry. Significance levels were set at a Bonferroni-corrected α value ($p < .05/17 = .003$) due to the number of correlations performed. To compare the bivariate correlation coefficients, effect sizes were calculated as Fisher's *z* coefficients. Power calculations were run for this analysis:

for a medium effect size, 80% power, and significance set at the level described above, the required sample size for bivariate correlations was 152.

The statistical analysis was conducted using the SPSS software version 21.00 and the AMOS software.

RESULTS

Confirmatory factor analysis (CFA)

Before conducting the CFA, the assumption of multivariate normality was investigated by the inspection of kurtosis and skewness indices. An absolute value on these indices falling out of the recommended range between -1 and $+1$, suggests a substantial deviance from normal distribution (Muthén & Kaplan, 1985). Twenty items (items 1, 2, 3, 5, 6, 8, 10, 11, 14, 16, 17, 18, 20, 22, 23, 24, 27, 28, 29, 30) showed kurtosis or skewness values out of the range, suggesting that data for these items were not normally distributed. Thus, the estimation method of Unweighted Least Squares was employed.

Models with two and five correlated factors were tested, as reported in the original validation study (Peters et al., 2014). Finally, a model with a single factor was assessed. The models with a single, two correlated factors or five correlated factors showed good or acceptable fit. As compared with the other models, the model with five correlated factors showed a slightly better fit (GFI = .95; AGFI = .94; NFI = .88; RFI = .87; RMR = .022).

In the single-factor model, nine items (1, 4, 5, 7, 9, 17, 19, 21, 29) had factor loadings lower than .30, in the five-factor model seven items had loadings below this cut-off score (1, 4,

5, 17, 19, 21, 29). In the model with two correlated factors, six items showed loadings lower than .30 (1, 4, 5, 29 for the Threat estimation factor; 17, 21 for the Anomalous perceptions factor). An overview of fit indices and factor loadings for all the three tested models is presented in Table 2 and in Table 3, respectively.

Internal consistency

Internal consistency for the CBQ-p total scores was good (Cronbach's alpha estimate = .83) according to Nunnally and Bernstein (1994). The CBQ-p Threat estimation subscale yielded only a modest internal consistency with a Cronbach's alpha estimate of .65 (range of item-total correlations = .10-.42). The CBQ-p Anomalous perception showed acceptable internal consistency with a Cronbach's alpha value of .75 (range of item-total correlations = .10-.54).

Thus, based on the original validation study (Peters et al., 2014) and the results from internal consistency and CFA, where either the five-, the two- and the single-factor models had good fit, a two-factor model was considered as preferred.

Differences between groups

Results of independent-sample t-tests indicated no difference for gender on the CBQ-p total and the two CBQ-p subscales (range of $t = .80-1.39$, p -values = .18-.42, range of Cohen's $d = .10-.12$). An overview of mean scores in the total group and in the two subgroups divided by gender is presented in Table 4.

Table 2 – Fit indices of the Italian CBQ-p factor models (n = 388)

| Tested models | χ^2 | df | p-value | χ^2/df | GFI | AGFI | NFI | RFI | RMR |
|------------------------------|----------|-----|---------|-------------|-----|------|-----|-----|------|
| One-factor model | 769.56 | 405 | .0001 | 1.90 | .94 | .94 | .87 | .86 | .023 |
| Two correlated factor model | 575.61 | 402 | .0001 | 1.43 | .95 | .94 | .87 | .86 | .023 |
| Five correlated factor model | 671.00 | 395 | .0001 | 1.69 | .95 | .94 | .88 | .87 | .022 |

Note. CBQ-p = Cognitive Biases Questionnaire for psychosis; GFI = Goodness of Fit Index; AGFI = Adjusted Goodness-of-Fit Index; NFI = Bentler-Bonett Normed Fit Index; RFI = Bollen's Relative Fit Index; RMR = Root Mean Squared Residual.

Table 3 – Factor loadings for the three tested models of the CBQ-p (n = 388)

| CBQ-p items | One factor | Two correlated factors | | Five correlated factors | | | | |
|-------------|------------|------------------------|-----|-------------------------|-----|-----|-----|-----|
| | | TE | AP | Int | Cat | DT | JTC | ER |
| Item 1 | .04 | .10 | | .02 | | | | |
| Item 2. | .41 | | .49 | | .42 | | | |
| Item 3. | .57 | | .68 | .55 | | | | |
| Item 4. | .18 | .19 | | | .25 | | | |
| Item 5. | .07 | .10 | | | | .10 | | |
| Item 6. | .33 | | .36 | | | | .39 | |
| Item 7. | .25 | .30 | | | .29 | | | |
| Item 8. | .47 | | .50 | | | | | .49 |
| Item 9. | .25 | .30 | | | | | .33 | |
| Item 10. | .49 | | .52 | | .52 | | | |
| Item 11. | .41 | .44 | | | | .39 | | |
| Item 12. | .37 | .41 | | | .41 | | | |
| Item 13. | .41 | .43 | | | | | | .38 |
| Item 14. | .35 | | .39 | | | .34 | | |
| Item 15. | .44 | .46 | | | | .44 | | |
| Item 16. | .40 | | .43 | | | | | .43 |
| Item 17. | .17 | | .20 | | | | .21 | |
| Item 18. | .51 | .54 | | | | | .55 | |
| Item 19. | .25 | .30 | | | | | | .24 |
| Item 20. | .71 | | .80 | .77 | | | | |
| Item 21. | .06 | | .10 | | | | .18 | |
| Item 22. | .60 | .71 | | .68 | | | | |
| Item 23. | .52 | | .56 | .56 | | | | |
| Item 24. | .47 | .50 | | | | | | .49 |
| Item 25. | .37 | | .41 | | .42 | | | |
| Item 26. | .41 | | .45 | | | | | .42 |
| Item 27. | .42 | | .47 | | | .43 | | |
| Item 28. | .51 | .57 | | .56 | | | | |
| Item 29. | .06 | .10 | | | | | .14 | |
| Item 30. | .38 | | .41 | | | .37 | | |

Note. TE = threatening events; AP = anomalous perceptions; Int = intentionalising; Cat = catastrophising; DT = dichotomous thinking; JTC = jumping to conclusions; ER = emotional reasoning.

Table 4 – Mean scores on the CBQ-p total and subscales in the total group and in the two subgroups divided by sex

| | Total group (n = 388) | Males (n = 175) | Females (n = 213) | $t_{(385)}$ | Cohen's d |
|----------------------------|--------------------------|--------------------|----------------------|-------------|-----------|
| | M (SD) | M (SD) | M (SD) | | |
| CBQ-p total | 46.12 (7.60) | 46.63 (8.06) | 45.71 (7.21) | 1.39 | .12 |
| CBQ-p Threat estimation | 24.15 (3.97) | 24.35 (4.03) | 24.02 (3.92) | .80 | .08 |
| CBQ-p Anomalous perception | 21.94 (4.32) | 22.27 (4.73) | 21.68 (3.96) | 1.33 | .10 |

Note. CBQ-p = Cognitive Biases Questionnaire for psychosis.

Convergent validity with psychotic-like experiences and cognitive factors

Scores on the CBQ-p total moderately correlated with scores on DACOBS belief inflexibility, external attribution bias, subjective cognitive problems, social cognition bias and safety behaviours scales, SPE, ASI, PSWQ and ICQ-EV, weakly with scores on DACOBS attention for threat and jumping to conclusions scales.

Scores on CBQ-p *Threat estimation* and *Anomalous perception* subscales moderately correlated with scores on DACOBS belief inflexibility, external attribution bias, subjective cognitive problems, social cognition bias and safety behaviours scales, SPE, and ICQ-EV, weakly with ASI, DACOBS attention for threat and jumping to conclusions scales. In addition, while scores on CBQ-p *Threat estimation* subscale moderately correlated with PSWQ scores, scores on CBQ-p *Anomalous perception* subscale weakly correlated with PSWQ scores.

CBQ-p total scores had the highest correlations with scores on DACOBS Safety behaviours scores (Fisher's $z = .63$), DACOBS External attribution bias (Fisher's $z = .45$), ICQ-EV (Fisher's $z = .54$), SPE (Fisher's $z = .55$). CBQ-p Threat estimation subscale scores had the highest correlations with scores on DACOBS Safety Behaviours (Fisher's $z = .44$), SPE (Fisher's $z = .47$), ICQ-EV (Fisher's $z = .51$), and PSWQ scores (Fisher's $z = .40$). CBQ-p Anomalous perception scores had the highest correlations with scores on DACOBS Safety behaviours scores (Fisher's $z = .66$), DACOBS External attribution bias (Fisher's $z = .44$), SPE (Fisher's $z = .46$) and ICQ-EV (Fisher's $z = .4$). An overview of bivariate Pearson's correlation coefficients between scores

on CBQ-p total/subscales and those on the other measures is presented in Table 5.

DISCUSSION

Synthesis of results

The current study investigated the psychometric properties of the Italian version of the CBQ-p, a self-report measure designed to assess cognitive biases related to psychotic symptoms and experiences. In the Italian literature, there is a lack of measures assessing these aspects. Current findings expanded previous knowledge on this measure, since a strength of the study was that it investigated the factor structure on adolescents and young adults, who are typically in an age range considered as a key stage for early identification of psychotic-like experiences. Starting from a theoretical model, where psychotic symptoms lie on a continuum (van Os & Linscott, 2012), this study was the first one investigating the relations of the CBQ-p with subthreshold psychotic symptoms and experiences and with other cognitive factors specific to psychosis, such as inferential confusion and aberrant salience.

Confirmatory analyses suggested that five-, two- and single-factor models yielded equally good fit to the data. Despite the model with five correlated factors evidenced a very slightly better fit compared with the other ones, and also the one with a single factor had good fit, the model including two correlated factors was considered as preferred for the Italian CBQ-p on the basis of evidence reported in the original validation study of the measure (Peters et al., 2014), where a

Table 5. – Bivariate Pearson's correlation coefficients (Fisher's z) between CBQ-p total and subscales and the other measures (n = 388)

| | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. | 11. | 12. | 13. | 14. |
|------------------------------------------|----------------|-----------------|----------------|-----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-----------------|----------------|
| 1. CBQ-p Threat estimation | .68** (.83) | .91** (1.51) | .40** (.42) | -.01 (-.01) | .30** (.31) | .20** (.20) | .36** (.38) | .36** (.38) | .39** (.41) | .44** (.47) | .47** (.51) | .51** (.56) | .27** (.28) |
| 2. CBQ-p Anomalous perception | | .92** (1.59) | .21** (.21) | .05 (.05) | .35** (.37) | .18** (.18) | .41** (.44) | .27** (.28) | .35** (.37) | .58** (.66) | .43** (.46) | .37** (.39) | .26** (.27) |
| 3. CBQ-p total | | | .34** (.35) | .03 (.03) | .36** (.38) | .21** (.21) | .42** (.45) | .34** (.35) | .40** (.42) | .56** (.63) | .50** (.55) | .49** (.54) | .29** (.30) |
| 4. PSWQ | | | | -.25** (.26) | .10 (.10) | .34** (.35) | .37** (.39) | .41** (.44) | .50** (.55) | .23** (.23) | .44** (.47) | .51** (.56) | .21** (.21) |
| 5. DACOBS Jumping to conclusions | | | | | .23** (.23) | .14** (.14) | .20** (.20) | -.04 (.04) | -.08 (.08) | 0.02 (.02) | -.10 (.10) | -.09 (-.09) | .0 |
| 6. DACOBS beliefs inflexibility bias | | | | | | .23** (.23) | .51** (.56) | .25** (.26) | .34** (.36) | .41** (.44) | .18 (.18) | .35** (.379) | .18 (.18) |
| 7. DACOBS Attention for threat bias | | | | | | | .33** (.34) | .42** (.45) | .25** (.26) | .24** (.25) | .24** (.25) | .49** (.54) | .32** (.33) |
| 8. DACOBS external attribution bias | | | | | | | | .47** (.51) | .41** (.44) | .40** (.42) | .33** (.34) | .41** (.44) | .15 (.15) |
| 9. DACOBS social cognition problems | | | | | | | | | .51** (.56) | .23** (.23) | .33** (.34) | .64** (.76) | .35** (.36) |
| 10. DACOBS subjective cognitive problems | | | | | | | | | | .31** (.32) | .41** (.44) | .63** (.74) | .35** (.36) |
| 11. DACOBS safety behaviours | | | | | | | | | | | .31** (.32) | .34** (.35) | .17 (.17) |
| 12. SPE | | | | | | | | | | | | .73** (.93) | .52** (.58) |
| 13. ICQ-EV | | | | | | | | | | | | | .59** (.68) |
| 14. ASI | | | | | | | | | | | | | 1 |

Note. ASI = Aberrant Salience Inventory, CBQ-p = Cognitive Biases Questionnaire for psychosis, DACOBS = Davos Assessment of Cognitive Biases Scale, ICQ-EV = Inferential Confusion Questionnaire-Extended Version, PSWQ = Penn State Worry Questionnaire, SPE = Screening for Psychotic Experiences.
** p<.003.

two-factor model was chosen as more reliable. The two-factor model was preferred also based on the factor loadings, which were higher than .30 for all the items except for six items, while nine and seven items had loadings lower than this value in the single- and five-factor models. Evidence that the items 1, 4, 17, 19 had insufficient loadings in all the three models of the Italian CBQ-p appeared quite consistent with evidence of factor loadings found in the English version (Peters et al., 2014), where factor loadings of these items ranged from .35 to .28, then were very close to the cut-off score of .30. Presence of factor loadings lower than .30 on six items in the two-factor solution was recognized as a limitation of the Italian CBQ-p, requiring further investigations in future studies; however, these items were not removed, in light of results from reliability analysis, where Cronbach's alpha estimates remained between .57 and .64 if each of these items were deleted for the *Threat estimation* factor and between .71 and .74 if deleted for the *Anomalous perception* factor.

Indeed, internal consistency appeared good for the CBQ-p total scale ($\alpha = .83$) and was in line with the value observed in the original study, where it was equal to .89 (Peters et al., 2014). According Peters and colleagues' definition (2014), the single factor of the CBQ-p total scale was considered as a general interpretation bias including features related to the two hypothesized cognitive distortions specific to psychotic symptoms and experiences. The two factors were loaded by the same items and were defined using the same labels as in the study of Peters and colleagues (2014): *Threat estimation* and *Anomalous perception* subscales. Regarding internal consistency of the subscales, while the *Threat estimation* subscale showed modest internal consistency, the *Anomalous perception* subscale had acceptable internal consistency.

Another finding was that no gender-related difference was found on cognitive distortions measured by the CBQ-p total and subscale scores. This evidence appeared somewhat in contrast with a commonly found result in the literature, where males typically showed an increased vulnerability for psychotic experiences (Barajas, Ochoa, Obiols & Lalucat-Jo, 2015). Thus, while gender-related differences were observed for some of the specific biases, no difference was found on the CBQ-p total scores suggesting that males and females do not endorse differently a general interpretation bias specific to psychotic-like experiences.

Regarding convergent validity, CBQ-p total and subscales all had significant correlations with measures of cognitive

distortions specific to psychotic symptoms and experiences, aberrant salience, inferential confusion, psychotic-like experiences and worry. CBQ-p total was the scale of the CBQ-p with the highest correlations with subthreshold psychotic experiences, followed by *Threat estimation* and *Anomalous perceptions*. In addition, none of the CBQ-p scales had significant correlations with jumping to conclusions bias measured by the DACOBS. This finding appeared in line with the evidence reported in the original study of the CBQ-p, where this scale had not robust correlations with self-report measures or experimental tasks related to cognitive biases, such as the Dysfunctional Attitudes Scale (Peters et al., 2014). Considering the DACOBS subscales, the two cognitive distortions, measured by the CBQ-p, had the highest correlations with safety behaviours assessed by the DACOBS. Overall, it could be hypothesized that a general interpretation bias measured by the CBQ-p is more specific to psychotic-like experiences (Fisher's $z = .55$) than the different cognitive biases measured by its subscales.

Among the cognitive biases measured by the DACOBS, this general bias appeared more strongly correlated with safety behaviours (Fisher's $z = .63$), and more strongly with inferential confusion (Fisher's $z = .54$) than salience (Fisher's $z = .30$) and worry (Fisher's $z = .35$). On one hand, this finding supported convergent validity of the CBQ-p, as inferential confusion is a cognitive construct associated with psychotic symptoms and experiences (Aardema et al., 2005); on the other hand, it was in contrast with evidence indicating that salience is a cognitive factor specific to psychotic features (Kapur, 2003). The current results, however, confirmed that the general interpretation bias measured by the CBQ-p, was specific to psychotic experiences rather than other constructs such as worry.

Considering the CBQ-p subscales, both *Threat estimation* and *Anomalous perception* had higher correlations with inferential confusion than aberrant salience. This finding could be explained by the fact that young individuals with subthreshold psychotic features, who frequently experience emotional states (e.g., anxiety, negative mood), would expect dangers more likely as consequences of their negative states; this could make certain dangerous stimuli salient leading the individuals to develop a catastrophising reasoning as a result of a vicious cycle. Another important finding was that *Threat estimation* had a higher correlation with worry than *Anomalous perceptions*, confirming that the first factor covers aspects more closely related to thinking (cognitive) processes, focusing on future dangers and negative events

than *Anomalous perceptions*. *Threat estimation* is a common cognitive bias among many of mental symptoms, including not only psychotic features, but also anxiety and depression symptoms (Beck & Clark, 1988). Overall, it could be stated that both the cognitive biases measured by the CBQ-p were related to subthreshold psychotic experiences. Moreover, *Anomalous perception* was more closely related to safety behaviours than *Threat estimation*, suggesting that when the young individual experiences more frequently anomalous perceptions, is more likely to adopt safety behaviours in order to cope with them than adopting them to manage threat estimation.

Limitation and conclusions

Some limitations should be considered. First, the study did not use a clinical group with individuals suffering from psychotic disorders. In addition, the use of non-help-seeking participants prevented to draw firm conclusions about the factor structure of the CBQ-p in help-seeking young individuals at risk of psychosis for subthreshold psychotic symptoms (Yung et al., 1998, 2005, 2006). Despite the use of a screening measure for psychotic experiences, the study did not use a semi-structured interview for subthreshold

psychotic symptoms, such as the Comprehensive of At Risk Mental States (CAARMS; Yung et al., 2005). Future research should evaluate the factor structure of the CBQ-p in young individuals seeking professional help at mental health centres and reporting subthreshold psychotic symptoms measured by the CAAMRS (Yung et al., 2006).

In addition, it could be interesting to compare individuals with subthreshold psychotic symptoms screened through interviews with a group of non-clinical controls, and a group of patients with full psychotic disorders. Another critical point was the internal consistency values for four subscales of the CBQ-p except intentionalising subscale, which resulted poor; this limitation should be further investigated in future research.

Finally, further investigations could assess predictive validity of the CBQ-p examining whether higher scores would predict the onset of a psychotic episode in help-seeking individuals for subthreshold symptoms through long-term evaluations (e.g., one-year follow-up).

In conclusion, the current study expanded knowledge on the cognitive biases specific to psychotic symptoms and experiences in the Italian context, demonstrating that the CBQ-p is a self-report questionnaire with good psychometric properties and convergent validity with constructs related to psychotic symptoms and experiences.

References

- AARDEMA, F., O'CONNOR, K.P., EMMELKAMP, P.M.G., MARCHAND, A. & TODOROV, C. (2005). Inferential confusion and obsessive-compulsive disorder: The inferential confusion questionnaire. *Behaviour Research and Therapy*, 43, 293-308.
- AARDEMA, F., TRIHEY, M., KLEIJER, T.M., O'CONNOR, K. & EMMELKAMP, P.M. (2006). Processes of inference, schizotypal thinking, and obsessive-compulsive behaviour in a normal sample. *Psychological Reports*, 99, 213-220.
- AARDEMA, F., WU, K.D., CAREAU, Y., O'CONNOR, K.P., JULIEN, D. & DENNIE, S. (2010). The Expanded Version of the Inferential Confusion Questionnaire: Further development and validation in clinical and non-clinical samples. *Journal of Psychopathology and Behavioral Assessment*, 32, 448-462.
- AMERICAN PSYCHOLOGICAL ASSOCIATION (2002). Ethical principles of psychologists and code of conduct. *American Psychologist*, 57, 1060-1073.
- BARAJAS, A., OCHOA, S., OBIOLS, J.E. & LALUCAT-JO, L. (2015). Gender differences in individuals at high-risk of psychosis: A comprehensive literature review. *The Scientific World Journal*, 1-13.
- BECK, A.T. & CLARK, D.A. (1988). Anxiety and depression: An information processing perspective. *Anxiety Research*, 1, 23-36.
- BEHLING, O. & LAW, K.S. (2000). *Translating questionnaires and other research instruments: Problems and solutions*. Thousand Oaks, CA: Sage.
- BLACKBURN, I.M., JONES, S. & LEWIN, R.J. (1986). Cognitive style in depression. *British Journal of Clinical Psychology*, 25, 241-251.
- BROOME, M.R., JOHNS, L.C., VALLI, I., WOOLLEY, J.B., TABRAHAM, P., BRETT, C., VALMAGGIA, L., PETERS, E., GARETY, P.A. & MCGUIRE, P.K. (2007). Delusion formation and reasoning biases in those at clinical high risk for psychosis. *The British Journal of Psychiatry*, 191, 38-42.
- CICERO, D.C., KERNS, J.G. & MCCARTHY, D.M. (2010). The Aberrant Salience Inventory: A new measure of psychosis proneness. *Psychological Assessment*, 22, 688-701.

- COMBS, D.R., PENN, D.L., WICHER, M. & WALDHETER, E. (2007). The ambiguous intentions hostility questionnaire (AIHQ): A new measure for evaluating hostile social-cognitive biases in paranoia. *Cognitive Neuropsychiatry*, 12, 128-143.
- DEROGATIS, L.R. (1992). *SCL-90-R: Administration, scoring and procedures manual for the Revised version and other instruments of the psychopathology rating scale series*. Wine Spring Ln, Towson: Clinical Psychometric Research.
- GARETY, P.A., BEBBINGTON, P., FOWLER, D., FREEMAN, D. & KUIPERS, E. (2007). Implications for neurobiological research of cognitive models of psychosis: A theoretical paper. *Psychological medicine*, 37, 1377-1391.
- GARETY, P.A., FREEMAN, D., JOLLEY, S., BEBBINGTON, P.E., KUIJPERS, E., DUNN, G., FOWLER, D.G. & DUDLEY, R. (2005). Reasoning, emotions, and delusional conviction in psychosis. *Journal of Abnormal Psychology*, 114, 373-384.
- KAPUR, S. (2003). Psychosis as a state of aberrant salience: A framework linking biology, phenomenology, and pharmacology in schizophrenia. *American Journal of Psychiatry*, 160, 13-23.
- KELLEHER, I. & CANNON, M. (2011). Psychotic-like experiences in the general population: characterizing a high-risk group for psychosis. *Psychological Medicine*, 41, 1-6.
- KENDLER, K.S., GALLAGHER, T.J., ABELSON, J.M. & KESSLER, R.C. (1996). Lifetime prevalence, demographic risk factors, and diagnostic validity of nonaffective psychosis as assessed in a US community sample: The National Comorbidity Survey. *Archives of General Psychiatry*, 53, 1022-1031.
- LOEWY, R.L., BEARDEN, C.E., JOHNSON, J.K., RAINE, A. & CANNON, T.D. (2005). The prodromal questionnaire (PQ): Preliminary validation of a self-report screening measure for prodromal and psychotic syndromes. *Schizophrenia Research*, 79, 117-125.
- MAGNANI, R., BARCELLA, M., BRONDINO, N., CAPPUCCIATI, M., MARTINELLI, V., ROCCHETTI, M., VECCHIA, L. & POLITI, P. (2010). Prevalenza di sintomi psicotici sotto-soglia in una coorte di studenti della scuola secondaria di secondo grado. *Bollettino della Società Medico Chirurgica di Pavia*, 123, 363-375.
- MEYER, T.J., MILLER, M.L., METZGER, R.L. & BORKOVEC, T.D. (1990). Development and validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy*, 28, 487-495.
- MORANI, S., PRICCI, D. & SANAVIO, E. (1999). Penn State Worry Questionnaire e Worry Domains Questionnaire. Presentazione delle versioni italiane ed analisi della fedeltà. *Psicoterapia Cognitiva e Comportamentale*, 5, 195-209.
- NUNNALLY, J. & BERNSTEIN, I. (1994). *Psychometric theory*. New York: McGraw-Hill.
- OHAYON, M.M. (2000). Prevalence of hallucinations and their pathological associations in the general population. *Psychiatry Research*, 97, 153-164.
- PETERS, E.R., MORITZ, S., SCHWANNAUER, M., WISEMAN, Z., GREENWOOD, K.E., SCOTT, J. & GARETY, P.A. (2014). Cognitive Biases Questionnaire for Psychosis. *Schizophrenia Bulletin*, 40, 300-313.
- SPAUWEN, J., KRABBENDAM, L., LIEB, R., WITTCHEN, H.U. & VAN OS, J. (2003). Sex differences in psychosis: Normal or pathological? *Schizophrenia Research*, 62, 45-49.
- STARTUP, H., FREEMAN, D. & GARETY, P.A. (2007). Persecutory delusions and catastrophic worry in psychosis: Developing the understanding of delusion distress and persistence. *Behaviour Research and Therapy*, 45, 523-537.
- VAN DAEL, F., VERSMISSEN, D., JANSSEN, I., MYIN-GERMEYS, I., VAN OS, J. & KRABBENDAM, L. (2006). Data gathering: Biased in psychosis? *Schizophrenia Bulletin*, 32, 341-351.
- VAN OS, J. & LINSKOTT, R.J. (2012). Introduction: The extended psychosis phenotype-relationship with schizophrenia and with ultrahigh risk status for psychosis. *Schizophrenia Bulletin*, 38, 227-230.
- VAN DER GAAG, M., NIEMAN, D. & VAN DEN BERG, D. (2013). *CBT for those at risk of a first episode psychosis: Evidence-based psychotherapy for people with an 'at Risk Mental State'*. New York: Routledge.
- WEISSMAN, A.M.B. & BECK, A.T. (1978). *Development and Validation of the Dysfunctional Attitudes Scale*. Paper presented at the Annual Meeting of the American Educational Research Association. Toronto, Ontario, Canada.
- YUNG, A.R., PHILLIPS, L.J., MCGORRY, P.D., MCFARLANE, C.A., FRANCEY, S., HARRIGAN, S., PATTON, G.C. & JACKSON, H.J. (1998). Prediction of psychosis: a step towards indicated prevention of schizophrenia. *The British Journal of Psychiatry*, 172, 14-20.
- YUNG, A.R., STANFORD, C., COSGRAVE, E., KILLACKEY, E., PHILLIPS, L., NELSON, B. & MCGORRY, P.D. (2006). Testing the ultra-high risk (prodromal) criteria for the prediction of psychosis in a clinical sample of young people. *Schizophrenia Research*, 84, 57-66.
- YUNG, A.R., YUEN, H.P., MCGORRY, P.D., PHILLIPS, L.J., KELLY, D., DELL'OLIO, M., FRANCEY, S.M., COSGRAVE, E.M., KILLACKEY, E., STANFORD, C., GODFREY, C. & BUCKBY, J. (2005). Mapping the onset of psychosis: The Comprehensive Assessment of at Risk Mental States (CAARMS). *Australian and New Zealand Journal of Psychiatry*, 39, 964-971.

Raven's Standard Progressive Matrices: Contribution to Italian standardization for subjects between ages 6 and 18

Laura Picone, Arturo Orsini, Lina Pezzuti

Department of Dynamic and Clinical Psychology,
Sapienza University of Rome

• **ABSTRACT.** Il lavoro presenta la taratura delle *Standard Progressive Matrices* di Raven su un campione italiano di 5438 ragazzi dai 6 ai 18 anni. Lo strumento è molto conosciuto e datato ma tuttora utile per una rapida valutazione delle abilità cognitive, confermata nella letteratura internazionale dal confronto con i risultati di altri test che misurano lo stesso costrutto. Il presente lavoro di taratura ha confermato l'aumento delle prestazioni dei ragazzi al test con il crescere dell'età ma anche l'influenza sulle stesse di variabili sociali, come la scolarità dei genitori presa come indicatore del livello socio-economico familiare. I punteggi grezzi ottenuti dal campione sono stati trasformati in punti standard per età (QI con $M = 100$ e $DS = 15$); sono stati calcolati l'attendibilità e l'errore standard di misura; tali parametri sono riportati nel Report di Giunti Testing e servono a dare una corretta interpretazione dei punteggi ottenuti.

• **SUMMARY.** This paper presents a standardization study of Raven's Standard Progressive Matrices on an Italian sample of 5438 young people aged 6 to 18 years. This instrument is well known, and although it is dated it is still useful for fast assessment of cognitive abilities, as confirmed in international literature a comparison of results from other tests which measure the same construct. The study confirms that young people performance is related with age, and it is influenced by social variables, such as parents' schooling level used as an indicator of social-economic level. The raw scores obtained by the sample were converted in standardized points for age (IQ with $M = 100$ and $SD = 15$). Reliability and standard measurement error were calculated. These parameters are included in the report by Giunti Testing; their purpose is to supply a correct interpretation of the scores obtained.

Keywords: Intelligence, Raven's Standard Progressive Matrices, Youth, Adolescence, Cognitive Function

INTRODUCTION

The first intelligence tests were created and spreaded at the beginning of the 1900s (Binet-Simon Scale, Binet and Simon, 1908; Stanford-Binet Intelligence Scales, Terman, 1916). They were criticized, however, because, being based primarily on verbal tests, they tended to penalize subjects belonging to culturally disadvantaged contexts. The first nonverbal tests were created by Otis (1936-1939), Wechsler (1939), and Raven (1938). In particular, the first version of Raven's Progressive Matrices (1938) was highly appreciated, coming at a time when the effort was to capture the essence of intelligence as independent of the knowledge which an individual has from schooling and from interaction with the surrounding environment.

John Carlyle Raven (1902-1970) had an idea for how to probe the question of intelligence without relying on verbal stimuli. He had been a student of Spearman's, the author of an intelligence theory (1927) which hypothesized the presence of a "general" and a "specific" factor. Together with the geneticist Penrose, Raven used Spearman's stimuli again; they did not, however, ask subjects to identify solutions verbally, but rather to find visually the solution which would fill in the incomplete abstract figures presented. Specifically, subjects had to identify a recurring pattern among the figure-stimuli based on an "inductive reasoning" method which apparently did not depend on scholastic or environmental education or knowledge.

Raven's Matrices are considered to be the best measure of Spearman's *g* factor, on a genetic and hereditary basis (Kaplan & Saccuzzo, 1997; Raven, 1938; Raven, Raven & Court, 1998).

The first version of the test, the *Standard Progressive Matrices* (SPM), was devised by Raven to measure cognitive capacity in children, adolescents, adults, and the elderly, even though national and international standardization data for this last category is more scanty.

Raven standardized them on a sample of Scottish young people (1407 subjects) aged 6 to 14 (1938, 1940). In 1944, Raven and Walshaw published another work on a sample with equivalent ages, from Colchester, above all to test correlations between the SPM and a vocabulary test (*Mill Hill Vocabulary Scale - MHV*). They found that the norms were 2 points lower than those from 1938. Raven attributed this datum to the restrictions and limitations the war-time period imposed, and this lead him to the conclusion that the test results were sensitive to environmental conditions and changes, both in

terms of nutritional deprivation and in terms of a reduction in significant stimuli to which the young people were exposed.

Raven's son, John Raven Junior, got involved with the studies and carried them forward after his father's death. In 1972, with his two brothers, he established a company to manage the Matrices' dissemination in countries around the world.

Raven's SPM have been highly successful throughout the world, above all in English-speaking countries. Many researchers have done studies to verify whether the British norms can be used as a reference standard. Here below we cite some of the most important ones. Adams (1952) records the norms of 11621 young people aged 12 from Surrey which, with the limits of measurement error, are very similar to those gathered for the Scottish young people in 1938. Tuddenham, Davis, Davison, and Schindler (1958) tested various classes of children in California, and concluded that the use of these normative data is acceptable. Byrt and Gill (1973), working with Raven, gathered a representative sample of children aged 5-11 (3464) in the Republic of Ireland; the results for city children were comparable to the 1938 standards, even if those for rural-based children were slightly lower. Kratzmeier and Horn (1979) reported standards from a broad study done in Germany, with results well above those obtained in England in 1938. As mentioned, this lead John Raven Junior to develop a new British standardization (1979, published in 1981) on a sample of 3569 youth aged 6 to 16, where greater attention was paid to the representativeness of the sample.

In the United States 50 standards studies have been done (between 1983 and 1989, reported in J. Raven et al, 1990/2000) on a population drawn from each school district. Approximately 60000 students aged 5 to 18 were tested. It was shown that American standards for the white population are similar to the British standardization, but some ethnic groups obtain lower average scores.

Other authors report similar results to the English norms of 1981: Holmes (1980) in British Columbia (Canada); de Lemos in New Zealand and in Australia (1984, 1989); Abdel-Khalek and Raven in Kuwait (2006). Zhang and Wang (1989) in China, gathering data on an urban and rural population, found similarities with the Chan standards (1981, 1989) for Hong Kong, which corresponded closely with those obtained elsewhere. The same result is to be found in Poland (Jaworowska & Szustrowa, 1991), in Spain (Raven, Court & Raven, 1995); and in Switzerland (Martinolli, 1990).

To conclude, research appears to confirm that the

reference standards found in other countries are similar to those from the last British standardization. Furthermore, studies also seem to confirm that test results are connected to social-cultural characteristics of the social group in which the subjects live, given that results from populations in privileged environments generally seem better than those from less-privileged ones. This point revives the old debate regarding the weight and importance of hereditariness or environment on the development of mental abilities. It appears that improvement (or worsening) of a population's living conditions may induce a positive (or negative) change in the development of its members' intellectual abilities.

In recent works, Raven Jr. (2000, 2008) discusses the stability and the variations in SPM standards among cultural, ethnic, and social-economic groups over the last 70 years, which were also revealed by other researchers. The comparison between the new British standards and the original ones highlights a score increase over time for the Standard Progressive Matrices.

Flynn's studies raised a lot of interest in this regard. An early study of his (1984), starting with an analysis of SPM results obtained from various cohorts of military recruits, revealed a consistent increase in test scores over the years. Later, in 1987, he compared results on a number of intelligence tests (in addition to the Standard Progressive Matrices, also the Wechsler scales) from 14 countries in Europe and America (including Raven's studies on 30000 students aged 5 to 18) with his studies did in Australia on a population of various ages, including children and adults. Flynn found that score increases over time were to be found above all in standards studies of the Standard Progressive Matrices and not in other tests, quantifying this increase to be from 5 to 25 IQ points every 10 years, with differences among the various countries. In any case, the extent of the increase varies among the studies of various authors. Some of them showed an increase of almost 7 points per decade in SPM scores (e.g. Flynn, 1998). This increase, which has since been called "the Flynn Effect", is thought to be due to environmental factors such as improved nutrition, living conditions, and stimuli among which could be included the advent and use of technologies which broaden the possibilities available to individuals for accessing and actively seeking out knowledge and information (such as the television and computer).

In Italy, too, the SPM are rather well-spread, but with the use of standardizations on smaller samples which have not always been representative of the Italian population.

Ferracuti and Groppelli (1954) and Boschi (1960) tested youth attending middle school (596 subjects between ages 11 and 13) and vocational-technical school (263 subjects between ages 12 and 14). The authors obtained the same results. Whereas gender showed no incidence, they did find significant differences for age and for the two types of school, with higher results among the youth from middle school who generally also belonged to a higher social-cultural class. They concluded, therefore, that the test's result is influenced by cultural variables. Other Italian studies highlight the effect one's level of schooling has on test performance. For example, Di Fiore and Renda (1968) (with a sample of 1560 youth aged 14 to 20) and Reda, Nencini and Riccio (1955) (with a sample of 400 subjects aged 16 to 29), observed higher performance among university graduates.

An Italian standardization study was published by Valseschini and Del Ton in 1973 on a sample of 1123 (990 males and 133 females) with ages ranging from 11 to 60, of which 857 subjects were between 11 and 20 years of age; their level of schooling went from illiterate to high school studies. In general the authors underline the absence of either a gender or age effect (the latter ostensibly being due to the sampling), and the presence of a schooling-level effect on the SPM. Starting with the group aged 31-40 years, a decrease in elderly people's performance is seen, which is mitigated by effect of schooling. Nevertheless, it should be noted that their Italian standardization sample, while being quite numerous, covered a broad age range, and the numbers in the various age ranges were not numerically balanced. Furthermore, subject distribution for gender was uneven, favoring the males.

One more recent Italian study (Giunti O.S., 2008) supplies reference standards for 825 non-clinical subjects aged 11 to 14, who go to middle or high school (defined as a non-clinical group) and 170 subjects aged between 10 and 84 years, defined as a clinical group given that they have various pathologies (e.g. trisomy 21, Alzheimer's). A standardization was calculated through division into percentiles for each of the two groups, leaving aside the fact that each clinical subject should always be compared to the standard. The clinical sample could be used only comparing it with the non-clinical sample and supplying the average performance of the two samples, after having verified that the two samples were paired at least for the most important variables like age, gender, and schooling.

The main results in the non-clinical group (11-14 years) are: an increase in scores compared with the 1973

standardization, just as was seen at the international level; the perception that the SPM are easy, based on the difficulty index calculated on the test's 5 series (naturally, the items which are perceived to be difficult are the last ones in each series and the last series); the discrimination indexes for satisfactory items; the calculated reliability looking at internal consistency (equal to .90 with the Kuder-Richardson coefficient calculator); there are no significant differences for gender except in series D.

The validity of Raven's matrices as an instrument for the measurement of mental abilities has been studied through comparison with the results of other tests with the same construct and has been confirmed by a substantial number of studies. The first data were summarized by Burke (1958) and then reported by Raven, Court and Raven (1983; 1986; 1992), and by Raven and Summers (1986). Burke found that the correlation between the SPM and the Wechsler scales (which provide a mental efficiency index, FSIQ) grows among elderly subjects, while the validity coefficient with development tests for children varies from .30 to .60. In the Anglo-Saxon studies, these values are lower than those found when correlating the SPM with the Wechsler and the Binet Scales, administered immediately before or after the SPM (they vary from coefficients of $r = .54$ to $r = .86$; Raven, Court and Raven, 1978). The correlation of the SPM and these instruments tends to be higher with the performance subtests than with the verbal ones. In a large study involving thousands of students, Saccuzzo and Johnson (1995) conclude that the SPM and the WISC-R have approximately the same predictive validity in regard to criteria such as school performance, and that there is no different validities among 8 different ethnic groups were found.

Factor analysis studies concerning the SPM have highlighted the presence of a *general* factor, the relevance of which is more or less important depending on the study, and a saturation in *s* (the ability to visualize spatial relationships) has also been found.

Van der Ven and Ellis (2000), in a study aimed at identifying talented youth, administering the SPM form, used the Rasch model to investigate the items' one-dimensionality. In items in the A, C, and D series one-dimensionality was confirmed, whereas in the B and E series the items reveal two different dimensions.

With a factor analysis study (on 2735 young people aged 12 to 18 in Estonia) Lynn, Allike, and Irwing (2004) explore and confirm the presence of a secondary factor *g*, but they

also come across the presence of 3 other factors: Gestalt continuation (which had already been found by Van der Ven & Ellis, 2000), verbal-analytical reasoning, and visual-spatial ability.

Also Mackintosh and Bennett (2005) find (among 97 college students aged 17/18 years) that the *g*-factor is mainly present in the SPM, but the easiest items measure a perception or Gestalt factor which is distinct from an analytic factor in the rest of the test.

In an Italian study by Picone (1996), with 2000 subjects aged 14 to 19 years, factor analysis on concrete and formal Piagetian tests and on the SPM shows that the latter are saturated for the factor which includes the formal tests (.54) and for those including concrete tests (.44), as further proof for the fact that the kind of reasoning found in the Matrices can be considered a general intelligence factor.

The aim of this paper is to describe the fundamental data which were used for a new Italian standardization of the SPM using a sample of 5438 subjects ranging from 6 to 18 years of age. Some validation studies done on the sample are also presented.

METHOD

Sample

The Italian standardization sample is made up of 5438 Italian subjects (2410 males and 3028 females) aged 6 to 18 years, devoid of any evident psycho-physical handicap. Table 1 reports subject distribution according to 13 age levels, of one year each (e.g. from 6 years 0 months and 0 days to 6 years 11 months and 30 days), and according to gender.

In addition, subjects were classified in three groups according to their father's years of schooling: 2538 subjects had fathers who did up to 8 years of schooling; 2235 subjects' fathers did between 9 and 13 years; and 665 subjects' fathers did more than 13 years.

For all subjects, Raven's Standard Progressive Matrices were administered collectively and with no time limit. For the younger children, administration was done in small groups.

Subjects for the sample were recruited in various regions of Italy as part of work on university theses on the intelligence efficiency of children and younger people. Subject groups were given other tests of similar construct, along with the SPM: Human figure drawing for cognitive assessment,

Table 1 – Study sample configuration by age and gender

| Age | Gender | | Total |
|------------|--------|------|-------|
| | F | M | |
| 6.0-6.11 | 141 | 138 | 279 |
| 7.0-7.11 | 175 | 163 | 338 |
| 8.0-8.11 | 224 | 188 | 412 |
| 9.0-9.11 | 186 | 220 | 406 |
| 10.0-10.11 | 163 | 190 | 353 |
| 11.0-11.11 | 145 | 141 | 286 |
| 12.0-12.11 | 84 | 74 | 158 |
| 13.0-13.11 | 197 | 173 | 370 |
| 14.0-14.11 | 285 | 208 | 493 |
| 15.0-15.11 | 676 | 438 | 1114 |
| 16.0-16.11 | 246 | 181 | 427 |
| 17.0-17.11 | 206 | 134 | 340 |
| 18.0-18.11 | 300 | 162 | 462 |
| 6.0-18.11 | 3028 | 2410 | 5438 |

Piagetian tests for concrete and formal thinking, aptitude tests for school orientation.

The area from which data was collected can be defined as central-southern Italy: approximately half of the subjects came from Lazio (47%), but Campania (13%) and Sardinia (15%) are also well-represented.

Instrument

– *Description of the SPM.* The Standard Progressive Matrices (SPM) were the first form developed by Raven, in 1938, and for this reason are also known as PM38.

The materials needed for administration of the test are made up of a test booklet, an answer sheet with a grid on it, and a pencil (so as to avoid having permanent marks which could reveal correct answers in the booklet).

The exercise which subjects take on concerns solving logic problems, using graphical stimuli which are organized in

matrices (2 x 2 or 3 x 3) made up of 60 items; these are divided in five series (A, B, C, D, E), increasing in difficulty within each series and from one series to the next.

The subject is asked to identify the figure which completes the item from among the alternatives. To do this, the subject has to observe the characteristics of the figures in the matrix both horizontally and vertically, as well as their relationship one to the other, given that the figures include graphical patterns which change from left to right, following some specific rationale, and from top to bottom, following some other rationale.

In general, the test requires that a series of concepts found in a set of problems of visual analogies be analyzed, constructed, and integrated with one another. Each entry (cell) contains one of the figurative elements, various kinds of shapes, various kinds of lines, or background patterns. Incorrect answers come from an unclear perception of the constant elements and the modified ones in the test figures. The younger children tend to look for equivalencies

and identify the correct answer in alternatives which have elements that are the same as those in the stimulus. Children aged 8 to 10 can keep only one peculiarity of the figure in mind, but not both at the same time. Older children and youth (between ages 11 and 18) can consider both of the abstract figure's characteristics which change, and so get to the correct answer. Examining these observations in light of Piaget's theory of cognitive development, one can identify a typical mental process of the *preoperational* stage in the first group; the period for the *concrete operational* stage in the second group; and the presence of logical processes belonging to the *formal operational* stage in the third group. Using these reference parameters, an in-depth qualitative analysis could be done on an individual's performance.

- *SPM Administration Process*. The SPM can be administered collectively or individually, with no time limit. For younger children and for the elderly, individual or small-group administration is recommended, so as to check that basic instructions are followed well, both for the task at hand and for the response procedures.

The examiner, in the case of collective testing, will give the following instructions:

"Open your booklet to the first page: you will find a figure like this one"; at the same time a booklet open to the first page is shown to the whole group, while adding:

"At the top, on this page, it reads: Series A, and in the same way, on your answer sheet, you have a column labeled "A". The problem I am showing you now is problem A1. This means it is the first problem of Series A. Notice what it looks like: the upper part is a figure which is missing a piece [the white space]. Each of the pieces which you find drawn in the lower part of the page [show them] has the correct shape for fitting into this white space, but these pieces cannot complete all of the figure. Number 1 [point to and show the figure to be completed also] provides a completely wrong answer. Numbers 2 and 3 are also not suitable; they fit in the space well, but they do not complete the figure. What do you think, on the other hand, of number 6? It has the correct shape [demonstrate that its shape is identical to that of the preceding numbers] but it does not cover the whole shape. Show me the correct piece (One makes sure that the test participants have found the correct answer. If necessary, supply further explanation, and proceed). Yes, this is number 4. So the answer to problem A1 is 4. Therefore, write 4 next to the number 1 in column A on your answer sheet. Do not turn the page yet".

The examiner waits until everyone has finished, and then continues:

"On each page of your booklet, there is a figure which is missing a piece; each time, you have to choose from among the pieces in the lower part, the one which will correctly complete the figure of the upper part. When you have found it, write the corresponding number on your answer sheet, next to the number which indicates the problem's page number. The problems will be easy at the beginning, but they will get steadily more difficult. There are no tricks. If you pay close attention to the way to solve the simple problems, you will find the following ones less difficult.

Solve each problem one after the other, without skipping any of them; start, therefore, at the beginning and continue without pausing until the end. Work at your normal pace. You will not get left behind; you have all the time you need. Now, please turn the page and begin the next problem".

When all the subjects have had time to write their answer for item A2, the examiner will say:

"The correct answer is obviously number 5. Check if you have all written the number 5 next to the number 2 in column A on your answer sheet. Continue on your own now, until the end of the booklet".

At this point, the instructions are complete and the subjects are left free to carry out their work without being pushed in any way, and without providing any further explanations.

- *Test duration*. Subjects are allowed to have all the time they need to complete all the test items. Usually, the average completion time for the SPM is approximately 20 minutes. It is good practice to note the completion time because it can supply important additional information. If the subject terminates in 10 minutes or less, his/her speed can be interpreted as an expression of rapid and concise thought processes, when his/her results are average or higher for his/her age; or it can be seen as an expression of anxiety or of an oppositional attitude to the test, when the results are less than average for his/her age. If the subject finishes after more than 45 minutes, there are two interpretations. If the performance is appropriate to his/her age, the subject may be one whose intelligence is of an analytical nature, which needs processing time to reach the correct answer (from an analysis of the handwriting in the response, should there be marks gone over multiple times or dotted or unsteady ones, one could hypothesize anxious or obsessive traits). If the results are below the average for his/her age and the answers were often erased, corrected, or missing, the

subject may have high performance anxiety and confused or contradictory thinking.

Finally, as far as the choice of when to interrupt testing is concerned, after 45 minutes the subjects are invited to give an answer anyway, and the answer sheets are collected.

- *Data analysis.* SPM's reliability in the Italian standardization sample was measured with the split-half method (correlation between the sums of the even scores and the odd ones) corrected for the length of the test with the Spearman-Brown prophecy formula.

ANOVA were carried out to study the effects of age and gender on the SPM raw scores.

For each of the 13 age range (from 6 to 18 years) of the Italian standardization sample ($N = 5438$), the SPM raw scores were transformed in standard equivalent points (IQ) with $M = 100$ and $SD = 15$, using the continuous norming approach proposed by Gorsuch (Gorsuch, 1983; Zachary & Gorsuch, 1985). This method requires regression analysis procedures which take into account the effect of age on both averages and standard deviations of the raw score distributions. These standards are those referred to by the Giunti Testing report, SPM-RA 6-18 (Standard Progressive Matrices - Report Aggiornato 6-18).

RESULTS

Reliability and standard measurement errors of the SPM

In Table 2, reliability coefficients (r_{tt}) and standard measurement errors (SME) are recorded for each age group, calculated for the SPM raw scores, for each of the ages, with their relative average coefficients. Reliability values vary between .85 (group of 18 year-olds) and .95 (group of 7 year-olds), with an average reliability of .91, which can be considered

excellent reliability values. The standard measurement errors tend to increase with age. This means that, contrary to what one might think, the performance of the younger children is more homogenous than that of the oldest young, for which the variability of the "real" score is higher.

Analysis of the effects of age and gender variables on SPM raw scores

First a variation analysis was done with age (13 levels) and gender (2 levels) as independent variables, and raw score obtained on the SPM as the dependent variable. The ANOVA revealed significant differences between the age groups ($F_{12,5412} = 539.20$; $p < .001$; $\eta^2 = .54$) but not between the two genders ($F_{1,5412} = .366$; $p = .545$; $\eta^2 = .00$), nor does a significant interaction effect between the two variables emerge ($F_{12,5412} = 1.732$; $p = .054$; $\eta^2 = .00$); related to the age effect, the post-hoc comparisons (Bonferroni method) between the age groups are generally significant ($p < .001$), apart from some contiguous age groups (11-12; 12-13; 13-14; 15-16; 17-18).

The trend of the series' average raw scores is the expected one: from the easiest series to the hardest one, results increase with age. Therefore, calculating the percentages of correct responses for age in the various series was possible; they will be taken into consideration in the report.

Analysis of the item difficulty index

The SPM have always been described as being made up of items which are presented in order of difficulty within the series and as being made up of series which get steadily more difficult. However, no one has ever experimentally proven this phenomenon, which has been thought to be true solely

Table 2 – Reliability coefficients and Standard Measurement Errors per age group

| Age | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | r_{tt} and average SME |
|----------|------|------|------|------|------|------|------|------|------|------|------|------|------|--------------------------|
| r_{tt} | .91 | .95 | .91 | .93 | .92 | .92 | .86 | .91 | .88 | .89 | .90 | .90 | .85 | .91 |
| SME | 3.82 | 3.92 | 4.60 | 4.55 | 4.31 | 4.16 | 4.98 | 4.28 | 4.95 | 5.01 | 5.13 | 5.22 | 5.80 | 4.70 |

on the basis of the items' content.

The difficulty index has been studied with this sample for Italian standardization, to supply a topic for consideration regarding the single items in each series. In specific, analyzing the percentages of correct answers to the 60 items as a function of age, some irregularities appear. Some items in certain ages are more difficult than items which follow them (e.g. items A8 and A9 at age 6), but those which reveal this irregular trend at all ages are items A11, B8, C4, C8, E8. We could hypothesize, therefore, that when subjects correctly solve one or more of these items (which reveal a low resolving percentage when compared with the items which follow them), this could be interpreted as one of their personal abilities or a strength among their cognitive abilities.

The effect of a social-cultural variable on IQ and the SPM

To study the effect of paternal schooling as an indicator of a family's social-cultural level, on full IQ and the SPM, subjects were divided into 3 levels based on the father's level of schooling (up to 8 years, between 9 and 13 years, and more than 13 years) and an ANOVA was done from which an effect emerges, even if not large, for paternal schooling ($F_{2,5435} = 71.218$; $p < .001$; $\eta^2 = .03$). Average performance results increase as the level of paternal schooling increases, up to a difference of approximately 7 IQ points between the lower level of schooling and the upper one, while still falling in the average level of interpretation. These values make up one of the standardizing variables in the Giunti Testing SPM report.

Analyzing the trend in the differences among the SPM IQ averages for the subject's age at the time of SPM testing and paternal schooling (see for example Table 3), a linear relationship can be seen between test results increase and paternal schooling increase from 6 years of age, reaching as many as 13 IQ points' difference at age 14. Such differences

then tend to decrease after 14 years of age, almost to the point of cancelling themselves out at age 18. Evidently, over time youth's school attendance makes these differences which come from the father's level of schooling less accentuated.

These results confirm what has been found in studies reported in the literature: test performance is correlated to the cultural level of the subject's family (represented in specific by the father's level of instruction) and so it cannot be affirmed that the SPM are a culture-free measure of cognitive capacity. This must be kept in mind when interpreting the results.

Interpretive report. Referring to the SPM-RA 6-18 report by Giunti Testing, after having inserted subjects' responses online, the administrator of the SPM, obtains:

- the percentages of correct answers for each series and for the total, of the test taker and of his/her age group;
- the test taker's total raw score with the corresponding percentile, IQ score, and IQ confidence intervals at 90% and 95%;
- interpretation of the IQ score and confidence interval at 95%;
- interpretation of the test taker's IQ score as a function of the father's years of schooling (if the related personal information field was filled in);
- a general interpretive comment.

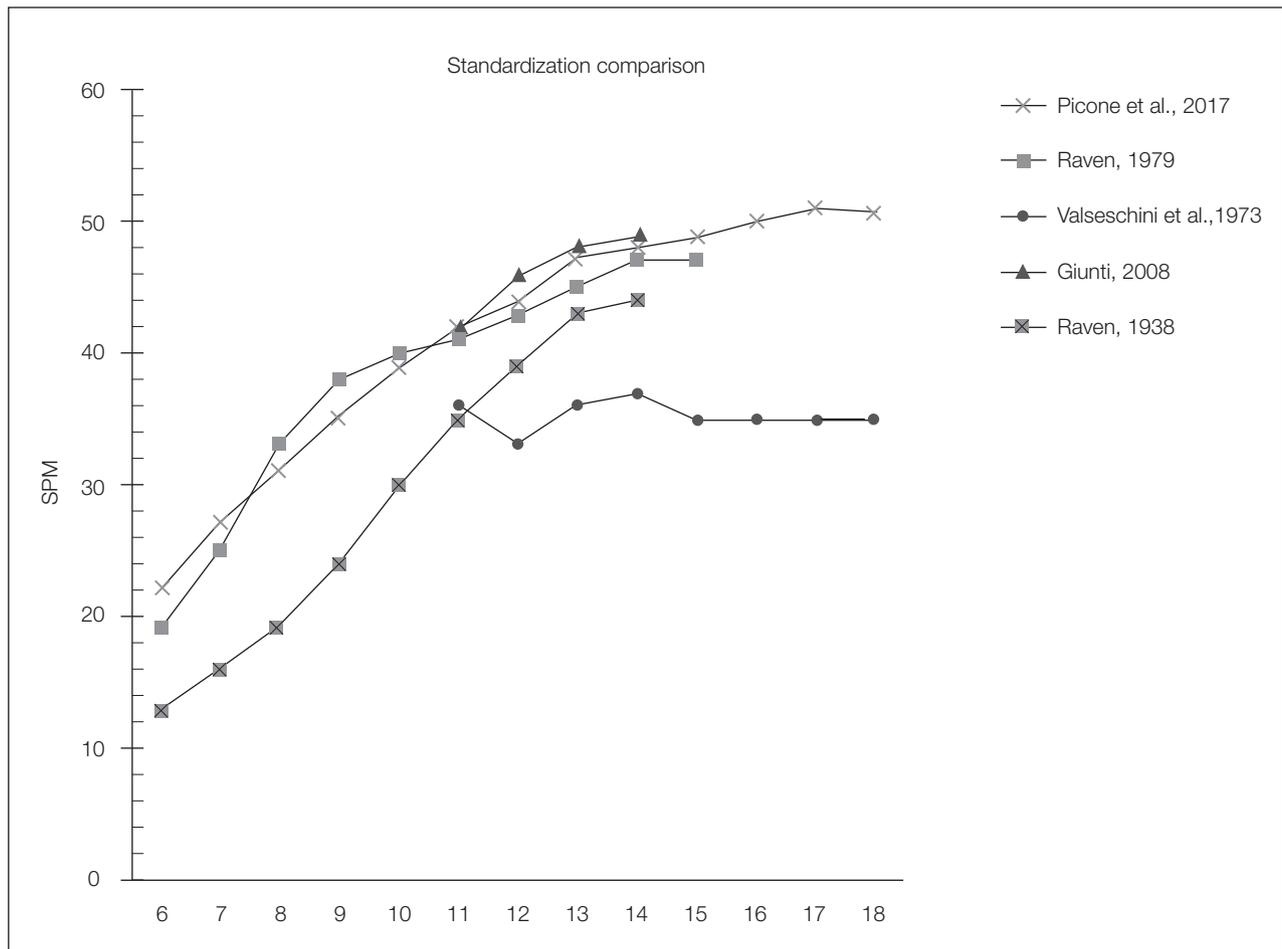
Comparison of the English and Italian calibrations

In figure 1 raw score trends at the 50th percentile for each age group are reported, for both the English standardization and the earlier Italian one, to compare them graphically with the current Italian one. The data used for the comparison are taken from the following studies:

- Raven (1938), on 1407 Scottish youth aged 6 to 14;
- Raven (1979), on 3569 Scottish youth aged 6 to 16;
- Valseschini and Del Ton (1973), on 857 subjects aged 11 to 20;

Table 3 – Maximum differences between IQ for subject's age and father's schooling level

| Age | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|---------------------------------|---|---|---|---|----|----|----|----|----|----|----|----|----|
| Maximum difference in IQ points | 4 | 4 | 6 | 7 | 4 | 6 | 7 | 10 | 13 | 5 | 5 | 5 | 1 |

Figure 1 – Comparison of the English and Italian standardizations

– Giunti O.S. (2008), on 825 subjects aged 11 to 14.

As can be seen, the trend our data follows is similar to that of Raven (1979), but, as to that study, our data highlights higher scores starting with age 11. In the most recent Italian study (Giunti O.S., 2008), on a group of young aged 11 to 14, comparable results to ours are found.

The earlier Italian standardization by Valseschini et al, in 1973, beyond offering generally inferior standardization references, included a limited age variation, whereas in the international literature this factor is identified as being the one with the most relevant effect on test performance.

In line with the international literature reports regarding score increase over time for this test according to the so-called “Flynn effect”, the differences between the medians found in the first English standardization (Raven, 1938), the subsequent ones (Raven, 1979; Giunti O.S., 2008), and the current Italian standardization are evident in the graph.

CONCLUSIONS

Research which has been published to date underlines how performance obtained on the SPM are correlated with all the cognitive abilities tests, and how the SPM are still used as the best test of abstract, non-verbal reasoning. This ability makes up the essence of the “fluid intelligence” factor, namely the ability to solve logic problems, which stands in contrast to “crystallized intelligence”, that being intelligence which uses knowledge people acquire from environmental stimulation and scholastic learning.

Moreover, the kind of fluid intelligence which would be measured displays precisely the problem-solving abilities which are not only present but also potential. It is possible that an individual not be aware of his/her own cognitive abilities, which may not necessarily be expressed in a particularly brilliant scholastic performance. However, if a person is

understood and stimulated, he/she may with time be able to use his/her abilities to progress and to reach important objectives and goals.

The predictive validity regarding scholastic progress at a distance of one or two years is good. Therefore, the SPM can be considered an effective test for predicting the likelihood of scholastic learning, in the absence of serious personality disorders. In this sense, they have been used to predict school success and to select subjects who are well-suited to undertaking university studies, regardless of the disadvantaged environment from which they came.

On the contrary, if individuals do not encounter those people or those things which are able to supply them with adequate motivation, they preserve abilities which remain silent, and bit by bit lose the propulsive thrust which could have lead them to grow and to establish themselves in any context. They are thus flattened in resigned adaptation to prevalent social models and values in their surroundings (above all if these are negative, such as the excessive importance given to consumerism, to establishing a social role, and to scarce interest in knowledge and culture). Thus the importance of the role of attentive, well-prepared educators (be they the parents themselves, teachers, or other adults) who are able to stimulate the energy and abilities present in young people is confirmed.

As to the contrast between hereditary and environmental factors which impact the development of cognitive abilities, the SPM have been considered a culture-free test and, therefore, well-suited to highlighting these abilities regardless of the characteristics of the environment in which the individuals live. The presence of non-verbal stimuli and the limited need for verbal instructions are special features which characterize the SPM as an appropriate test to be administered to subjects who come from others cultures. Currently, the preference is to consider them to be "culture fair" tests, that is "unbiased for all cultures"; they are seen as tests which all individuals can take on, while still being sensible to cultural factors. Research in this area with data gathered from disadvantaged countries compared with more

advanced ones which provide greater resources for education, highlights the differences between performance on cognitive ability tests done with urban or rural populations, which have or have not had formal instruction. To close the discussion of the conflict between inheritance and environment, one can consider a review of 30 years of research on ethnic differences in cognitive abilities in which Rushton and Jensen (2005) hypothesize, in agreement with contemporary social sciences, that all children are born with potentially similar intellectual and learning abilities. The inequalities which can be found among various groups are the result of social, economic, and political factors. This world view has generated many strategies for intervention in the family, work, mass media, and criminal justice systems, to the point of including the entire social-economic system. We could think, therefore, that improvement (or worsening) of a population's living conditions can induce a positive (or negative) change in its members' performance on tests for the development of intelligence abilities.

To conclude, the SPM can be used:

- for a quick screening, done collectively, of individuals' level of cognitive efficiency;
- for an assessment of cognitive efficiency in subjects with hearing, language, and motor disabilities;
- for education or professional guidance recommendations (done together with a battery of specific tests);
- regardless of language comprehension problems (e.g., with foreign subjects);
- individually for subjects with comprehension and verbal production deficits or with suspected attention deficits (e.g., subjects diagnosed with ADHD);
- also for subjects with serious personality disorders for whom administration of a tool like the SPM may be useful because it uses abstract and "affectively neutral" stimuli, compared with other cognitive tests;
- for adult and elderly subjects as an indicator of neurophysiological deterioration and as a predictor of degenerative processes (with Alzheimer's, for example).

References

- ABDEL-KHALEK, A.M. & RAVEN, J. (2006). Normative data from the standardization of Raven's Standard Progressive Matrices in Kuwait in an international contest. *Social Behavior and Personality*, 34 (2), 169-180.
- ADAMS, E.A. (1952). *Analysis of Raven's Matrices scores: Preliminary report*. Surrey, England: Surrey Educational Research Association.
- BINET, A. & SIMON, T. (1908). Le développement de l'intelligence chez l'enfant. *L'Année psychologique*, 14, 1-94.
- BOSCHI, F. (1960). Contributo alla taratura del reattivo PM 38 di J.C. Raven. *Bollettino di Psicologia Applicata*, 37, 173-186.
- BURKE, H.R. (1958). Raven's Progressive Matrices: A review and Critical Evaluation. *Journal of Genetic Psychology*, 93, 199-228.
- BYRT, E. & GILL, P.E. (1973). *Standardisation of Raven's Standard Progressive Matrices and Mill Hill Vocabulary for the Irish population: Ages 6-12*. National University of Ireland, University College Cork.
- CHAN, J. (1981). Correlates of parent-child interaction and certain psychological variables among adolescents in Hong Kong. In J.L.M. Binnie-Dawson (Ed.), *Perspectives in Asian cross-cultural psychology* (pp. 117-131). Lisse, Netherlands: Swets and Zeitlinger.
- CHAN, J. (1989). The use of Raven's Progressive Matrices in Hong Kong: A critical review. *Psychological Test Bulletin*, 2 (2), 40-45.
- de LEMOS, M.M. (1984). A note on the Australian norms for the Standard Progressive Matrices. *Bulletin for Psychologists (ACER Melbourne)*, 36, 9-12.
- de LEMOS, M.M. (1989). The Australian re-standardisation of the Standard Progressive Matrices. *Psychological Test Bulletin*, 2 (2), 17-24.
- DI FIORE, E. & RENDA, S. (1968). Studio sulle applicabilità di alcuni test in orientamento professionale. *Securitas*, 9-10, 193-208.
- FERRACUTI, F. & GROPELLI, A. (1954). Contributo preliminare alla taratura del test "Progressive Matrices 1938" di Raven su soggetti di 11, 12, 13 anni di età. *Infanzia Anormale*, 9, 740-750.
- FLYNN, J.R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, 95, 29-51.
- FLYNN, J.R. (1987). Massive IQ gains in 14 nations. What IQ tests really measure. *Psychological Bulletin*, 101, 171-191.
- FLYNN, J.R. (1998). *IQ gains over time: Toward finding the causes*. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures*. Washington, DC: American Psychological Association.
- GIUNTI O.S. (2008). *SPM Standard Progressive Matrices. Standardizzazione italiana*. Firenze: Giunti O.S. Organizzazioni Speciali.
- GORSUCH, R.L. (1983). The theory of continuous norming. In R.L. Gorsuch (chair), *Continuous norming: An alternative to tabled norms?* Symposium conducted at the 91st Annual Convention of the American Psychological Association, Anaheim, CA.
- HOLMES, B.J. (1980). *British Columbia norms for Wechsler Intelligence Scale for children-Revised; Peabody Picture Vocabulary Test; Slosson Intelligence Test; Standard Progressive Matrices, and Mill Hill Vocabulary Scale*. Unpublished manuscript, University of British Columbia: Faculty of Education.
- JAWOROWSKA, A. & SZUSTROWA, T. (1991). *Podrecznik Do Testu Matryc Ravena*. Warsaw: Pracownia Testow Psychologicznych Polskiego Towarzystwa Psychologicznego.
- KAPLAN, R.G. & SACCUZZO, D.P. (1997). *Psychological testing: Principles, applications and issues*. Pacific Grove (CA): Brooks Cole Publishing Company.
- KRATZMEIER, H. & HORN, R. (1979). *Manual: Raven-Matrizen-Test, Standard Progressive Matrices*. Weinheim, Germany: Beltz Test.
- LYNN, R., ALLIK, J. & IRWING, P. (2004). Sex differences on three factors identified in Raven's Standard Progressive Matrices. *Intelligence*, 32, 411-424.
- MACKINTOSH, N.J. & BENNETT, E.S. (2005). What do Raven's Matrices measure? An analysis in terms of sex differences. *Intelligence*, 33, 663-674.
- MARTINOLLI, L. (1990). *Étude et Réévaluation des Matrices Progressives Couleur*. Unpublished manuscript, L'Institut de Psychologie, Université de Fribourg.
- OTIS, A.S. (1936-39). *Otis Quick-Scoring Mental Ability Test*. New York: Harcourt, Brace & World.
- PICONE, L. (1996). *Il pensiero formale nell'adolescente. Studio metrico della Scala ECDL di Longeot-Hornemann*. Padova: CEDAM.
- RAVEN, J.C. (1938). *Progressive Matrices*. London: Lewis.
- RAVEN, J.C. (1941). Standardization of Progressive Matrices 1938. *British Journal Medical Psychology*, 19, 137-150.
- RAVEN, J. (1981). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Research supplement no. 1: The 1979 British standardization of the Standard Progressive Matrices and Mill Hill Vocabulary Scales, together with comparative data from earlier studies in the UK, US, Canada, Germany, and Ireland*. Oxford, England: Oxford Psychologists Press; San Antonio, TX: The Psychological Corporation.
- RAVEN, J. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology*, 41, 1-48.
- RAVEN, J. et al. (1990/2000). *Manual for Raven's Progressive Matrices*

- and Vocabulary Scales. Research supplement no. 3 (2nd/3rd edition): A compendium of international and North American normative and validity studies together with a review of the use of the RPM in neuropsychological assessment. Oxford, England: Oxford Psychologists Press; San Antonio, TX: The Psychological Corporation.
- RAVEN, J.C., COURT, J.H. & RAVEN, J. (1983). *Manual for Raven's Progressive Matrices and Vocabulary Scales. (Section 3) Standard Progressive matrices*. London: Lewis.
- RAVEN, J.C., COURT, J.H. & RAVEN, J. (1986). *Manual for Raven's Progressive Matrices and Vocabulary Scales. (Section 3) Research supplement n.3*. London: Lewis.
- RAVEN, J.C., COURT, J.H. & RAVEN, J. (1992). *Standard Progressive Matrices, 1992 Edition*. Oxford: Oxford Psychologists Press.
- RAVEN, J.C., COURT, J.H. & RAVEN, J. (1995). *Raven, Matrices Progressives (Escalas: CPM, SPM, APM)*. Madrid: TEA Ediciones, S.A.
- RAVEN, J. & RAVEN, J. (2008) *Uses and Abuses of Intelligence: Studies Advancing Spearman and Raven's Quest for Non-Arbitrary Metrics*. Unionville, New York: Royal Fireworks Press; Edinburgh, Scotland: Competency Motivation Project; Budapest, Hungary: EDGE 2000; Cluj Napoca, Romania: Romanian Psychological Testing Services SRL.
- RAVEN, J., RAVEN, J.C. & COURT, J.H. (1978). *Manual for Raven's progressive matrices and vocabulary scales*. London: Lewis
- RAVEN, J., RAVEN, J.C. & COURT, J.H. (1998). *Raven manual: Section 3, Standard Progressive Matrices*. Oxford, England: Oxford Psychologists Press.
- RAVEN, J.C. & SUMMERS, B. (1986). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Research Supplement No. 3*. London: Lewis.
- RAVEN, J.C. & WALSHAW, J.B. (1944). Vocabulary tests. *British Journal of Medical Psychology*, 20, 185-194.
- REDA, G.C., NENCINI, R. & RICCIO, D. (1955). Contributo alla taratura delle Progressive Matrices 1937 e 1938 di J.C. Raven. *Bollettino di Psicologia Applicata*, 7-8, 15-22.
- RUSHTON, J.P. & JENSEN, A.R. (2005). Thirty years of research on race differences in cognitive Ability. *Psychology, Public Policy and Law*, 11 (2), 235-294.
- SACCUZZO, D.P. & JOHNSON, N.E. (1995). Traditional psychometric test and proportionate representation: An intervention and program evaluation study. *Psychological Assessment*, 7 (2), 183-194.
- SPEARMAN, C. (1927). *The abilities of man*. New York: McMillan.
- TERMAN, L.M. (1916). *The measurement of Intelligence*. Boston: Houghton Mifflin.
- TUDDENHAM, R.D., DAVIS, L., DAVISON, L. & SCHINDLER, R. (1958). *An experimental group version for school children of the Progressive Matrices*. Unpublished manuscript, University of California. (See also Abstract: *Journal of Consultant Psychology*, 22, 30.)
- UNITED STATES GOVERNMENT, BUREAU OF THE CENSUS (1984). *Statistical Abstract of the United States, 1983*. Washington, DC: U.S. Government Printing Office.
- VALSESCHINI, S. & DEL TON, F. (1973). *Le Matrici Progressive di Raven. Contributo critico alla taratura su 1123 soggetti e considerazioni sulla validità e attendibilità della prova*. Firenze: O.S. Organizzazioni Speciali.
- VAN DER VEN, A.H.G.S. & ELLIS, J.L. (2000). A Rasch analysis of Raven's Standard Progressive Matrices. *Personality and Individual Differences*, 29, 45-64.
- WECHSLER, D. (1939). *The measurement of adult intelligence*. Baltimore: Williams & Wilkins.
- ZACHARY, R.A. & GORSUCH, R.L. (1985). Continuous norming: Implications for the WAIS-R. *Journal of Clinical Psychology*, 41 (1), 86-94.
- ZHANG, H.C. & WANG, X.P. (1989). Chinese standardisation of Raven's Standard Progressive Matrices. *Psychological Test Bulletin*, 2 (2), 36-39.