19.00 €

vol. LXV. May-August 2017

N° 279

BPA

2/3

Four-monthly Journal
ISSN 0006-6761

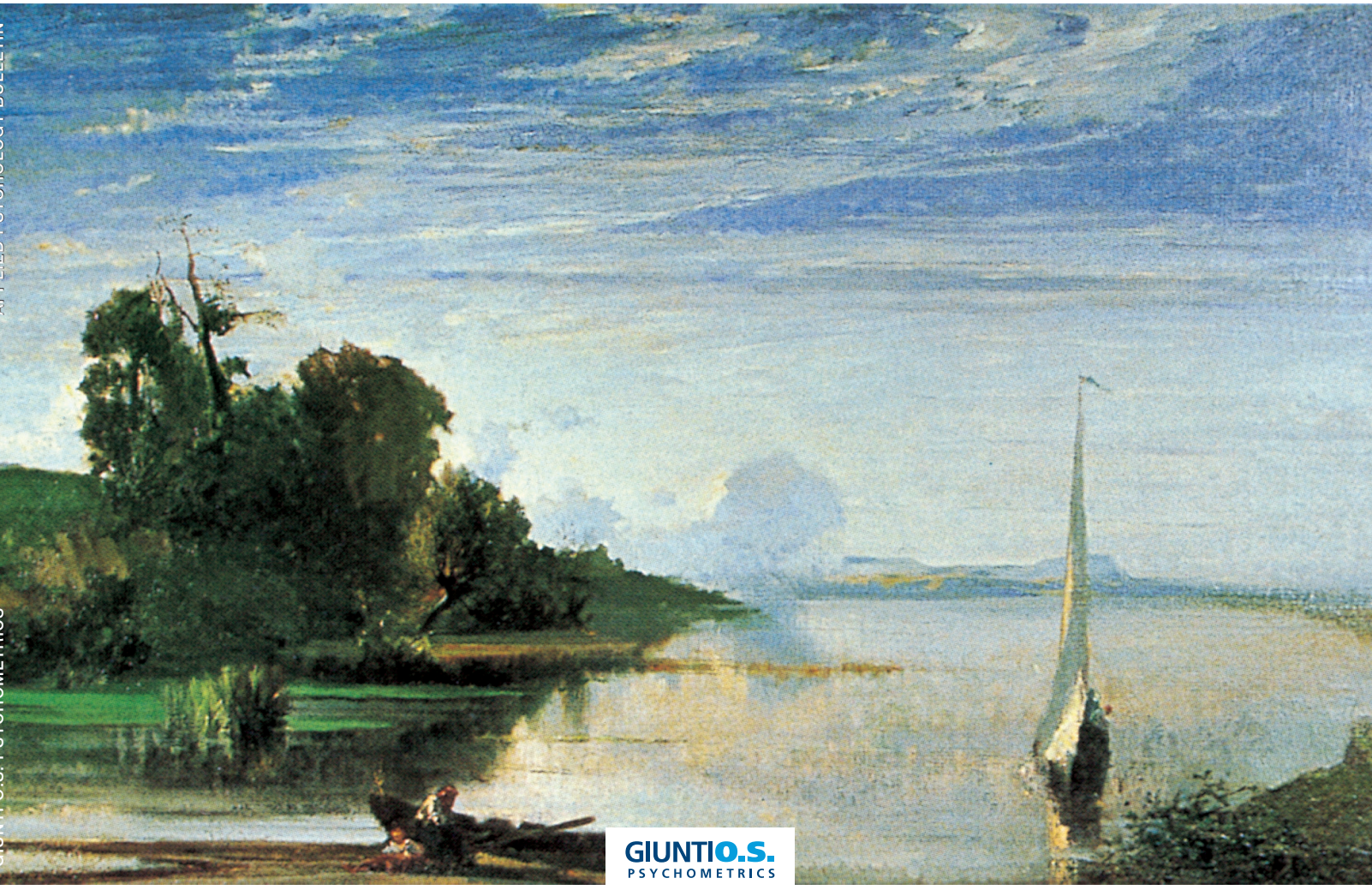
Bollettino di Psicologia Applicata

APPLIED PSYCHOLOGY BULLETIN

Indexed in PsycINFO® – Scopus Bibliographic Database

Research

Experiences & Tools



GIUNTI O.S. PSYCHOMETRICS

Scientific Director Alessandro Zennaro

CONTENTS

⬢ Research

▲ Experiences & Tools

Is a web video effective in increasing intention to use condoms? A test based on the Health Action Process Approach

Andrea Gragnano, Massimo Miglioretti

Department of Psychology, University of Milano-Bicocca

ABSTRACT. Nonostante la grande diffusione, l'efficacia dei video via Internet nella promozione della salute non è ancora stata dimostrata. Utilizzando come riferimento teorico il modello HAPA il presente studio valuta, su 352 partecipanti reclutati attraverso Facebook, l'efficacia di un video via Internet che promuove l'utilizzo del condom per incrementarne l'intenzione d'uso. Mediante modelli di equazioni strutturali lo studio analizza anche la bontà della fase motivazionale del modello HAPA rispetto al comportamento di utilizzo del condom. I risultati mostrano come i video via Internet possano essere strumenti efficaci nella promozione dell'uso del condom ma anche come questi debbano essere inseriti in campagne di promozione più strutturate per generare cambiamenti importanti. Il modello HAPA spiega il 33% della varianza nell'intenzione d'uso del condom mostrando la sua validità anche con questo tipo di comportamento.

SUMMARY. Introduction: The effectiveness of internet video in the promotion of healthy sexual behaviours is still not proved despite the large diffusion of this tool on the web. It is useful to rely on the theories of behaviour change when tailoring and evaluating a health promoting video. The HAPA model is a well-known theory of behaviour change still not tested in the field of promotion of healthy sexual behaviour. This study tested the usefulness of the motivational phase of the HAPA model to the promotion of healthy sexual behaviour and the effectiveness of a video delivered via web in increasing the intention to use condoms. Methods: The data were collected among 352 Italian participants recruited through Facebook. Participants were randomly assigned to the vision of the video or to the control group. Assessments were pre and post the intervention. SEMs were used to test the hypothesis. Results: The HAPA model explains 33% of the variance of Intention to use condoms and fit the data well. Outcome expectancies and self-efficacy are significant predictors of intention and the role of risk perception is discussed. The video increases the intention to use condoms ($\beta = .11$ $p<.05$). The video was more effective among non-intenders than intenders. The moderating effect of the initial level of intention had a p value of .07. Conclusion: Web videos are potentially effective and efficient tools to use in broader campaigns promoting the use of condom. Theories of behaviour change have to be used to design effective and tailored web videos. The HAPA model demonstrated its validity for the behaviour of condom use.

Keywords: *Health promotion, Video, Behaviour change*

INTRODUCTION

AIDS and sexually transmitted infections (STI) prevention is a crucial fact for the health of people, even in the more developed countries. The vision of the Joint United Nations Program on HIV/AIDS (UNAIDS, 2011) 'zero discrimination, zero new infections and zero AIDS-related deaths' is an unrealized purpose: 27% more people living with HIV from 1999 to 2009 and since 2010 there have been no declines in new HIV infections among adults (UNAIDS, 2016). Although the largest epidemic hits the sub-Saharan region even in the high income countries the situation is not as positive as it might be expected. The number of newly infected people in North America increased from 66,000 in 2001 to 70,000 in 2009 (UNAIDS, 2010). In 2015, there were an estimated 91,000 new HIV infections in Western and Central Europe and North America (UNAIDS, 2016). Along with the problem of the HIV diffusion there is also the issue of the sexually transmitted infections (STI). WHO estimates 357 million new cases of curable STI like syphilis, gonorrhea, chlamydia, and trichomoniasis every year (World Health Organization, 2016).

Since unsafe sexual intercourses are the main cause of AIDS and STI, it is clear that the role prevention plays in contrasting these diseases is very important, but prevention interventions in the high income countries need to be more sophisticated and precise (UNAIDS, 2010). A ten-year review of HIV/AIDS mass campaigns (Noar, Palmgreen, Chabot, Dobransky & Zimmerman, 2009) showed that great improvements have been achieved in this field but also that the effects and efficacy of every single tool implemented in the campaigns is unclear. A better comprehension of the efficacy of the tools implemented is crucial if we consider the development of internet and new technology that bring new ways of communication and new tools for prevention that seem especially appealing for younger, the more threatened by HIV and STI.

One technique identified as important in the process of formation of the intention to change behaviour is the use of persuasive videos. This type of videos is commonly used in health promotion interventions and campaigns together with other techniques and with the diffusion of the Internet and the broadband they are gaining even more importance. However, because the exclusive use - not associated with other technique - of this tool is rare we do not have reliable data about the effect of video on people when delivered via web and not via TV. Examining the findings in social psychology and marketing it is possible to identify a set of characteristics that a video has to have to be effectively persuasive (Table 1). Internet has a great dissemination in the developed countries even if not equal to that one of the Television, and like this last one allows the transmission of attractive dynamic audio/video messages. The Internet is cost-effective and on-demand, namely it is cheap for the number of people that can be reached by and it is the subject that decides when and where to use the intervention (Roberto, Zimmerman, Carlyle & Abner, 2007).

Like any other tool of a health promotion campaign, the video to be effective has to be tailored, that is to provide personally relevant messages (Krebs, Prochaska & Rossi, 2010). To tailor the message it is necessary to rely on the theories of behaviour change that indicate what dimension of the person has to be persuade. The Health Action Process Approach (HAPA; Schwarzer, 2008) divides the changing process in a *preintentional motivational* phase and a *postintentional volitional*. In the preintentional motivational phase individuals develop an intention throughout a series of weighting and evaluations. The factors involved at this stage are risk perception, outcome expectancies, and self-efficacy. The most important and general custom tailoring that can be made following the HAPA model is the distinction between the messages addressed to *intenders* and the ones addressed to *nonintenders*. That is, the effect of a video promoting the use condom will have different effect on people with a low or high initial intention to use condom.

The HAPA has been tested with many behaviours like alcohol consumption, eating habits, physical activity, performing regular breast self-examination, or seat belt use (Băban & Crăciun, 2007) but only three studies tested the HAPA model with the health behaviour condoms use (Carvalho, Alvarez, Barz & Schwarzer, 2015; Teng & Mak, 2011; Tsui, 2010).

AIMS OF THE STUDY

This study has three related purpose. First, to verify if the motivational phase of the HAPA adequately predicts the formation of intention of condoms use. Second, to test the efficacy of a video delivered via web in increasing the intention to use condoms. Finally, we aimed at support the utility of the HAPA model to design tailored health

Table 1 – Characteristics of an effectively persuasive video

Elements	Characteristics	How they should be	Sources
Source of the message / main character of the narration	Authority	Perceived as real, not pretended	Miller, Collins & Brief (1995)
	Credibility	Competence Reliability Good faith	McCroskey & Teven (1999)
	Social attractiveness	Pleasantness Similarity with the audience Physical attractiveness	Rhoads & Cialdini (2002) Anderson & McMillon (1995) Chaiken (1979)
Message	Rational/Emotional appeal	Effective if the central/ peripheral route of thinking of the audience is activated	Petty, Wheeler & Tormala (2003)
	Fear appeal	Only if it is followed by a message that persuade the audience to be able to face the threat	Roberto et al. (2007)
	Arguments in support	Present	Reynolds & Reynolds (2002)
	Story + emotional appeal	More effective than quantitative and factual evidences	Gray & Harrington (2011)
Audience	Previous cultural norms and attitudes	Recall them	Giles & Street (1994)
	Characters of the audience Features of the context	Careful evaluation	Burgoon, Denning & Roberts (2002)

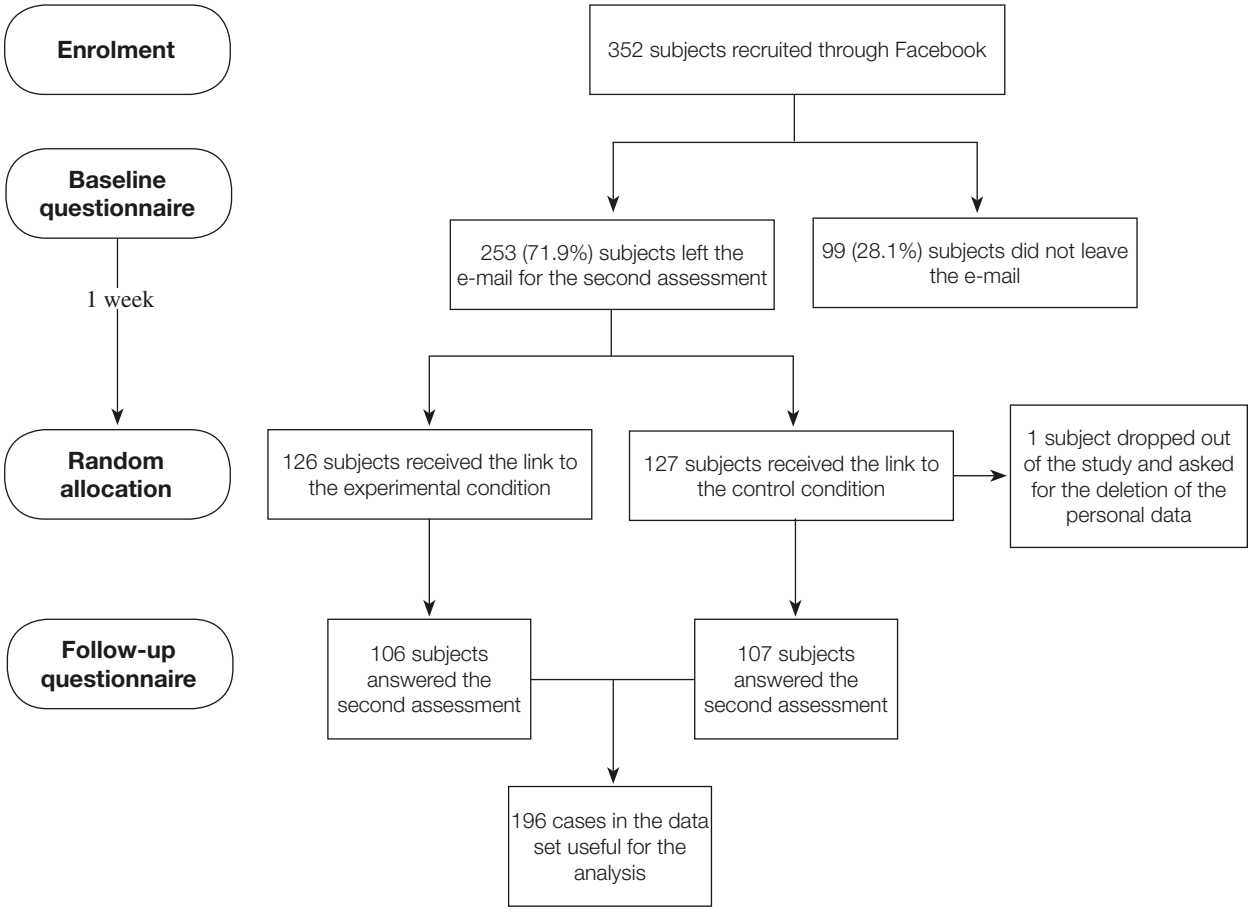
promotion web video. That is, the video will be more effective for non-intenders as identified by the HAPA model. Therefore, we hypothesized that:

- *H1*: The components of the HAPA model, namely risk perception, outcome expectancies and action self-efficacy, will predict the level of intention to use condom at baseline;
- *H2*: The vision of the video delivered via web will increase the intention to use condoms;
- *H3*: The level of initial intention to use condom will moderate the efficacy on intention of the web video. The video will be more effective among subjects with a lower intention (non-intenders).

PARTICIPANT

We recruited 352 participants at the baseline (Figure 1). Of these the 8% had not had a complete sexual intercourse at the time of the baseline assessment and so automatically skipped the questions about the past use of condom. The sample was investigated for demographic characteristics, sex (45% of male), age (68% between 16 and 25 years, 23% between 26 and 40, 9% more than 40), education (56% has a bachelor degree or higher), and occupation (56% were students, 36% workers, 5% unemployed and 3% were homemakers and pensioners). Only 10% of the participants were married or cohabiting

Figure 1 – Participant flow chart



however 60% has a committed relationship.

Two-hundred-thirteen (60,5%) of the people involved at the baseline participated also to the follow-up. Chi-square analysis of the sociodemographic characteristics between the group of people that completed only the baseline and that one that completed also the follow-up did not revealed any significant difference (gender $\chi^2_{(1)} = .81$, $p = .43$; education $\chi^2_{(1)} = .025$, $p = .91$; profession $\chi^2_{(4)} = 7.78$, $p = .1$; civil state $\chi^2_{(1)} = 3.63$, $p = .07$; age $\chi^2_{(2)} = 2.97$, $p = .23$; state of relationship $\chi^2_{(1)} = .017$, $p = .91$).

PROCEDURES AND MEASURES

We created a new questionnaire through the adaptation of others available in the literature. The questionnaire lasted about 15 minutes and consisted of two parts. The first part was composed by 10 socio-demographic items plus the request of an email address and the past use of condom. The second part contained 22 items assessing the social cognitive factors indicated by the HAPA model (Table 2). At the follow-up we submitted only the second part of the questionnaire.

We tested this set of questions through a confirmatory factorial analysis followed by a specification search (MacCallum, 1986) and a test for the invariance over time of the measurement model (Pitts, West & Tein, 1996). All the models were fitted using maximum likelihood estimate in LISREL (Jöreskog & Sörbom, 1996). The measurement model at the baseline and at the follow-up presented a good fit, $\chi^2(91, N = 196) = 204.33$, $p<.05$; $\chi^2/df = 2.25$, CFI = .93, RMSEA = .074; 90% CI [.059, .089] and $\chi^2(91, N = 196) = 190.70$, $p<.05$; $\chi^2/df = 2.10$, CFI = .94, RMSEA = .073; 90%

Table 2 – Social cognitive and behavioural measures used in the research

Variable	Citation	n. of items	Items
Past condom use (α at t0 = .94)	Based on Albarracín, Gilette, Earl et al. (2005) adapted following Schwarzer, Sniehotta, Lippke et al. (2003).	2	Ordinal-polytomous, from 0 times to 4 times (e.g., "Think about your last four sexual intercourses, how many times did you use condom?")
Intention to use condom (α at t0 = .93) (α at t1 = .90)	Based on Schwarzer & Renner (2000) and Schwarzer, Luszczynska, Ziegelmann et al. (2008)	2	8-points Likert scale (e.g., "Referring to your next sexual intercourse, do you think you will use a condom?")
Perceived absolute vulnerability (α at t0 = .95) (α at t1 = .96)	Based on Schwarzer, Sniehotta, Lippke et al. (2003)	4	7-points Likert scale (e.g., "I think my chances of contracting AIDS are…")
Perceived relative vulnerability (α at t0 = .97) (α at t1 = .98)	Based on Schwarzer, Sniehotta, Lippke et al. (2003)	4	7-points Likert scale (e.g., "If compared to the other people of the same age and sex of mine, I think my chances of contracting Syphilis are…")
Perceived severity (α at t0 = .56)* (α at t1 = .68)	Based on Schwarzer, Sniehotta, Lippke et al. (2003)	3	7-points Likert scale (e.g., "How much do you think Hepatitis is severe?")
Positive outcome expectancies (α at t0 = .58)* (α at t1 = .57)	Based on a research of the Centers for Disease Control and Prevention (2009)	3	7-points Likert scale (e.g., "If I use condom, I will feel more relaxed")
Negative outcome expectancies (α at t0 = .63)* (α at t1 = .65)	Based on a research of the Centers for Disease Control and Prevention (2009)	3	7-points Likert scale (e.g., "If I use condom, the pleasure will decrease")
Action self-efficacy (α at t0 = .87) (α at t1 = .89)	Based on Schwarzer, Renner, (2000) and Schwarzer, Luszczynska, Ziegelmann, et al. (2008)	5	4-points Likert scale (e.g., "Even if it can affect the relationship, I am confident that I am able to use condom")

Note. * The relative low level of Cronbach's Alpha is due to a low number of items in conjunction with a violation of tau equivalence (the items assessed different facets of the construct). Cronbach's α approaches reliability under condition of essential tau-equivalence (Cortina, 1993), otherwise it underestimates reliability (Graham, 2006). In these cases a better index of reliability is the composite reliability index (Bagozzi & Yi, 2011). The lowest composite reliability index was .65 for Positive outcome expectation at t1. This value is sufficient if there is a good fit of the model (Bagozzi & Yi, 2011).

CI [.058, .088] respectively. All the factor loadings in both models were statistically significant and both presented a mean item reliability of .65. Longitudinal metric invariance were respected since we obtained a ΔCFI lower than .010 (Cheung & Rensvold, 2002) between the model with the two measurement time together with no constrain and another one equal but constraining all factors loading involving the same indicator to be equal.

Procedure

Participant recruited for the study had to be older than 16 due to the sensitive data of the survey through Facebook. Using a social network service (SNS) as Facebook with research purposes is a relatively new practice which brings a lot of advantages but also some problems, furthermore, the studies that use this method are not so much widespread (Bull et al., 2011). With the SNSs the recruitment proceed like an accelerated snowball sampling with a very high number of potential respondents reached in a short time and in a very cost efficient way (Baltar & Brunet, 2012).

During the phase one, we posted the advertisement with the link to the questionnaire as a personal message of a profile with 130 friends so that this message appeared into the virtual wall of each friend of the source profile. We posted the same message every 2 days for 11 days. In the phase two, during the 12^{th} and 13^{th} day, we sent a Facebook's email with a call for the research to every friend of the source profile that had not yet replied to the messages in the virtual wall. In each phase the message contained beyond the link to the questionnaire a call for sharing the message with their own friends using the Facebook's button "share".

The first page that a participant found visiting the link to the survey was the information and agreement form. The respondents at the baseline that left the email at the end of the questionnaire were randomly assigned to the control or experimental group. Chi-square analyses of the sociodemographic characteristics and t-tests for mean differences in risk perception, outcome expectancies, self-efficacy and initial intention revealed no significant differences between the control and experimental group.

The control condition included only the compilation of the two on-line questionnaires whereas the experimental condition comprised also an intervention just before the follow-up. The link to the online questionnaire of the follow-up was sent 1 week after the baseline assessment.

The intervention in the experimental group consisted in a video lasting about 40 sec. shown just before the follow-up assessment in a natural context since it was delivered via web. Two judges selected the video concerning sexually transmitted disease among 123 available on the web. The judges separately evaluated the web videos with a two steps procedure. In the first step the two judges separately evaluated half of the videos each. The videos with at least one characteristic for each elements of Table 1 were selected. In the second step the judges selected the video that best fits the prescription in Table 1 with a consensus procedure.

The selected video showed a chaste sexual scene of a young couple underlined by a slightly anxious music. Parts of human body appear in the foreground: a shoulder, a back together with hands of outsiders, not visible in the first sequence. Meanwhile a voice start saying "Carol sleeps with Mark, Mark has slept with Sandra and Peter, Peter with Ann and Tracy and Susan, Susan with Dave, Steven and Roger." The number of hands increases so much that last image is a chaotic and confused interlacing of men's and women's hands. Then the video blackens and a notice recites: "Your past always sleeps with you". The following image is a blow-up photo of a condom with this subtitle: "Protect yourself in the future" (Solid & Hallerfilm AG, 2005). This video had many of the characteristics of Table 1. The simplicity of the message gave confidence in the honesty and credibility of the source of the message that did not appear as pretending to have a scientific truth or an institutional authority. The indefiniteness of the characters made the identification easy for the audience with an active sexual life. The message had both a rational and emotional appeal with a moderate fear appeal followed by a message that persuaded the audience to be able to face the threat. This last characteristic was the most difficult to find among the videos we examined, fear appeals were usually not followed by messages that persuade the audience to be able to face the threat. Finally, the selected video indirectly recalled the values of individual (sexual) freedom and responsibility that are widespread in contemporary western society.

RESULTS

According to the HAPA model *H1* predicted that risk perception, outcome expectancies and action self-efficacy are

significant determinants of intention. To verify this we tested the model presented in Figure 2.

We obtained a good fit of the model to the data with $\chi^2(86, N = 196) = 137.21$, $p<.05$; $\chi^2/df = 1.60$, CFI = .97, RMSEA = .050; 90% CI [.031, .068]. All the predictors of intention but *risk perception* had a statistically significant effect on *intention*. This model explained 33% of the variance of *intention to use condoms*, the 25% was explained by *positive outcome expectancies*, the 6% by *negative outcome expectancies* and the 2% by *action self-efficacy*. To understand if the role of risk perception in the behaviour of condom use is truly null and following the statement that risk perception is a sort of antecedent of self-efficacy and outcome expectancies (Schwarzer, 2001) we tested other modified versions of the HAPA model. At the end of this process, we obtained a model equal to the first except for the set of a covariance between *risk perception* and *self-efficacy* (Figure 3). This model showed a good fit to the data $\chi^2(85, N = 196) = 130.43$, $p = .0011$; CFI = .97, RMSEA = .047; 90% CI [.026, .065] with a significant negative correlation between *self-efficacy* and *risk perception* of −.20. Since the first model is nested into the second we compared the two models through the $\Delta\chi^2$ test. This test revealed a $\Delta\chi^2_{(1)} = 6.78$ with a $p<.05$ indicating that removing the correlation between *risk perception* and *self-efficacy* significantly worsens the fit of the model to the data. Given these results, we can sustain that *H1* has not been falsified.

To test *H2* with structural equation modelling we comply with the "group code" approach (Aiken, Stein & Bentler, 1994) which consists in including a dummy variable reflecting group membership in the model. Figure 4 presents the model used for *H2*. The latent variable *Treatment* is the dummy variable with the values of 1 for the non-vision of the video control group- and 2 for the vision of the video -experimental group-. The measurement model was constrained to be equal between the baseline and the follow-up. The overall model resulted in a good fit: $\chi^2(3, N = 196) = 2.14$, $p = .54$; CFI = 1, RMSEA = .0; 90% CI [.0, .11]. The *t* test for the correlation between the error of the *Intention* at the baseline and the *Treatment* was not significant ($\beta = -.13$) meaning that these two factors are uncorrelated and consequently there was not a significant difference between the control and experimental group in terms of level of intention to use condoms at the baseline. This was as expected, given the random assignment to the experimental condition.

H2 posited a positive effect of the video on the level of intention to use condoms. The path from *Treatment to Intention* at the follow-up was statistically significant ($\beta = .10$) meaning that the vision of the video led to a higher level of intention to use condoms. The model explained 63% of the variance of the intention to use condoms at the follow-up, with the initial intention explaining the majority of the variance of the final intention to use condoms. Therefore, also *H2* were not falsified.

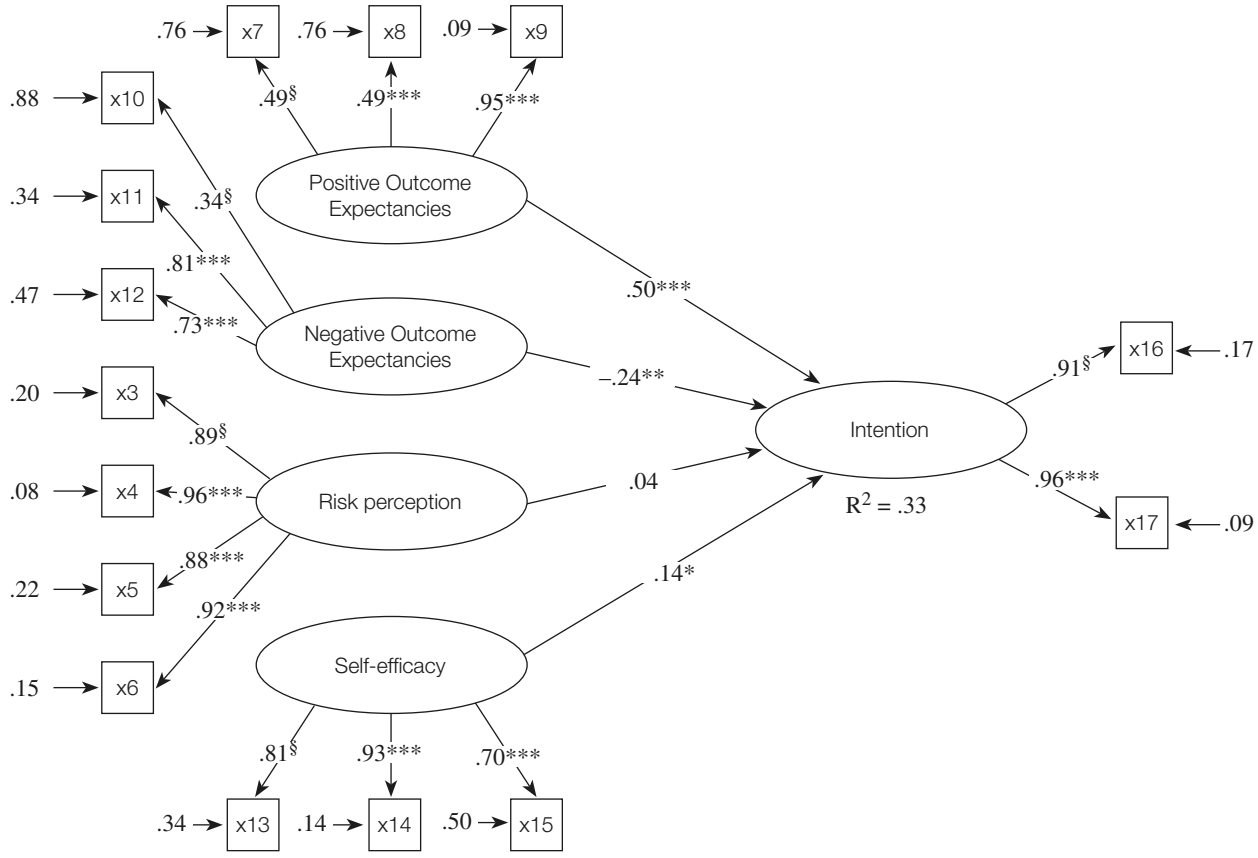
H3 stated a greater effect of the video for people with a low initial level of intention to use condoms – non-intenders – so we searched for a moderation effect of the initial level of intention on the effect of the video on the final level of intention. Specifically, we expected higher effect of the video at lower levels of initial intention. To test this hypothesis we followed the two-step Ping's approach (1995) which suggests introducing a multiplicative term in the original model after a mean-centring of variables. The model of the second step (Figure 5), the one with the multiplicative term, showed a good fit: $\chi^2(4, N = 196) = 3.702$, $p = .448$; CFI = 1, RMSEA = .0; 90% CI [.0, .104].

The path from the multiplicative term to the intention at the follow-up has a t-value of −1.767 so its statistical significance is $p = .077$, a little bit higher than .05 but still relevant. This indicates that probably there is the hypothesized moderating effect of the initial level of intention. Specifically, the unstandardized regression coefficient is −.77 and its effect on the relation is presented in Figure 6. This plot suggests that the vision of the video results in a higher level of intention to use condoms at the follow-up but this effect decreases until it will disappear with the increase of the initial level of intention. Given the level of significance of the moderating effect *H3* cannot be considered as confirmed but nevertheless as completely rejected.

DISCUSSION

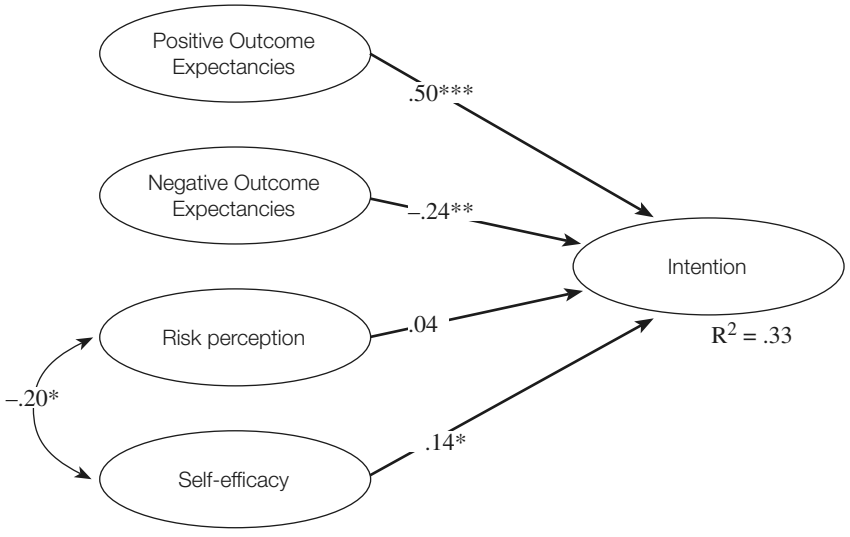
In this study we verified the applicability of the motivational phase of the HAPA model to the behaviour of condom use. For what is our knowledge HAPA model has been applied to the condom use behaviour only with the population of men who have sex with men (Teng & Mak, 2011), male drug user (Tsui, 2010), and recently in the general population but testing in the motivational phase only the outcome expectancies (Carvalho et al., 2015). We tested all the components of the motivational phase of the HAPA in a general population. The model has a good fit to the data

Figure 2 – Structural equation model of the motivational phase of the HAPA model



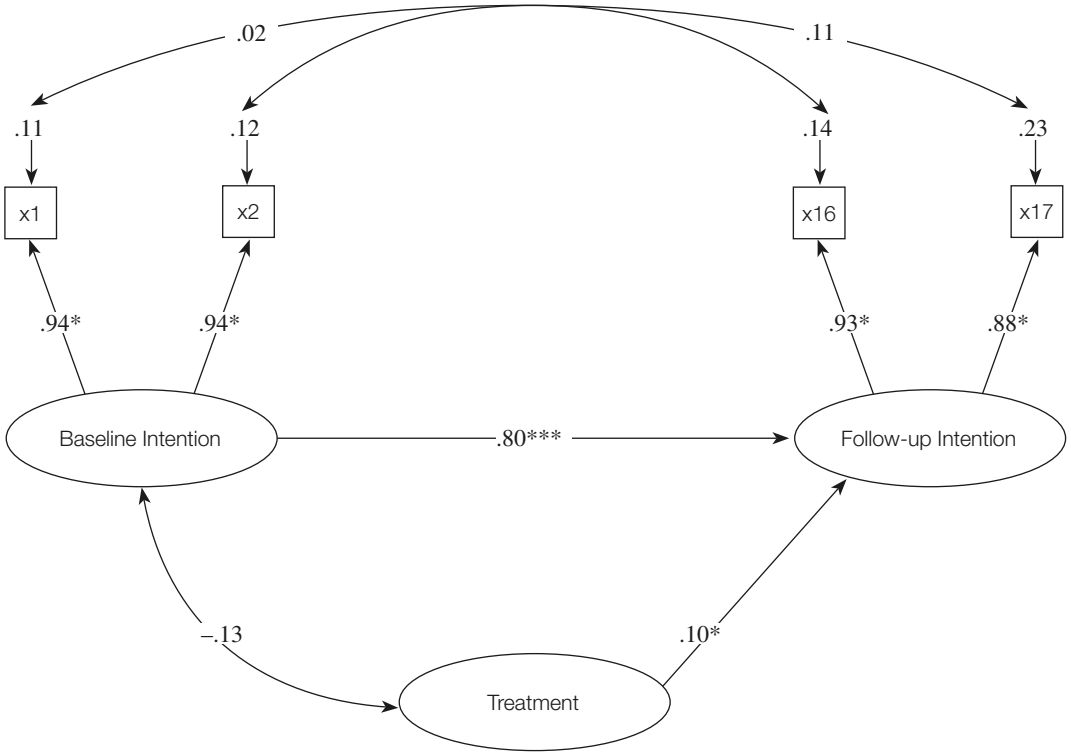
Note. * $p<.05$; ** $p<.01$; *** $p<.001$; § = value specified by the researcher

Figure 3 – Structural equation model of a modified version of the motivational phase of the HAPA model



Note. * $p<.05$; ** $p<.01$; *** $p<.001$

Figure 4 – Model used to test *H2*. All the paths present standardized parameters



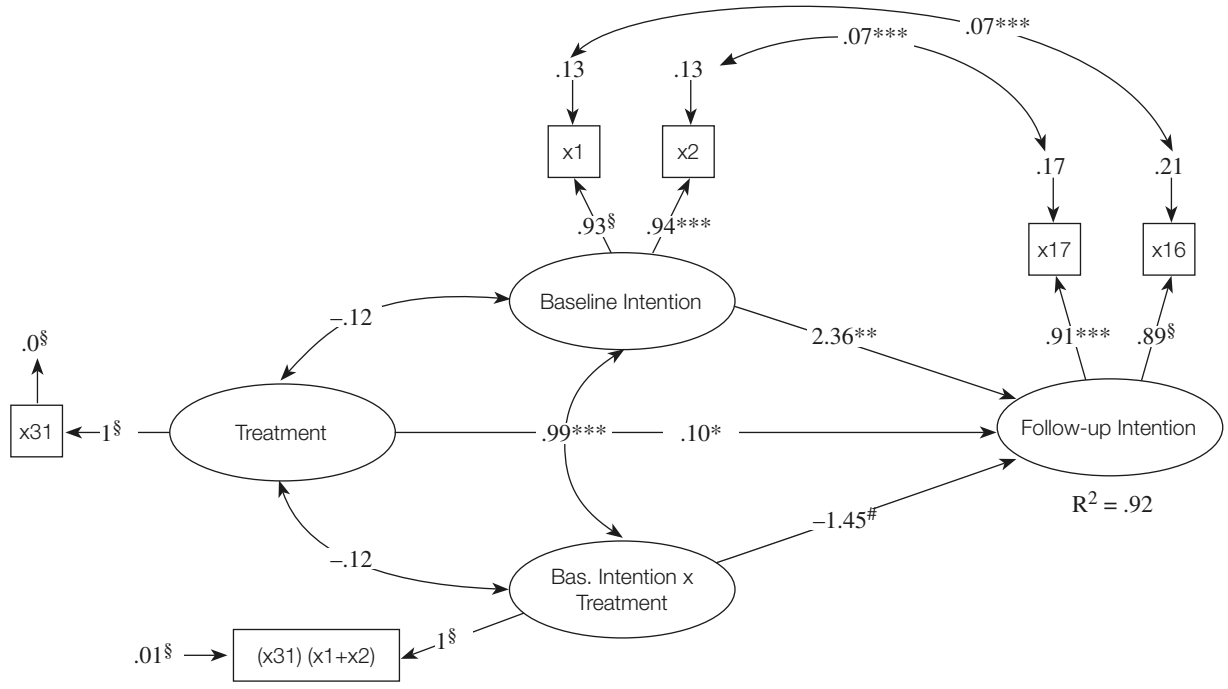
Note. * $p<.05$.

and the percentage of variance of intention explained by the model is in line with the one explained in other behaviours (Barg et al., 2012; Payaprom, Bennett, Alabaster & Tantipong, 2011) and by other models of behaviour change (Armitage & Conner, 2001). In line with the result of Carvalho et al. (2015) the crucial aspects that seem to determine if a person has or not the intention to use condoms are the outcome behaviour expectancies and substantially the positive one, accounting for the 25% of the variance of the intention. An intervention aimed at increasing the intention to use condoms should focus its efforts at increasing the positive outcome expectancies of the use of condom and partly at reducing the negative one. In this study, the "classical" role attributed to risk perception on intention in the HAPA model is absent. However, we tested a modified version of the model finding a statistically significant correlation between risk perception and self-efficacy. This could indicate the presence of two mind-sets that are not causally related with the actual intention to use condoms. One characterized by a great perceived ability in the use of condom that may indicate a "confident" approach to the sexual matters predisposing one to feel able to cope with the STI with a low risk perception. The other defined by a perceived difficulties in the management of the condom use that may indicate a "complex" approach to the sexual matters predisposing one to feel at risk of STI with a high risk perception. However, future researches should test this interpretation.

With the second hypothesis, we tested the efficacy of a video delivered via web in increasing the intention to use condoms. Our data supported the possibility to use internet to deliver effective health promotion videos. There is a statistically significant increase in the level of intention due to the vision of the video but the magnitude of this effect is low and the most part of the variance of the intention to use condoms at the follow-up was explained by the initial levels of intention.

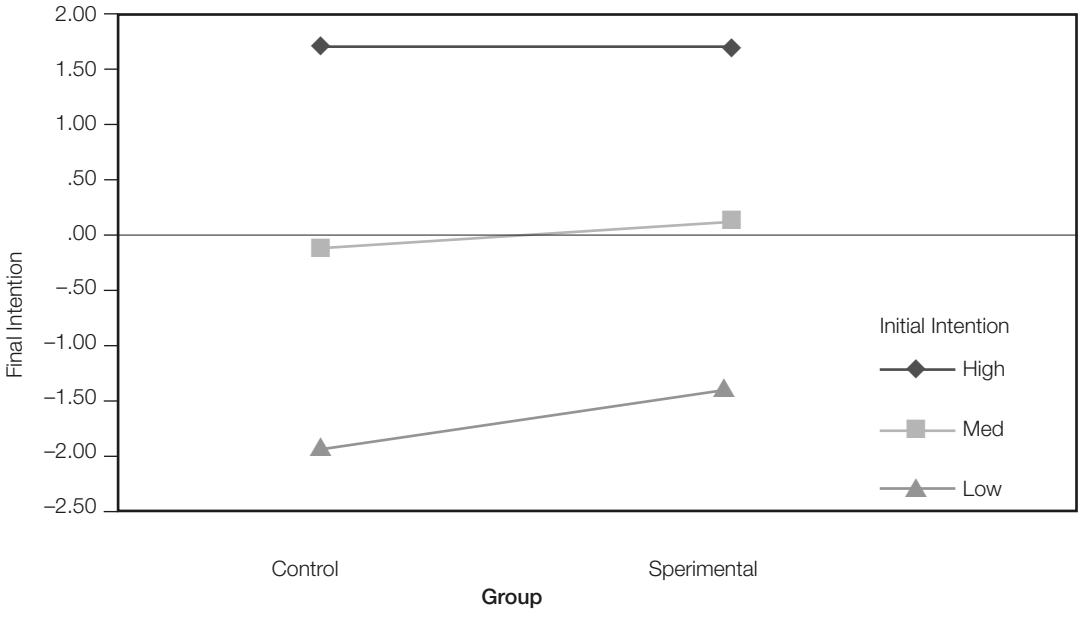
Nevertheless, this result is very important for two reasons. Firstly, the video and the questionnaires were delivered via

Figure 5 – Model of the second step in the Ping's approach for the moderation of the Treatment on Final Intention by the level of Initial Intention



Note. *p<.05, ** p<.01, *** p<.001, #p = .077, § = value specified by the researcher

Figure 6 – Moderation of the Effect of the Video on Intention by the level of Initial Intention



web and every subject saw the video in their real life setting, not in a laboratory. This means the effect we found accounted for all the situations that can potentially distract the enduser enjoying health-promoting videos through a SNS in his daily life. Secondly, we did not want to demonstrate that a video of 40 seconds is the way to increase the intention to use condoms but that the common use of health promoting videos on the web is a useful and potentially effective practice. However, the small effect clearly indicates that the use of video alone is not enough to produce really impacting health promotion campaigns.

The third and last hypothesis recalls the needs of tailored messages in to produce persuasive messages. The HAPA model state that there is a qualitative difference between people in the motivational stage and the ones in the volitional stage so in order to be effective a health promotion intervention has to be tailored on the stage in which the target population is. That is, if the intervention is aimed at increasing the motivation in the use of condom it will be effective for the people in the motivational stage and useless for the people in the volitional stage. As exposed before the distinction between the two stages is based on the level of intention. A high level of intention place a person in the volitional phase conversely a low level of intention place a person in the motivational phase. We founded a moderator effect of the initial level of intention to use condoms, that corresponds to the phase in which the people are, with a significance of .077. This level of significance is a critical one since it is difficult to refuse the hypothesis of moderation with this figure but at the same time it is not possible to conclude with confidence that moderation does exist. More studies are needed. We consider our data as an indication that moderation does exist. This means that the difference between the control and experimental group for the final level of intention to use condoms is higher for the people in the motivational phase and it decrease until disappearing for the people in the volitional phase. That is, the video is effective only for the people in the motivational phase.

The study presents some limitations that need to be taken into consideration. These are related with the adopted online data collection. Besides the limits of the classic snowball sampling, the use of SNSs leads to a selection bias because the Internet population constitutes a biased sample of the total population (Baltar & Brunet, 2012) and because of the self-selection of participants, the so called "volunteer effect" (Eysenbach & Wyatt, 2002).

In addition, the sample was highly heterogeneous, and this led to the sample being composed of individuals with very different ages, type of relationships and sexual habits. This made it possible to test only general hypothesis about the efficacy of the web video and the fit of the HAPA model but left unsolved a series of interesting issues. For example, it was not possible to compare the efficacy of the web video among specific subpopulations (e.g., young and elderly, homosexual and heterosexual, male and female) or the differential fit of the HAPA model. However, these issues were beyond the initial aims of the present research and the heterogeneity of the sample was also positive. We are more confident that the web video and the motivational phase of the HAPA model are appropriate for a varied population.

CONCLUSION

This study revealed that persuasive videos promoting healthy sexual behaviours can be effectively delivered via web. However, the video has to be part of a bigger and structured campaign to obtain relevant results.

Our results clearly indicate the applicability of the motivational phase of the HAPA model to the condom use behaviour and it remarks the need of a deeper knowledge about the role of risk perceptions and positive outcome expectations in developing the intention to use condoms. Using the HAPA model as our theoretical framework allowed us to discover the importance of the positive, and to a lesser extent of the negative, expectancies about the use of condom on intention. Without using a behaviour change model these insights would be impossible. As shown by Noar et al. (2009) the greatest progress in the HIV/AIDS mass campaigns during the ten years consisted in a more careful and wider use of the theories of behaviour change during the phase of planning of the campaign. A theoretically embedded campaign presents other advantages beyond a better efficacy. Specifically, a campaigner using a theory as conceptual framework is able to rationally and tidily determine the factors that may determine the behaviour change and consequently design and tailor the best messages. The information that may be inferred from the evaluation of a theory-based campaign is richer than those from a standard campaign. Through a test of the changes in the factors indicated by the theory as determinants of the targeted behaviour it is possible to identify which part of the campaign does or does not work and why (Rothman, 2004).

References

AIKEN, L.S., STEIN, J.A. & BENTLER, P.M. (1994). Structural equation analyses of clinical subpopulation differences and comparative treatment outcomes: Characterizing the daily lives of drug addicts. *Journal of consulting and clinical psychology*, *62*(3), 488-99.

ALBARRACÍN, D., GILETTE, J.C., EARL, A.N., GLASMAN, L.R., DURANTINI, M.R. & HO, M.H. (2005). A test of major assumptions about behaviour change: A comprehensive look at the effects of passive and active HIV-prevention interventions since the beginning of the epidemic. *Psychological Bulletin*, *131*(6), 856-897.

ANDERSON, R.B. & McMILLON, P.Y. (1995). Effect of similar and diversified modeling on African American women's efficacy expectations and intention to perform breast self-examination. *Health Communication*, 7(4), 324-343.

ARMITAGE, C.J. & CONNER, M. (2001). Efficacy of the theory of planned behaviour: A structured review. *British Journal of Social Psychology*, *40*(4), 471-499.

BĂBAN, A. & CRĂCIUN, C. (2007). Changing healthrisk behaviors: A review of theory and evidence-based interventions in health psychology. *Journal of Cognitive and Behavioral Psychotherapies*, *7*(1), 45-66.

BAGOZZI, R.P. & YI, Y. (2011). Specification, evaluation, and interpretation of structural equation models. *Journal of the Academy of Marketing Science*, *40*(1), 8-34. http://doi.org/10.1007/s11747011-0278-x

BALTAR, F. & BRUNET, I. (2012). Social research 2.0: Virtual snowball sampling method using Facebook. *Internet Research*, *22*(1), 57-74.

BARG, C.J., LATIMER, A.E., POMERY, E., RIVERS, S.E., RENCH, T., PRAPAVESSIS, H. & SALOVEY, P. (2012). Examining predictors of physical activity among inactive middle-aged women: an application of the Health Action Process Approach. *Psychology & health*, *27*(7), 829-845.

BULL, S., BRESLIN, L., WRIGHT, E., BLACK, S., LEVINE, D. & SANTELLI, J. (2011). Case Study: An Ethics Case Study of HIV Prevention Research on Facebook: The Just/Us Study. *Journal of Pediatric Psychology*, *36*(10), 1082-1092.

BURGOON, M., DENNING, V.P. & ROBERTS, L. (2002). Language expectancy theory. In J.E. Dillard & M. Pfau (Eds.), *The persuasion handbook: Developments in theory and practice*. Thousand Oaks: Sage.

CARVALHO, T., ALVAREZ, M.J., BARZ, M. & SCHWARZER, R. (2015). Preparatory behavior for condom use among heterosexual young men: A longitudinal mediation model. *Health Education & Behavior, 42*(1), 92-99.

CENTERS FOR DISEASE CONTROL AND PREVENTION (2009). *Sexually Transmitted Disease Surveillance*, *2008*. Atlanta, GA: U.S. Department of Health and Human Services.

CHAIKEN, S. (1979). Communicator's physical attractiveness and persuasion. *Journal of Personality and Social Psychology*, *37*(8), 1387-1397.

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

CORTINA, J.M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98-104.

GILES, H. & STREET, R.L. (1994). Communicator characteristics and behaviour. In M.L. Knapp & G.R. Miller (Eds.), *Handbook of interpersonal communication*. Thousand Oaks: Sage.

GRAHAM, J.M. (2006). Congeneric and (Essential) Tau-Equivalent Estimates of Score Reliability. What They Are and How to Use Them. *Educational and Psychological Measurement*, *66*(6), 930-944.

GRAY, J. & HARRINGTON, N. (2011). Narrative and framing: A test of an integrated message strategy in the exercise context. *Journal of Health Communication*, *16*(3), 264-281.

JÖRESKOG, K. & SÖRBOM, D. (1996). *LISREL 8: User's reference guide*. Chicago: Scientific Software International.

KREBS, P., PROCHASKA, J.O. & ROSSI, J.S. (2010). A meta-analysis of computer-tailored interventions for health behavior change. *Preventive medicine*, *51*(3), 214-221.

LIPPKE, S., SCHWARZER, R., ZIEGELMANN, J.P., SCHOLZ, U. & SCHÜZ, B. (2010). Testing stage-specific effects of a stage-matched intervention: A randomized controlled trial targeting physical exercise and its predictors. *Health Education & Behaviour*, *37*(4), 533-546.

MacCALLUM, R. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin*, *100*(1), 107-120.

McCROSKEY, J.C. & TEVEN, J.J. (1999). Goodwill: A reexamination of the construct and its measurements. *Communication Monographs*, *66*(1), 90-103.

MILLER, A.G., COLLINS, B.E. & BRIEF, D.E. (1995). Perspectives on obedience to authority: The legacy of Milgram experiments. *Journal of Social Issues*, *51*(3), 1-19.

NOAR, M., PALMGREEN, P., CHABOT, M., DOBRANSKY, N. & ZIMMERMAN, R.S. (2009). A 10-year systematic review

of HIV/aids mass communication campaigns: Have we made progress? *Journal of Health Communication*, *14*(1), 15-42.

PAYAPROM, Y., BENNETT, P., ALABASTER, E. & TANTIPONG, H. (2011). Using the Health Action Process Approach and implementation intentions to increase flu vaccine uptake in high risk Thai individuals: A controlled before-after trial. *Health psychology*, *30*(4), 492-500.

PETTY, R.T., WHEELER, S.C. & TORMALA, Z.L. (2003). Persuasion and attitude change. In T. Millon & M.J. Lerner (Eds.), *Handbook of psychology: Vol. 5*. Hoboken: Wiley.

PING, R.A. (1995). A parsimonious estimating technique for interaction and quadratic latent-variables. *Journal of Marketing Research*, *32*(3), 336-347.

PITTS, S.C., WEST, S.G. & TEIN, J.Y. (1996). Longitudinal measurement models in evaluation research: Examining stability and change. *Evaluation and Program Planning*, *19*(4), 333-350.

REYNOLDS, R.A. & REYNOLDS, J.L. (2002). Evidence. In J.E. Dillard & M. Pfau (Eds.), *The persuasion handbook: Developments in theory and practice*. Thousand Oaks: Sage.

RHOADS, K.V.L. & CIALDINI, R.B. (2002). The business of influence: Principles that lead to success in commercial settings. In J.P. Dillard & M. Pfau (Eds.), *The persuasion handbook: Developments in theory and practice*. Thousand Oaks: Sage.

ROBERTO, A.J., ZIMMERMAN, R.S., CARLYLE, K.E. & ABNER, E.L. (2007). A computer-based approach to preventing pregnancy, STD, and HIV in rural adolescents. *Journal of Health Communication*, *12*(1), 53-76.

ROTHMAN, A. (2004). "Is there nothing more practical than a good theory?": Why innovations and advances in health behaviour change will arise if interventions are used to test and refine theory. *International Journal Of Behavioural Nutrition And Physical Activity*, *1*(11), Retrieved from http://www.ijbnpa.org/content/1/1/11

SCHWARZER, R. (2001). Social-Cognitive Factors in Changing Health-Related Behaviours. *Current Directions in Psychological Science*, *10*(2), 47-51.

SCHWARZER, R. (2008). Modeling health behavior change: How to predict and modify the adoption and maintenance of health behaviors. *Applied Psychology*, *57*(1), 1-29.

SCHWARZER, R., LUSZCZYNSKA, A., ZIEGELMANN, J.P., SCHOLZ, U. & LIPPKE, S. (2008). Social-cognitive predictors of physical exercise adherence: Three longitudinal studies in rehabilitation. *Health Psychology*, *27*(1), 54-63.

SCHWARZER, R. & RENNER, B. (2000). Social-cognitive predictors of health behaviour: Action self-efficacy and coping self-efficacy. *Health Psychology*, *19*(5), 487-495.

SCHWARZER, R., SNIEHOTTA, F.F., LIPPKE, S., LUSZCZYNSKA, A., SCHOLZ, U., SCHÜZ, B. et al. (2003). *On the assessment and analysis of variables in the Health Action Process Approach: Conducting an investigation*. Unpublished manuscript, Health Psychology Department, Freie Universeitat Berlin, Berlin, Germany. Retrieved from http://web.fu-berlin.de/gesund/hapa_web.pdf

SOLID & HALLERFILM AG (Producer), (2005). *Zürcher Aids Hilfe - Vergangenheit* [Videofile]. Retrieved from http://www.youtube.com/watch?v=W6BslQ4NZnM

TENG, Y. & MAK, W.S. (2011). The role of planning and self-efficacy in condom use among men who have sex with men: An application of the Health Action Process Approach model. *Health psychology*, *30*(1), 119-128.

TSUI, H.Y. (2010). *Risk perceptions, cognitive behavioural models and hiv-related risk behaviours among non-institutionalized male injecting drug users in china. Dissertation abstracts international*. Retrieved from http://search.barnesandnoble.com/Risk-Perceptions-Cognitive-Behavioural-Modelsand-HIV-related-Risk-Behaviours-among-Non-institutionalized-Male-Injecting-Drug-Users-in-China/Hi-Yi-Tsui/e/9781124497792

UNAIDS (2010). *Global report: UNAIDS report on the global AIDS epidemic 2010*. Joint United Nations Program on AIDS/HIV, Geneva, Switzerland.

UNAIDS (2011). *World AIDS day report 2011*. Joint United Nations Program on AIDS/HIV, Geneva, Switzerland.

UNAIDS (2016). *Fact sheet November 2016*. Joint United Nations Program on AIDS/HIV, Geneva, Switzerland. Retrieved from http://www.unaids.org/sites/default/files/media_asset/UNAIDS_FactSheet_en.pdf

WORLD HEALTH ORGANIZATION (2016). Global health sector strategy on sexually transmitted infections, 2016-2021. Geneva, Switzerland: WHO.

Norms for Letter and Number Sequencing, Figure Weights and Cancellation subtests for the elderly Italian population

Lina Pezzuti, Serena Rossetti

Department of Dynamic and Clinical Psychology,
Faculty of Medicine and Psychology,
Sapienza University of Rome

ABSTRACT. La somministrazione dei tre subtest supplementari *Riordinamento di lettere e numeri*, *Confronto di pesi* e *Cancellazione* della WAIS-IV a soggetti con età superiore ai 69 anni non è prevista in alcuna taratura degli altri Paesi del mondo. Poiché non siamo riusciti a trovare una giustificazione di tale decisione, in fase di taratura della WAIS-IV i curatori hanno deciso di somministrare tutti e 15 i subtest anche ai soggetti dai 70 ai 90 anni di età. In questo lavoro si presenta la tabella di conversione dei punteggi grezzi in punti ponderati per quattro fasce di età tra i 70 e i 90 anni. Inoltre, si forniscono le attendibilità e gli errori standard di misura dei tre subtest per le quattro fasce di età. Si discute sull'utilità per il clinico dell'utilizzo di tali subtest con gli anziani.

SUMMARY. The administration of the three supplementary subtests Letter and Number Sequencing, Figure Weights and Cancellation in WAIS-IV to people over 69 years of age is not included in the scaling of any other countries around the world. Since it was not possible to find a justification for this decision, the curators decided to administer all 15 subtests to people aged 70 to 90 during WAIS-IV scaling. This work presents the table converting raw scores into weighted points for four age groups between the ages of 70 and 90. It also provides the reliability and the standard measurement errors of the three subtests for the four age groups. The usefulness to clinicians of using these subtests among the elderly is discussed.

Keywords: *WAIS-IV, Ageing, Letter and Number Sequencing, Figure Weights, Cancellation*

INTRODUCTION

The fourth edition of the *Wechsler Adult Intelligence Scale* (*WAIS-IV,* Wechsler, 2008a, 2008b) is a clinical tool that can be administered to individuals to measure the intelligence of adolescents and adults aged between 16 years, 0 months and 0 days and 89 Years, 11 months and 30 days. The structure of WAIS-IV involves composite scores in specific cognitive areas such as the Verbal Comprehension Index (VCI), the Perceptual Reasoning Index (PRI), the Working Memory Index (WMI), and the Processing Speed Index (PSI). Each of these indices contributes to the composite score represented by total IQ, alternatively defined as the expression of general intelligence ability.

The 15 subtests belonging to each of the indices are subdivided into 10 labelled fundamental and 5 supplementary. To obtain the 4 indices and the total IQ it is sufficient to administer the 10 fundamental subtests, while the 5 additional subtests can be administered in two circumstances. The first circumstance is when the clinician needs to substitute a fundamental subtest with one of the supplementary ones (e.g. if a person has physical or sensory limitations, or if the scores from a fundamental subtest are invalidated due to administration errors or because the person always answers "don't know"); the second is when there is seen to be the need for a clinical investigation of a particular cognitive ability, completing the diagnosis by analysing discrepancies between several subtests.

The Verbal Comprehension Index includes three fundamental subtests (*Similarities*, *Vocabulary* and *Information*) and a supplementary subtest (*Comprehension*). The Perceptual Reasoning Index contains three fundamental subtests (*Block Design*, *Matrix Reasoning* and *Visual Puzzles*) and two supplementary subtests (*Figure Weights* and *Picture Completion*). The Working Memory Index includes two fundamental subtests (*Digit Span* and *Arithmetic*) and one supplementary subtest (*Letter* and *Number Sequencing*). Finally, the Processing Speed Index includes two fundamental subtests (*Symbol Search* and *Coding*) and one supplementary subtest (*Cancellation*).

Letter and Number Sequencing, Figure Weights and *Cancellation* are three supplementary subtests to be used only with people between the ages of 16 and 69, and therefore not with people aged 70 or over. The decision to not administer these subtests to these older age groups was not discussed or explained in the *Technical and Interpretive Manual* of WAIS-IV (Wechsler, 2008b), or in the literature regarding the Wechsler scale.

Letter and Number Sequencing (a supplementary subtest in the Working Memory Index) can be a valid substitute for the *Digit Span* subtest or the *Arithmetic* subtest for situations in which the performance of one of these two subtests is invalidated. In this subtest a sequence of mixed numbers and letters is read to the subject, who is asked to repeat the numbers in ascending order and the letters in alphabetical order. The task requires attention, concentration, mental manipulation, sequential processing, and short-term auditory memory capacity. It may also call for the processing of information, cognitive flexibility and fluid intelligence (Crowe, 2000; Groth-Marnat, 2003; Kaufman & Lichtenberger, 1999, 2006; Sattler, 2008). This subtest also includes a processing score which consists of the memory span of letters and numbers, i.e. the number of elements that make up the last correctly recalled sequence. This score can be particularly informative in cases in which performance is variable, for example if a person makes a mistake in one or two tests in various sequences.

Figure Weights is a supplementary subtest in the Perceptual Reasoning Index in which the person being tested observes some scales in equilibrium, one of which has a missing weight on a pan, and must select (within a time limit) the option that keeps the scales in equilibrium. This subtest was designed to measure a specific aspect of fluid reasoning: non-verbal and analogical quantitative reasoning. According to Carroll (1993), quantitative reasoning tasks involve reasoning processes that can be expressed mathematically through the use of inductive or deductive logic. Although the subtest also requires memory work to some extent, its weight is lower than that of classical quantitative reasoning tasks (such as in the *Arithmetic* subtest); this is because the *Figure Weights* subtest takes place through a visual presentation of the stimuli (the scales and weights), which enables the person to keep them constantly in view - and thus alive in his or her short-term memory - while seeking the solution to the problem. To find the correct solution to the various subtest tasks, however, the working memory is involved more as the difficulty of the items increases: the more difficult items require the consideration of a higher number of quantitative relationships between the shape and weight of the stimuli. This relationship between reasoning and working memory is unsurprising given the results of studies that suggest a

dynamic interaction between fluid reasoning, working memory and processing speed (Kyllonen & Christal, 1990; de Ribaupierre & Lecerf, 2006; Salthouse & Pink, 2008; Unsworth & Engle, 2007).

The *Cancellation* subtest is an additional subtest in the Processing Speed Index in which the people being tested analyse a structured system of figures and, within a predetermined time limit, have to mark the target figures. This subtest can function as a good substitute for the *Coding* subtest in situations in which the person being tested has low fine motor skills, or as a substitute for one of the two fundamental subtests within the Index (*Coding* or *Symbol Search*) in the event that one of these is invalidated. The *Cancellation* subtest was designed to measure processing and perception speed, selective visual attention, and vigilance and visual-motor skills (Bate, Mathias & Crawford, 2001; Geldmacher, Fritsch & Riedel, 2000; Sattler, 2008; Wojciulik, Husain, Clarke & Driver, 2001).

Cancellation tasks are mainly used in neuropsychological contexts such as the measurement of visual neglect and inhibition of motor response and perseverance (Adair, Na, Schwartz & Heilman, 1998; Geldmacher et al., 2000; Lezak et al., 2004; Na et al., 1999). Some studies have highlighted an age-related reduction in performance in the processing of visual information that is not associated with the presence of illness. These changes have been described as declines in visual research, in the identification processes of target stimuli and in the rapidity of attention shifts (Folk & Hoyer, 1992; Plude & Doussard-Roosevelt, 1989). According to Geldmacher et al. (2000) "Examining the effects of age on cancellation tasks is important, since these are normally used for the clinical evaluation of the visuospatial function in age-related illnesses such as stroke and dementia" (p.196).

These three subtests (*Letter and Number Sequencing*, *Figure Weights* and *Cancellation*) were administered to all subjects in the Italian WAIS-IV scaling sample, including those aged 70 to 90 years. The objective of a recent international article by Pezzuti & Rossetti (2017) was to study the factorial invariance and mean differences in factors of the hierarchical model (4 first-order factors corresponding to the 4 Indices and a second-order factor corresponding to general intellectual ability) of WAIS-IV among Italian adults and elderly people belonging to the standardisation sample. The results confirmed the configural and structural invariance of WAIS-IV between the two age groups, showing that the second-order hierarchy model is equally valid among the elderly population even when all 15 WAIS-IV subtests are considered. In the conclusions of the above-mentioned article a forthcoming Italian publication was announced that would enable Italian clinicians to use a table to convert the raw scores from the three subtests *Letter and Number Sequencing*, *Figure Weights* and *Cancellation* for 4 age groups within the 70 to 90 age range.

The objective of this work, therefore, is to provide this table as well as a table with the distribution of the cumulative frequencies of the *Letter and Number Sequencing Spans* (LNS) for the four age groups. It also provides the reliability and the standard measurement errors of the three subtests for the four age groups.

METHODOLOGY

Sample and instrument

The sample of reference for this article is the Italian WAIS-IV scaling sample consisting of 750 subjects aged 70 to 90, sub-divided into 4 age groups: 200 subjects (100 males and 100 females) aged 70 to 74; 200 (100 males and 100 females) aged 75 to 79; 200 subjects (100 males and 100 females) aged 80 to 84; and 150 subjects (75 males and 75 females) aged 85 to 90 (Orsini & Pezzuti, 2015). To make the four age groups representative of the Italian population in terms of educational level, each contains a number of subjects proportionate to the percentage of people in the Italian population possessing one of the following four educational levels: a) up to primary school certificate; b) middle school certificate; c) high school diploma; d) degree. This representative design was drawn up using data from the 15th ISTAT National Census (ISTAT, 2011).

Data analysis

To obtain the tables for converting raw scores into weighted scores, for each of the four age groups the raw scores of the 3 supplementary subtests *Letter and Number Sequencing*, *Figure Weights* and *Cancellation* were converted into weighted scores with $M = 10$ and $SD = 3$, using the continuous norming method proposed by Gorsuch (Gorsuch, 1983; Zachary & Gorsuch, 1985). A method that requires the

use of specific procedures to take into account the effect of age on both the means and the standard deviations of the raw score distributions. The weighted scores of the subtests vary between 1 and 19.

This procedure was not used for the *Letter and Number Sequencing Span*, for which the frequency distribution of the raw scores for each of the four age groups is reported.

As regards the study on the reliability of the subtests, with the exception of *Cancellation* - for which the test-retest method was used - the reliability of *Letter and Number Sequencing* and *Figure Weights* was calculated using the split-half method, correcting for the length of the test with the Spearman-Brown prophecy formula.

RESULTS

Table 1 shows the data for converting the raw scores of the three supplementary subtests into weighted scores according to the age of the subject. For example, if a 75-year-old person has obtained a raw score of 26 in the *Cancellation* supplementary subtest, this corresponds to a weighted score of 11.

Table 2, on the other hand, shows the cumulative frequencies of the *Letter and Number Sequencing Span*. For example a Span of 6 in the *Letter and Number Sequencing* subtest is obtained by 6% of the population aged 70 to 74.

Finally, Table 3 shows the results for reliability and standard measurement error (*SME*) of the three subtests by the 4 age groups. As can be observed, the reliability of the three subtests ranges from .86 for the *Cancellation* subtest among the 75-79 age group, and .92 for the *Letter and Number Sequencing* subtest among the 85 to 90 age group. These indices, according to Nunnally & Bernstein (1994), are from very good to excellent.

CONCLUSION AND DISCUSSION

The administration of the three supplementary subtests *Letter and Number Sequencing*, *Figure Weights* and *Cancellation* in WAIS-IV to people aged 70 to 90 is not included in the scaling of any other countries in the world, the instruction for this age group being to administer 12 subtests out of 15. Since it was not possible to find a scientific justification for this decision, the curators decided to administer all 15 subtests to people aged 70 to 90 during the Italian WAIS-IV scaling. However, when the Italian scaling was published no information was given regarding those three supplementary subtests administered to the elderly. Pezzuti & Rossetti (2017) subsequently set out to study whether, when using all 15 subtests, the hierarchical structure with 4 first-order factors (verbal comprehension, perceptual reasoning, working memory and processing speed) and a second-order factor (general intellectual ability) remained the same also for the elderly. Having obtained confirmation of this, this work provides the norms for using these subtests also with this age group.

Having the norms of these three subtests also for the elderly can increase the usefulness of WAIS-IV in evaluating the cognitive abilities of this population, improving its potential for use clinically and in research. We agree with Wechsler (2008a) when he says that each of the three subtests can provide additional information on intellectual functioning for clinical interpretation and decision-making, but it should be added that this would also seem to be true not only for adults but also for the elderly. Additionally, having these supplementary subtests also for people aged 70 to 90 provides clinicians with the possibility of using them as substitutes if needed (for example in the event of an invalid score in one of the fundamental subtests, or when a subject presents particular characteristics or limitations). It can be reasonably concluded that possessing the norms for all 15 subtests also for older age groups will give an opportunity to reflect more thoroughly on the typical performance of the elderly and on their strengths and weaknesses in intellectual abilities.

Table 1 – Conversion of raw scores into weighted scores for 4 age groups

	70-74 years			75-79 years			80-84 years			85-90 years			
PP	**LN**	**FW**	**CA**	**FW**	**CP**	**CA**	**LN**	**FW**	**CA**	**LN**	**FW**	**CA**	**PP**
19	27-30	18-27	58-72	25-30	17-27	51-72	24-30	15-27	45-72	22-30	13-27	39-72	**19**
18	26	17	54-57	24	16	48-50	22-23	14	42-44	21	–	36-38	**18**
17	24-25	16	51-53	23	15	45-47	21	13	40-41	19-20	12	34-35	**17**
16	23	15	47-50	21-22	14	42-44	19-20	12	37-39	18	11	31-33	**16**
15	21-22	14	44-46	20	13	39-41	18	11	34-36	16-17	10	29-30	**15**
14	20	13	40-43	18-19	12	36-38	17	–	31-33	15	9	26-28	**14**
13	18-19	12	37-39	17	11	32-35	15-16	10	28-30	13-14	–	24-25	**13**
12	17	11	33-36	15-16	10	29-31	14	9	25-27	12	8	21-23	**12**
11	15-16	10	30-32	14	9	26-28	12-13	8	22-24	11	7	19-20	**11**
10	14	8-9	26-29	12-13	8	23-25	11	7	20-21	9-10	6	16-18	**10**
9	12-13	7	23-25	11	7	20-22	9-10	6	17-19	8	5	14-15	**9**
8	11	6	19-22	9-10	6	17-19	8	5	14-16	6-7	4	11-13	**8**
7	9-10	5	16-18	8	5	13-16	6-7	4	11-13	5	–	9-10	**7**
6	8	4	12-15	6-7	4	10-12	5	3	8-10	3-4	3	6-8	**6**
5	6-7	3	9-11	5	3	7-9	3-4	2	5-7	2	2	4-5	**5**
4	5	2	5-8	3-4	2	4-6	2	0-1	2-4	0-1	0-1	0-3	**4**
3	4	0-1	2-4	2	0-1	0-3	0-1	–	0-1	–	–	–	**3**
2	2-3	–	0-1	0-1	–	–	–	–	–	–	–	–	**2**
1	0-1	–	–	–	–	–	–	–	–	–	–	–	**1**

Legenda. LN = Letter and Number Sequencing; FW = Figure Weights; CA = Cancellation.

Table 2 – Cumulative percentages, averages, standard and median deviations of the Letter and Number Sequencing Span by age group

Span L-N	70-74	75-79	80-84	85-90	Span L-N
8	.0	.0	.0	.0	**8**
7	.0	.0	.0	.0	**7**
6	6.0	.7	.5	.5	**6**
5	26.5	4.0	9.5	5.5	**5**
4	60.5	14.0	41.5	25.5	**4**
3	92.5	76.7	88.0	85.0	**3**
2	99.5	98.7	99.0	100.0	**2**
1	100.0	99.4	99.5	100.0	**1**
0	100.0	100.0	100.0	100.0	**0**
Average	3.85	2.93	3.38	3.16	**Average**
SD	1.04	.77	.87	.76	***SD***
Median	4	3	3	3	**Median**

Table 3 – Reliability and standard measurement errors of the three subtests by age group

		LN	FW	CA
r_{el}	70-74	.90	.90	.89
	75-79	.90	.88	.86
	80-84	.90	.87	.89
	85-90	.92	.87	.88
SME	70-74	1.44	1.04	3.56
	75-79	1.36	1.06	3.35
	80-84	1.32	.94	3.10
	85-90	1.30	.92	2.53

Legenda. LN = Letter and Number Sequencing; FW = Figure Weights; CA = Cancellation; r_{el} = Reliability; SME = Standard Measurement Error.

References

ADAIR, J.C., NA, D.L., SCHWARTZ, R.L. & HEILMAN, K.M. (1998). Analysis of primary and secondary influences on spatial neglect. *Brain and Cognition*, *37*, 351-367.

BATE, A.J., MATHIAS, J.L. & CRAWFORD, J.R. (2001). Performance on the test of everyday attention and standard tests of attention following severe traumatic brain injury. *The Clinical Neuropsychologist*, *15*, 405-422.

CARROLL, J.B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, England: Cambridge University Press.

CROWE, S.F. (2000). Does the letter number sequencing task measure anything more than digit span? *Assessment*, 7, 113-117.

FOLK, C.L. & HOYER, W.J. (1992). Aging and shifts of visual spatial attention. *Psychology and Aging*, *7*, 453-465.

GELDMACHER, D.S., FRITSCH, T. & RIEDEL, T.M. (2000). Effects of stimulus properties and age on random-array letter cancellation tasks. *Aging, Neuropsychology, and Cognition*, *7* (3), 194-204.

GORSUCH, R.L. (1983). The theory of continuous norming. In R.L. Gorsuch (chair), *Continuous norming: An alternative to tabled norms?* Symposium at the 91st Annual Convention of the American Psychological Association, Anaheim, August 26-30.

GROTH-MARNAT, G. (2003). *Handbook of psychological assessment* (4th ed.). New York: Wiley.

ISTAT (Istituto Nazionale di Statistica) (2011). *15° Censimento generale della popolazione e delle abitazioni*. In Internet (9 marzo 2015): http://censimentopopolazione.istat.it.

KAUFMAN, A.S. & LICHTENBERGER, E.O. (1999). *Essentials of WAIS-III assessment*. New York: Wiley.

KAUFMAN, A.S. & LICHTENBERGER, E.O. (2006). *Assessing adolescent and adult intelligence* (3rd ed.). Hoboken, NJ: Wiley.

KYLLONEN, P.C. & CHRISTAL, R.E. (1990). Reasoning ability is (little more than) working memory capacity? *Intelligence*, *14* (4), 389-433.

LEZAK, M.D., HOWIESON, D.B. & LORING, D.W. (2004). *Neuropsychological assessment* (4th ed.). New York: Oxford University Press.

NA, D.L., ADAIR, J.C., KANG, Y., CHUNG, C.S., LEE, K.H. & HEILMAN, K.M. (1999). Motor perseverative behavior on a line cancellation task. *Neurology*, *52*, 1569-1576.

NUNNALLY, J.C. & BERNSTEIN, H. (1994). *Psychometric Theory*. New York: McGraw-Hill.

ORSINI, A. & PEZZUTI, L. (2015). *WAIS-IV. Contributo alla taratura italiana (70-90 anni)*. Firenze: Giunti O.S. Organizzazioni Speciali.

PEZZUTI, L. & ROSSETTI, S. (2017). Letter-Number Sequencing, Figure Weights, and Cancellation subtests of WAIS-IV administered to elders. *Personality and Individual Differences*, *104*, 352-356.

PLUDE, D.J. & DOUSSARD-ROOSEVELT, J.A. (1989). Aging, selective attention, and feature integration. *Psychology and Aging*, *4*, 98-105.

RIBAUPIERRE, A. DE & LECERF, T. (2006). Relationships between working memory and intelligence from a developmental perspective: Convergent evidence from a neo-Piagetian and a psychometric approach. *European Journal of Cognitive Psychology*, *18* (1), 109-137.

SALTHOUSE, T.A. & PINK, J.E. (2008). Why is working memory related to fluid intelligence? *Psychonomic Bulletin & Review*, *15* (2), 364-371.

SATTLER, J.M. (2008). *Resource guide to accompany assessment of children: Cognitive foundations (5th ed.)*. San Diego, CA: Author.

UNSWORTH, N. & ENGLE, R.W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133* (6), 1038-1066.

WECHSLER, D. (2008a). *WAIS-IV: Wechsler Adult Intelligence Scale. Administration and Scoring Manual*. San Antonio, TX: Pearson.

WECHSLER, D. (2008b). W*AIS-IV: Technical and Interpretive Manual*. San Antonio, TX: Pearson.

WEISS, L.G., SAKLOFSKE, D.H., COALSON, D.L. & RAIFORD, S.E. (2010). *WAIS-IV Clinical Use and Interpretation*. USA: Elsevier.

WOJCIULIK, E., HUSAIN, M., CLARKE, K. & DRIVER, J. (2001). Spatial working memory deficit in unilateral neglect. *Neuropsychologia*, *39*, 390-396.

YUAN, K.H. & BENTLER, P.M. (2000). Three likelihoos-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, *30*, 165-200.

ZACHARY, R.A. & GORSUCH, R.L. (1985). Continuous norming: Implications for the WAIS-R. *Journal of Clinical Psychology*, *41* (1), 86-94.

Early reading treatment in children with developmental dyslexia improves both reading and spelling

Anna Judica[1], Maria De Luca[1], Paola Angelelli[2], Donatella Spinelli[1,3], Pierluigi Zoccolotti[1,4]

[1] Neuropsychology Unit, IRCCS Fondazione Santa Lucia, Rome
[2] Department of History, Society and Human Studies, Lab. of Applied Psychology and Intervention, University of Salento, Lecce
[3] Department of Human Movement, Social and Health Sciences, University of Rome "Foro Italico"
[4] Psychology Department, Sapienza University of Rome

ABSTRACT. I disturbi specifici della lettura e scrittura sono i più comuni disturbi di apprendimento e causano frequenti insuccessi scolastici. Lo studio ha lo scopo di verificare l'efficacia di un trattamento del disturbo di lettura in bambini italiani con dislessia che frequentano la terza classe della scuola primaria e la sua possibile generalizzazione alla prestazione di scrittura. Hanno partecipato allo studio 10 bambini con dislessia evolutiva. La prestazione di lettura pre- e post-trattamento è stata valutata mediante due test di lettura standardizzati e la registrazione dei tempi di reazione vocale. La prestazione di scrittura è stata valutata mediante una misurazione pre- e post-trattamento ad un test standardizzato di competenza ortografica. La principale procedura di trattamento consisteva nella presentazione tachistoscopica di parole singole, scelte sulla base delle caratteristiche psicolinguistiche. Ciascun bambino ha effettuato 35 sessioni di trattamento. Il miglioramento dei parametri di accuratezza e velocità nella lettura di testi e di parole singole (diverse da quelle usate nel corso del trattamento) è stato significativamente superiore al miglioramento riscontrato nei coetanei normolettori nello stesso periodo di tempo. Per la comprensione della lettura non si sono riscontrati cambiamenti significativi. I risultati mostrano l'efficacia di un trattamento precoce della dislessia evolutiva e la stretta relazione tra abilità di lettura e scrittura in questa fascia d'età.

SUMMARY. Developmental reading and spelling disorders are the most common learning disabilities and they have severe scholastic consequences. The present study aimed to test the efficacy of reading treatment in Italian third-grade children with dyslexia and its possible generalization to writing performance. A group of 10 third-grade children with dyslexia participated in the study. The main procedure of the treatment was a computerized program that included a tachistoscopic presentation of single words. Each child underwent 35 treatment sessions. Pre- and post-training reading performance was measured by two standard reading tests and vocal reaction time recordings. Pre- and post-training spelling ability was also measured using a standard spelling test. The percentage of improvement in accuracy and speed in reading texts and lists of words (not included among the trained items) was greater than the developmental increase characteristic of non-disabled children. Reading comprehension was only moderately affected and did not change appreciably after therapy. Results indicate the effectiveness of a reading treatment in the early stages of reading acquisition and the strict relationship between reading and spelling systems at this age.

Keywords: *Dyslexia, Reading, Treatment, Learning disability*

INTRODUCTION

Developmental disorders in reading and writing are relatively frequent deficits with important consequences on scholastic achievement and personal adjustment. In recent years greater attention has been given to reading and writing deficits in young native speakers of languages with relatively regular grapheme-to-phoneme correspondences, such as German and Italian, as opposed to the previous prevalent focus on English orthography (e.g. Share, 2008). The present paper describes a study that investigated the efficacy of a reading treatment in Italian third-grade children with dyslexia and its possible generalization to writing performance.

There is evidence that deficits in literacy acquisition in Italian may be expressed differently than in languages with opaque orthographies, such as English or French (e.g. Wimmer & Goswami, 1994). In the latter cases dyslexia is marked by a large number of reading errors. By contrast, in orthographically regular languages such as German or Italian children with dyslexia may read in a relatively correct fashion but their reading is characteristically slow and laborious (Wimmer, 1993; Zoccolotti et al., 1999). This pattern has been interpreted as reflecting weakness in the lexical procedure resulting in prevalent reference to the sub-lexical route (Zoccolotti et al., 1999).

Reliance on a sequential mode of processing is clear when vocal latencies to words are examined: children with dyslexia depend on word length while non-disabled children show a nearly flat function (De Luca, Barca, Burani & Zoccolotti, 2008; Judica, De Luca, Spinelli & Zoccolotti, 2002; Spinelli et al., 2005; Zoccolotti, De Luca, Judica & Spinelli, 2008). In non-disabled children this pattern is present only in the early learning phases (Zoccolotti, De Luca, Di Pace et al., 2005). Eye movement analysis in children with dyslexia confirmed the presence of slow, fragmented scanning when they read meaningful texts (De Luca, Di Pace, Judica, Spinelli & Zoccolotti, 1999) and lists of words and non-words (De Luca, Borrelli, Judica, Spinelli & Zoccolotti, 2002).

Although reading data are available for languages with regular orthographies, such as Italian, relatively little is known about writing deficits in these languages. In Italian (as in other relatively consistent orthographies) there is a certain degree of uncertainty regarding phoneme-to-grapheme correspondence. Some phonological strings have more than one possible orthographic solution, though only one is correct. For example, the phonemic group [kw] may be transcribed by the orthographic sequences QU, CU, or CQU; there is no definite rule for choosing among these alternatives and reference to a lexical entry is required.

All these cases of unpredictable spelling have been successfully used to assess lexical spelling in children with dyslexia (Angelelli, Judica, Spinelli, Zoccolotti & Luzzatti, 2004; Angelelli, Marinelli & Zoccolotti, 2010; Angelelli, Notarnicola, Judica, Zoccolotti & Luzzatti, 2010), thus confirming poor orthographic lexical knowledge and prevalent reliance on phoneme-to-grapheme processing also for the spelling process.

In recent years a number of studies have investigated the effect of training on the reading of Italian children with dyslexia (Judica et al., 2002; Lorusso, Facoetti & Molteni, 2004; Lorusso, Facoetti, Paganoni, Pezzani & Molteni, 2006; Lorusso, Facoetti, Toraldo & Molteni, 2005; Tressoldi, Lonciari & Vio, 2000; Tressoldi, Vio & Iozzino, 2007). Based on the interpretation that dyslexia is due to weakness of the lexical procedure leading to prevalent reliance on the use of the sub-lexical route (Zoccolotti et al., 1999), Judica et al. (2002) tried to foster parallel processing in reading by using a tachistoscopic presentation of stimuli. By presenting words for a shorter time than the minimum necessary to start a saccadic eye movement, this procedure impedes the sequential scanning of the visual target. By the end of the treatment, the children with dyslexia were faster and more accurate in reading both meaningful texts and lists of words than a group of untreated children. The latter group showed similar improvements when submitted to therapy a year later. The efficacy of tachistoscopic presentation in treating dyslexia was also confirmed by studies carried out using the model of dyslexia proposed by Bakker (1992), which uses a lateralized presentation. However, no clear effect of the side of presentation was consistently present, presumably indicating that the crucial aspect of the intervention was linked to the tachistoscopic presentation of stimuli *per se* (Lorusso et al., 2004, 2005, 2006). A procedure that focuses on the recognition of successive syllables within a word also proved effective, but only if the time devoted to each syllable was forced by the automatic procedure (Tressoldi et al., 2007).

Although these studies were based on different theoretical premises, they all emphasize the importance of time constraints in stimulus presentation in modulating the effectiveness of training. Notably, all of these studies examined seventh graders (Judica et al., 2002) or groups of mixed-age children (from second to eighth grade: Tressoldi et al., 2007; from seven to 16 years of age in Lorusso and co-workers' studies). Therefore, these data do not allow establishing the age at which rehabilitation training can be effectively started.

The aim of the present study was twofold. First, we wished to ascertain the efficacy of the tachistoscopic treatment program (Judica et al., 2002) in young children who are learning to read. In general, a reliable distinction can be made between proficient and impaired readers as early as third grade (e.g., Zoccolotti, De Luca, Di Pace et al., 2005). If treatment is effective at such an early age, it could have a critical impact on a child's overall scholastic achievement.

To better evaluate the improvement of decoding abilities, we used vocal reaction times (RT) to single word reading as well as standard reading tests. Vocal RTs provide an indication of decoding time independent of pronunciation time and are particularly sensitive in detecting the word length effect. This indicates the reliance of children with dyslexia on a sequential mode of processing, which is an indication of the prevalent use of the sublexical procedure (De Luca et al., 2008; Judica et al., 2002; Spinelli et al., 2005; Zoccolotti et al., 2008). Reading training aimed at fostering more global word analysis should reduce the length effect and encourage lexical processing.

The second aim of the study was to evaluate whether the reading treatment also has an effect on spelling performance. We were interested in evaluating this possibility because of the association between reading and writing deficits in Italian children with dyslexia (Angelelli et al., 2004; Angelelli et al., 2010). However, only a few studies have examined the spelling outcomes of reading interventions and they were carried out in opaque languages (Wanzek et al., 2006; Williams, Walker, Vaughn & Wanzek, 2016). Brunsdon, Hannan, Coltheart & Nickels (2002) made an in-depth single-case study of a ten-year-old English-speaking child who was suffering from severe mixed dyslexia and co-morbid spelling difficulties. The child underwent two 10-week treatment periods in which the highest frequency words that were read incorrectly at the baseline were trained by presenting flash cards to increase the child's visual word recognition skills. The authors found a significant treatment effect particularly in the spelling of the treated target words (only treated words were tested). In their meta-analysis, Wanzek et al. (2006; Williams et al., 2016) cited three other studies of English-speaking children that adopted reading intervention and included measures of spelling. However, in the first study (Torgesen et al., 2001), a moderate effect size (ES = .46) on a standardized spelling measure was reported in a group of secondary school students. The second study (Keel, Slaton & Blackhurst, 2001) examined the effect of reading training on infrequent words and their spelling as well as the spelling of other words in a single case: the effect was much larger for the studied items. However, as very few words were examined the results have to be interpreted with caution. Finally, the third study (Jitendra et al., 2004) examined the effects of reading intervention in two second grade participants with learning disabilities. Following the intervention, the performance of both participants improved on the spelling measure. However, in this study the intervention was an extensive and systematic program with explicit reading instructions in phonological awareness, phonics, fluency, vocabulary and comprehension.

Also in Italian children with dyslexia, Lorusso et al. (2004) reported unexpected spelling improvement on a standard spelling test following a tachistoscopic reading treatment, with items displayed centrally. In a subsequent study (age range: 7-15 years), spelling skills also improved following lateralized training. However, this outcome was not different from that observed after a reading treatment based on metaphonological tasks, perceptual prerequisites and word reading (Lorusso et al., 2006). Overall, the limited (and partially contradictory) evidence in Italian children indicates the importance of further evaluating the possible effect of reading treatment on spelling in this population. This conclusion parallels the considerations of Wanzek et al. (2006) based on their meta-analysis of studies on English-speaking children.

The generalization of reading treatment to spelling has interesting theoretical implications. The first studies that investigated the generalization of a lexical reading treatment to spelling involved adults with acquired impairment and they failed to demonstrate improvements in spelling (Scott & Byng, 1989; Weekes, 1996). However, generalization from reading to spelling (and viceversa) might be more effective early in development because of the greater interaction between the reading and spelling systems, as suggested by developmental theories (Ehri, 1997; Frith, 1985). Therefore, it would be useful to examine generalization by testing untreated items. Note that the results of the cited studies do not allow drawing conclusions about this. In fact, these studies either tested spelling of the same words used in the reading treatment (Brunsdon et al., 2002) or did not explicitly say whether the words in the assessment tasks were used during training (Lorusso et al., 2004; Lorusso et al., 2006; Torgesen et al., 2001). In Keel and coworkers' study (2001), untrained items were tested but training was too limited to be conclusive. According to Weekes (1996), generalization of a reading treatment to the reading and spelling of untreated items indicates a general effect on access to orthography, and thereby improved access and use of untreated word representations. To evaluate the generalization of treatment, we tested spelling with

words not used during the reading treatment.

Overall, we administered a reading training that was aimed at fostering global processing of single words to a group of third-grade children with dyslexia. Pre- and post-treatment examination included measures of text, word and non-word reading, vocal reaction times to single word presentation and spelling of regular and ambiguous words as well as non-words.

MATERIALS AND METHODS

Participants

The group of children with dyslexia included 10 third-graders (9 males, 1 female). Ages ranged from 8.0 to 8.7 years (Mean = 8.4; *SD* =.2 years). Criteria for inclusion in the sample of children with dyslexia were the following: a) marked reading delay on a standard reading test: performance 2 standard deviations (*SD*s) below the norms for either accuracy or speed on the MT *Reading test* (Cornoldi, Colpo & Gruppo MT, 1998); b) performance in the normal range on *Raven's Coloured Progressive Matrices* (above 10^{th} percentile = 19 for third grade; Pruneti et al., 1996); c) normal or corrected-to-normal visual acuity. Table 1 summarizes the main demographic and clinical characteristics of the sample. Note that mean group performance on the Vocabulary sub-test of the WISC-R was about average, indicating normal lexical ability from verbal input; similarly performance on Raven's test indicated average nonverbal skills.

The study was part of a research agreement between schools in Nettuno (near Rome) and the institutions the authors are affiliated with. As requested by the research agreement, all children screened who had a reading deficit were submitted to training; therefore, we did not have a control group with no treatment. The parents were given a description of the study and had to approve their child's participation. The study conformed to the standards of the Declaration of Helsinki and was approved by the Ethical Committee of the institutions the authors were affiliated with.

Study design

The study was carried out over a nine-month period. Performances of the children with dyslexia were measured pre-treatment in October. Reading performance was measured with two standard tests (*MT Reading test* and *Word and Non-word Reading test*) and the Vocal reaction times test; spelling ability with a standard instrument (DDO-2 *Spelling test*). Values of the pre-training evaluation of the standard reading and spelling tests are presented in Table 1 (for descriptions of these tests see Reading assessment and Spelling assessment, respectively).

Following the pre-training evaluation, the children underwent the reading treatment in their schools from November to May. A speech therapist administered the treatment program to each participant individually in a quiet room during school hours. We expected to carry out two one-hour sessions per week. However, for various reasons (holidays, special school activities and children's school absences), fewer sessions were actually performed. In practice, we were able to administer only 35 sessions to each child who participated over the seven-month training period. An effort was made to ensure that each child was submitted to the 35 sessions; in a few cases, this required additional training during the first days of June. In each session, the main treatment procedure was a computerized program featuring the tachistoscopic presentation of single words (see below).

The post-treatment evaluation was carried out in June using the same evaluation battery as in the pre-treatment assessment.

Reading assessment

- *Passage reading*: reading level was examined using a standard reading achievement test (*MT Reading test,* Cornoldi et al., 1998). The participant had to read aloud one passage within a 4-min time limit; speed (time in seconds per syllable read) and accuracy (number of errors, adjusted for the amount of text read) were scored. To measure comprehension, the participant read a second passage without a time limit and responded to ten multiple-choice questions. Stimulus materials (and related reference norms) varied according to grade; in the same grade, they were different at the beginning and the end of the year. Raw scores were converted to *z* scores according to standard reference data (Cornoldi et al., 1998).
- *Word and non-word reading*: three lists (each containing 30 items) from the *Word and Non-word Reading test* (Zoccolotti, De Luca, Di Filippo, Judica & Spinelli, 2005) were used: short (four-to-five letter) and long (eight-to-

Table 1 – Summary statistics for children with dyslexia at the assessment phase

		Raw scores		Standard scores	
		Mean	**SD**	**Mean**	**SD**
Age	Years	8.4	.2	–	–
Gender	Proportion M/F	9/1		–	–
Raven test	Correct responses (CR)	23.2	3.7	–	–
WISC-R	Vocabulary subtest	–	–	10.11	1.36
MT Reading test	Speed (s/syllable)	1.22	.51	–2.45	1.72
	Accuracy (errors)	21.3	4.7	–3.22	.91
	Comprehension (CR)	4.3	2.5	–.80	.92
Word and Non-word Reading test	Short words (s/item)	2.1	.7	–3.41	2.10
	Long words (s/item)	3.9	1.2	–4.12	2.09
	Short non-words (s/item)	2.5	.9	–2.38	1.78
	Short words (% errors)	19.0	12.3	–2.52	2.09
	Long words (% errors)	22.3	8.0	–2.28	1.20
	Short non-words (% errors)	38.7	13.6	–2.79	1.36
Spelling test	Regular words (CR/70)	55.9	8.9	–3.56	2.8
	Reg. words syll. conv. (CR/10)	6.0	1.9	–3.01	1.8
	Unpredictable words (CR/55)	29.7	7.3	–2.27	1.3
	Non-words (CR/25)	16.6	4.8	–3.25	2.3

Legenda. CR = correct responses.

Note. For gender, proportion of M/F is indicated. For all other reading and cognitive parameters, means (and SDs) for raw and standard scores are reported. For raw scores, specific parameters are indicated. For standard scores, *z*-values (with 0 +/–1) are reported in all cases except the Vocabulary test (where the expected mean is 10 +/–3).

nine letter) high frequency words, and short (four-to-five letter) non-words. Participants were asked to read each list of stimuli as quickly and accurately as possible. A short practice list was presented separately for words and non-words. The time needed to complete the task was measured separately for each list. The dependent measure was the reading time in seconds per item. Thus, we measured the time needed by the reader independently from accuracy. Errors were also measured during the test administration (1 = passed; 0 = failed). Omissions, insertions, reversals or substitutions of letters and wrong stress assignment were considered pronunciation errors. Also self-corrections (but not hesitations) were scored as errors. For off-line checks of reading times and errors the participant's vocal output was also tape-recorded. *Z*-scores (one value per grade) are available from a normative sample (Zoccolotti, De Luca, Di Filippo et al., 2005).

– V*ocal reaction times test*: vocal reaction times (RT) at stimulus onset were detected with a microphone and recorded using an Apple Performa computer. Both stimulus presentation and RT recording were controlled by SuperLab Pro 1.75 software. Stimuli were single words (black letters on a white background) of 2, 3, 4 and 5 letters (54 words for each length for a total of 216 stimuli as in Zoccolotti, De Luca, Di Filippo et al., 2005), displayed at the center of the computer screen. Each letter

subtended .4 degrees horizontally at a viewing distance of 57 cm. The word-length sets were matched on the initial phoneme and the median frequency value. This was 7105 in 10,000,000 occurrences, which indicates a generally high frequency value (VELI, 1989). Frequency did not vary among different word lengths (Kruskal Wallis H (4) = 2.82, n.s.). A fixation point was displayed for 750 ms; after that, a word was displayed and remained on the screen until the participant responded (otherwise, within a 6-sec. time limit, in case of no response). Then a 250 ms blank screen followed. One block of 10 practice stimuli and six experimental blocks of 36 stimuli were administered interspersed with brief pauses. Word length was randomized within each block.

The task was to read the words aloud as quickly and accurately as possible when they appeared on the screen. Vocal RTs to correctly read words were measured. Errors (incorrectly named words) were also computed. In a few instances, trials were not valid because of outside noise or technical failures.

Spelling assessment

The *Spelling test* (DDO-2, Angelelli et al., 2016) consists of four sections: Section A: regular words with complete one-sound-to-one-letter correspondence (N = 70); Section B: regular words requiring syllabic conversion rules (N = 10). Syllabic conversion is required when the orthographic realization of a consonant is determined by the vowel that follows it; Section C: words with unpredictable transcription along the phonological-to-orthographic conversion routine (e.g. [kwo] in [kwota], the quota: QUOTA and not *CUOTA) (N = 55); Section D: non-words with one-sound-to-one-letter correspondence (N = 25).

Each participant was tested individually. Words and non-words were presented in random order on separate lists. The examiner read each item aloud in a neutral tone. The child was asked to repeat each item before he/she wrote it to ensure he/she had perceived it correctly. No feedback was provided on the correctness of the responses. Spontaneous repairs were accepted as correct responses. The number of correctly spelled words in each category was the dependent measure. *Z*-scores (one value per year) are available from a normative sample. Spelling performances were available for 7 out of the 10 children studied.

Reading treatment

– *Computerized training*: training was controlled by the Tachistoscopio software (Morchio, Ott, Pesenti & Tavella, 1989). A single word (white letters on a blue background) was briefly presented (60 to 150 ms) in the center of the PC screen, followed by a mask (to prevent the support of iconic memory). The screen was set at a 45 cm viewing distance. Mean character width (center-to-center letter distance) was .6 degrees of visual angle.

  Lists of words from different categories (nouns, verbs, adjectives) were used. Word length varied from 2 to 6 letters and word frequency (De Mauro & Moroni, 1996) could be "very high" (from a pool of 2000 words including items at the core of the Italian language; hereafter, "core" words) or "high" (from a list of 3000 words used with high frequency in both speaking and writing).

  About 800 words were selected and used in the present study to generate 40 lists of 20 words (20 lists with "core" words and 20 with "high" frequency words). None of these stimuli were present in the *Word and Non-word reading test*, in the V*ocal reaction time* set of words, or in the *Spelling test* used for the assessment. From six to ten lists were administered in each session. Overall, about 6000 stimuli were administered to each child. On five to nine lists, the participant's task was to read the word aloud. On one list the task was to read it silently and print it on the keyboard. Treatment difficulty was adjusted individually for the children so that they started with stimuli yielding optimal performance (e.g. three-letter very high frequency words with a 150 ms presentation time). Then, we increased the difficulty of the computerized training by using shorter presentation times or longer and less frequent words. Manipulation of the materials and exposures was adjusted to keep the number of correct responses relatively high (60-70% in each session).

– *Additional training*: to facilitate reading, functional exercises were also given in each session (Judica, Baldoni, Chirri, Cucciaioni & Del Vento, 2006). The training is based on various game-like exercises (word reading, crossword puzzles, "memory" like exercises) that forced the children to practice the same stimuli in different contexts. Stimuli were words grouped in different categories (e.g. colors, animals, fruit, body parts etc.) of high frequency (De Mauro & Moroni, 1996) and appeared regularly in elementary textbooks (Marconi, Ott, Pesenti, Ratti & Tavella, 1994). For more details, see Judica et al. (2006).

Data analysis

Pre- and post-training measures from the reading assessment tests were used to evaluate the efficacy of the reading treatment program. Pre- and post-training measures on the spelling test were used to evaluate the effect of reading training on spelling performance.

Regarding the *MT Reading test*, *z*-scores of speed, accuracy and comprehension from pre- and post-treatment were submitted to a MANOVA with treatment (pre- and post) as repeated measure. Univariate tests for each dependent variable (speed, accuracy and comprehension) were also carried out. The presence of separate norms in the *MT Reading test* for the beginning and end of the school year allows directly evaluating treatment gains with respect to normal reading acquisition.

In the *Word and Non-word Reading test*, reading time and accuracy z-scores from pre- and post-treatment were entered in a MANOVA with treatment (pre- and post-) and type of stimulus (short words, long words and short non-words) as repeated measures. Univariate tests for both reading time and accuracy measures were also carried out and multivariate and univariate effect sizes are reported. Note that separate norms for the beginning and end of the year are unavailable for the *Word and Non-word Reading test*. Therefore, to evaluate the impact of treatment with respect to normal acquisition, the change in performance (in terms of raw values) as a function of treatment was graphically compared to the performances of non-disabled readers in a larger age range. Data of second (N = 40), third (N = 55) and fourth (N = 44) grade non-disabled readers from a previous study of our group (Zoccolotti, De Luca, Di Filippo, Judica & Martelli, 2009) were used.

Vocal RTs at onset (raw data in msec.) to correctly responded items were analyzed by an ANOVA with treatment (pre-, post-) and word length (two-, three-, four-, five-letter words) as repeated measures. Errors were also examined to evaluate a possible trade-off in performance. However, they were too few (percentage of errors < 1 in the control group) and variable from condition to condition to allow for parametric analysis. Data of children with dyslexia as a function of treatment (raw values) were graphically compared with those of a group of 28 non-disabled third-grade children who were tested with the same materials (Zoccolotti et al., 2005).

As for the *Spelling test*, *z*-scores were submitted to an ANOVA with treatment (pre-, post-) and stimulus category (regular words, regular words with syllabic conversion, unpredictable transcription words and non-words) as repeated measures. Separate norms were used for the beginning and the end of the year. The improvement of performance (total raw score) as a function of treatment was graphically compared to the performances of non-disabled children in second (N = 74), third (N = 110) and fourth (N = 136) grade (data from Angelelli et al., 2016).

Size effects for main effects and interactions were evaluated for both multivariate (i.e. multivariate eta-squared; Gall, Gall & Borg, 2011) and univariate analysis (i.e. partial eta-squared; Cohen, 1988). Conventional reference values for small, medium, and large effects are considered to be .01, .06 and .13, respectively. However, it has been proposed that empirical benchmarks can be identified for gauging effect sizes of the achievement outcomes of educational interventions. According to Bloom, Hill, Black & Lipsey (2008), interventions should be compared relative to the magnitudes of normal increases in performance in a given cognitive skill. Therefore, we used recently established normative values to determine the expected increase in reading/writing performances in the various tests used. Changes in performance passing from third to fourth grade were computed to establish normal annual increases (Bloom et al., 2008) and were used as benchmarks to evaluate the treatment effect (with the exception of the *Vocal reaction times test* for which only data for second and third grade were available).

RESULTS

Effect of training on reading performance

– *Passage reading*: at the pre-treatment evaluation, reading speed and accuracy were severely affected on the *MT Reading test*, with group means exceeding −2 *SD*; by contrast, comprehension was within normal values (Table 1).
The MANOVA on these data indicated a significant effect of treatment ($\lambda = .31$, $F_{(3, 7)} = 5.21$; $p<.05$; multivariate eta-squared = .44): performance improved from −2.15 to −1.11. In proficient readers, the mean performance increase passing from third to fourth grade was .38 (Tressoldi, 2008); therefore, the change observed after treatment in children with dyslexia was 174% of normal reading acquisition. The

univariate tests showed that the improvement regarded both reading accuracy (passing from −3.22 to −1.92; $F_{(1, 9)} = 10.81$; $p<.01$; partial eta-squared = .81) and speed (passing from −2.45 to −.99; $F_{(1, 9)} = 8.64$; $p<.05$; partial eta-squared = .49); however, comprehension was not affected even prior to treatment and did not show significant modification following the reading training (passing from −.80 to −.42; $F_{(1, 9)} = 2.75$; n.s).

– *Comments*: at the pre-treatment evaluation, reading speed and accuracy were severely affected, whereas comprehension was only mildly affected (Table 1). This pattern is common in Italian children with dyslexia (Judica et al., 2002). At the post-test, performance improved significantly for both accuracy and speed. By contrast, text comprehension did not improve as a function of training. Therefore, the treatment seemed to have a specific effect only on decoding skills. This general pattern seems consistent with the characteristics of the training, which emphasized practice in correctly identifying isolated words. The finding that performance improvement was present in reading meaningful passages indicates that improvement due to training on single words generalized to some extent to functional reading.

  The improved performance after treatment cannot be explained by the normal increase due to age/school attendance, because separate norms are available for this test for different periods during the school year. In fact, the effect of the training was 1.74 times larger than the normal annual improvement in performance on the MT test (note that the training lasted ca. seven months).

– *Word and non-word reading*: at the pre-treatment evaluation (Table 1) reading time and accuracy on the *Word and Non-word Reading* test were both severely affected for all stimulus categories.

  The MANOVA on reading time and accuracy *z*-scores indicated a main effect of treatment ($\lambda = .21$, $F_{(2, 8)} = 15.16$; $p<.01$; multivariate eta-squared = .54) and type of stimulus ($\lambda = .21$, $F_{(4, 6)} = 5.94$; $p<.05$; multivariate eta-squared = .55). The interaction effect of treatment x type of stimulus was also significant ($\lambda = .21$, $F_{(4, 6)} = 5.97$; $p<.05$; multivariate eta-squared = .55).

  The univariate test on reading times indicated a main effect of treatment ($F_{(1, 9)} = 18.52$; $p<.01$; partial eta-squared = .67): across stimulus materials, reading times improved from −3.30 to −1.70 *z*-values. The mean performance increase passing from third to fourth grade in non-disabled readers was .50 (Zoccolotti et al., 2009); therefore, the change following treatment in children with dyslexia was 134% which occurs in normal reading acquisition. The main effect of stimulus type was significant ($F_{(2, 18)} = 14.09$; $p<.0005$; partial eta-squared = .61): performance on long words (−3.13) was more impaired than performance on both short words (−2.36; $p<.01$) and short non-words (−2.01; $p<.001$); these two latter conditions did not differ from each other. The treatment x type of stimulus interaction was significant ($F_{(2, 18)} = 5.31$; $p<.05$; partial eta-squared = .37): post-hoc comparisons showed an effect of treatment over performance on short (passing from −3.41 to −1.31) and long (passing from −4.12 to −2.14) words (both $p<.001$). Performance on short non-words did not change significantly after treatment.

  The univariate test on accuracy showed the significance of the main effect of treatment ($F_{(1, 9)} = 11.49$; $p<.01$; partial eta-squared = .56): performance improved from −2.53 to −1.40 *z* values. In non-disabled readers, the mean performance increase passing from third to fourth grade was .19 (data from Zoccolotti et al., 2009); therefore, the change following treatment in children with dyslexia was 294% with respect to normal reading acquisition. The main effects of type of stimulus ($F_{(2, 18)} = 1.94$; n.s) and the treatment x type of stimulus interaction ($F_{(2, 18)} = .62$; n.s) were not significant.

  To graphically illustrate the effect of treatment versus normal reading acquisition, Figure 1 shows the performances of non-disabled readers in second, third and fourth grade (Zoccolotti et al., 2009) and of the present study's children with dyslexia. Reading time (upper graphs) is expressed as seconds per item; accuracy (lower graphs) is expressed as percentage of errors. Inspection of the figure indicates that, for reading time, the change in performance of children with dyslexia as a function of treatment, is steeper than that of non-disabled children for short and long words (but not non-words). In the case of accuracy, changes in performance during treatment are greater than the annual improvement of reading performance for all types of stimuli.

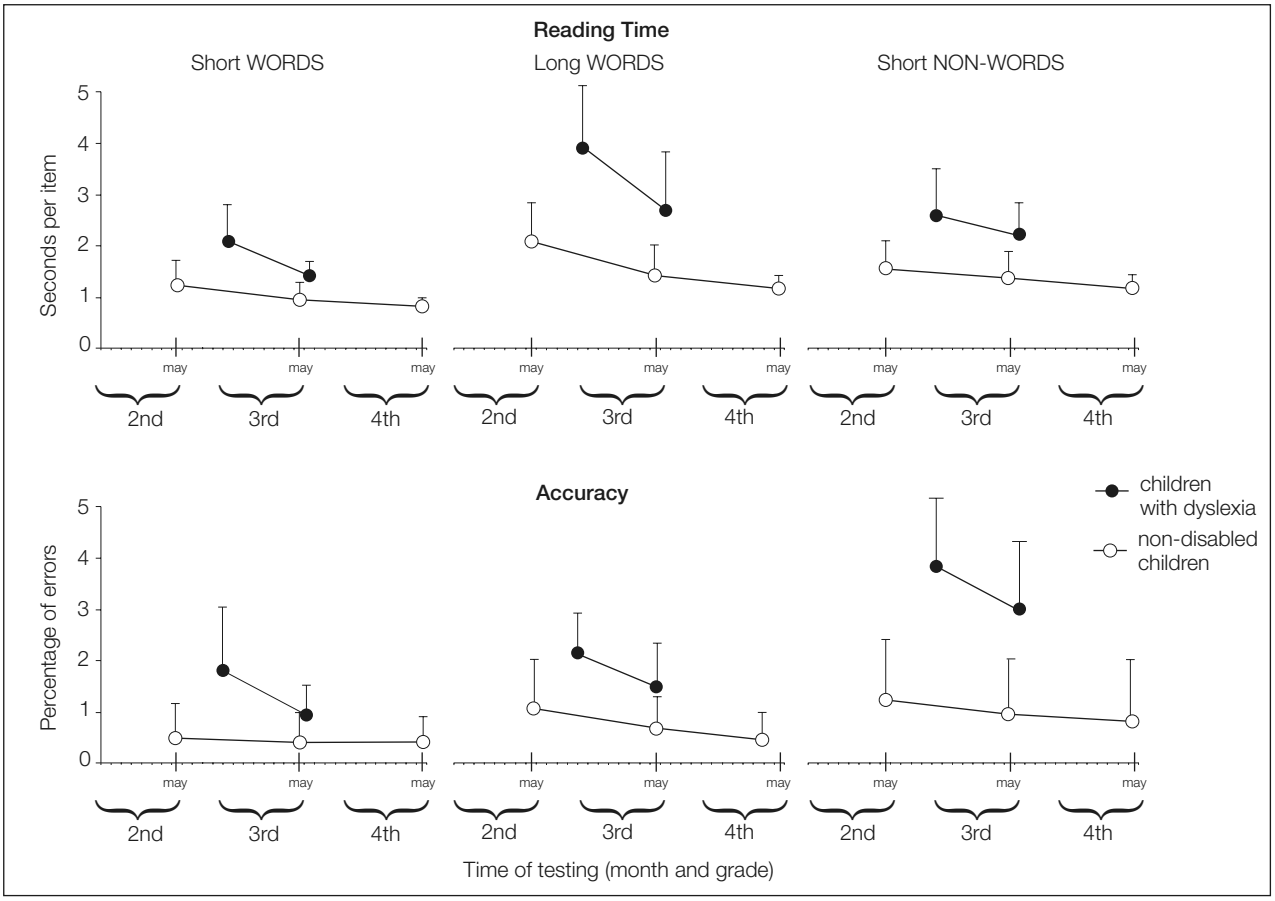
– *Comments*: before training the children were severely impaired in all reading measures. Non-word reading was not more affected than word reading (in fact, it was somewhat less in the case of reading time). This finding confirms previous observations of non-selective deficits in reading non-words in Italian children with dyslexia (e.g.,

Zoccolotti et al., 1999). A length effect was also evident for reading time, i.e., children with dyslexia were more severely impaired on long than on short words.

At the post-test, performance improved significantly both for reading time and accuracy. For reading time, the performance improvement was limited to words, whereas for accuracy it was present for both words and non-words. The graphical comparison with reading acquisition over a three-year period indicated that training had a larger effect than the normal annual increase in performance in this test. In the case of accuracy it must be noted that Italian readers reach relatively high accuracy in early grades (Zoccolotti et al., 2009) and show only modest increases in performance after third grade. Accordingly, children with dyslexia (who still make many errors) show an extremely large improvement in performance compared with what occurs in normal reading acquisition.

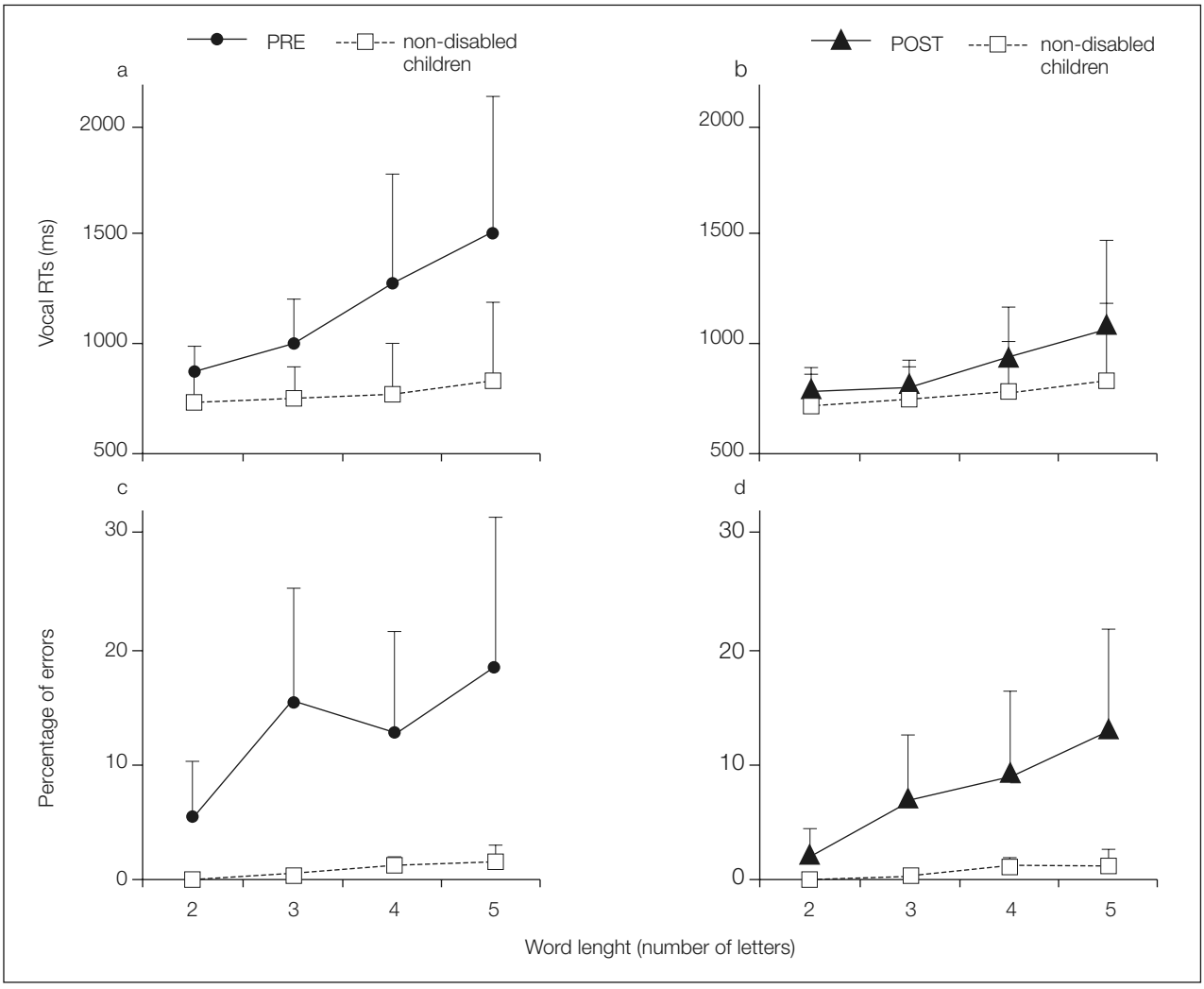
– *Vocal Reaction Times*: the vocal RTs in naming words for children with dyslexia (pre- and post-testing) and non-disabled children are presented at the top of Figure 2 (panels a and b) as a function of word length. The bottom part of the figure (panels c and d) shows a similar plot based on the percentage of errors. An inspection of the figure indicates that before treatment RTs increase with a steep slope as a function of stimulus length. The pattern of errors is similar to that of vocal RTs; therefore, differences in RTs cannot be easily interpreted in terms of a speed/accuracy trade-off. For the sake of comparison, data from a group of 28 non-disabled third-grade children who were tested with the same materials are also presented (Zoccolotti

Figure 1 - The pre- and post-treatment mean performances of children with dyslexia on the *Word and Non-word Reading test*



Note. Reading time (upper graphs) is expressed as seconds per item; accuracy (lower graphs) is expressed as percentage of errors. Bars depict upward SDs. Control data are based on performances of 40 non-disabled children in second grade, 55 in third grade and 44 in fourth grade (Zoccolotti et al., 2009).

Figure 2 – Mean pre- and post-treatment performances of children with dyslexia on the *Vocal reaction time test*



Note. Vocal reaction times are plotted as a function of the number of letters in a word (i.e., word length). For comparison, data of non-disabled peers (Zoccolotti et al., 2005) are also presented. Panels a and b: mean vocal RTs as a function of word length for children with dyslexia (a: pre-treatment; b: post-treatment) and non-disabled children. Panels c and d: mean percentage of errors as a function of word length for children with dyslexia (c: pre-treatment; d: post-treatment) and peer non-disabled children.

et al., 2005). After treatment, vocal RTs of children with dyslexia improve (and depend less on word length) but are still slower than those of non-disabled children.

The ANOVA showed a main effect of treatment ($F_{(1, 9)} = 7.09$, $p<.05$; partial eta-squared = .44): vocal RTs were shorter post- (922) than pre-treatment (1182). Information on the mean performance increase in non-disabled readers was available only from second to third grade (i.e., it was .41; data from Zoccolotti, De Luca, Di Pace et al., 2005); therefore, the change observed in children with dyslexia following treatment was 107% that of normal reading acquisition. The effect of word length was significant ($F_{(3, 27)} = 14.2$, $p<.001$; partial eta-squared = .61): vocal RTs increased with increasing word length. The treatment by word length interaction was only marginally reliable ($F_{(3, 27)} = 2.25$, $p = .10$; partial eta-squared = .20): inspection of Figure 2 indicates a trend for RTs of children with dyslexia to be less dependent upon word length after treatment.

– *Comments*: before treatment the vocal RTs of children with dyslexia were greatly influenced by word length. After treatment, these children significantly reduced their vocal RTs; however, their performance still partially reflected the influence of word length. The graphical comparison

showed that treatment had a slightly greater effect than the increase in performance passing from second- to third-grade. Note that changes in performance characteristically decrease with increasing age (Bloom et al., 2008); therefore, reference to this age change probably leads to an underestimation of the actual effect.

Naming isolated words with the instruction of being fast is a task somewhat similar to that used during the treatment, where words were tachistoscopically presented, and the child named the words (without a time constraint). Consequently, some of the improvement may be task specific. However, the words used in the RT task were different from those used during training; therefore, the changes in performance indicate improvements in word decoding not in item-specific learning.

– *Effect of training on spelling performance*: at the pre-treatment evaluation (Table 1), spelling performance was severely affected in all spelling categories.

The ANOVA showed a main effect of treatment ($F_{(1, 6)}$ = 7.73, p<.05; partial eta-squared = .56): accuracy improved across stimulus materials from −3.02 to −1.53. The mean performance increase in proficient readers passing from third to fourth grade is .46 (Angelelli et al., 2016); therefore, the change observed following treatment in children with dyslexia was 122% that of normal spelling acquisition. The main effect of type of stimulus and its interaction with treatment were not significant.

To evaluate the effect of treatment versus normal acquisition of writing skills, the spelling performance (total raw score) of children with dyslexia is presented in Figure 3 along with that of non-disabled children in second, third and fourth grade (Angelelli et al., 2016). Inspection of the figure indicates that the increase in the performance of children with dyslexia was greater than that observed in the non-disabled children.

– *Comments*: before training the children were severely impaired in all sub-sets of the spelling test. This is a common finding in children with dyslexia, who usually have spelling deficits (Angelelli et al., 2004). After the reading treatment the children's spelling performance generally improved but was still defective.

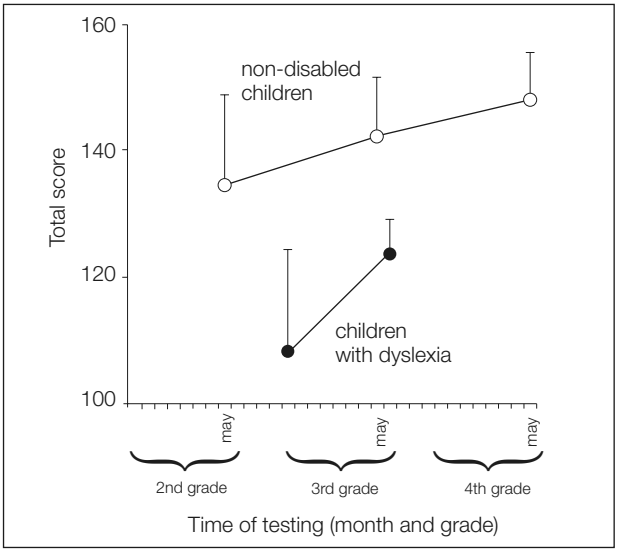
The graphical comparison with spelling acquisition over a three-year period indicates that the effect of training was larger than the normal annual increase in performance. Note that the words of the spelling test were not trained in the reading training.

DISCUSSION

Two main findings were observed. First, the present results show the efficacy of a tachistoscopic treatment program in improving the reading of children with dyslexia at an early stage of reading acquisition. Second, they indicate that a treatment program focussed on reading also has positive effects on spelling.

After treatment, the Italian children with dyslexia showed improved reading performance in terms of both speed and accuracy. Confirming previous work (Judica et al., 2002; Lorusso et al., 2004, 2005, 2006), the use of a tachistoscopic presentation proved helpful for children who are native speakers of a transparent language. Presenting words for an insufficient amount of time to sequentially scan the stimulus apparently improves the child's ability to analyze the stimulus as a whole. Improvement was evident not only in the recognition of words presently singly on a PC but also in reading meaningful texts and lists of words. Furthermore, it should be noted that stimuli used in the pre- and post-treatment assessment were not presented during treatment. Therefore, the changes after therapy indicate a general improvement in decoding skills rather than in stimulus specific learning. Overall, these findings confirm and extend previous evidence found in older children (Judica et al., 2002).

Figure 3 – Mean pre- and post-treatment total raw scores on the *Spelling test* for children with dyslexia



Note. Control data are based on performances of 74 non-disabled children in second grade, 110 in third grade and 136 in fourth grade (Angelelli et al. 2016).

To evaluate the efficacy of an intervention, the effect of treatment must be disentangled from normal reading acquisition expected as an effect of school attendance and general age changes. Particularly in the early school years non-disabled children improve their decoding skills quite rapidly. Children with dyslexia also improve in the absence of a specific treatment, although at a slower rate than non-disabled children (Tressoldi, Stella & Faggella, 2001). As an effect of this differential learning slope, the gap in performance between non-disabled children and children with dyslexia characteristically increases with age/reading experience (Tressoldi et al., 2001).

One way to evaluate treatment efficacy is to compare the experimental group with an untreated group of children with dyslexia. However, in a developmental perspective a control group of children with dyslexia who receive no treatment raises deontological problems. In our particular case, the research agreement with the schools required that all children with a reading deficit be given immediate treatment. Therefore, we evaluated treatment efficacy in comparison with the changes observed as an effect of normal reading acquisition. In particular, as suggested by Bloom et al. (2008), we compared the improvement in performance due to the treatment to that observed in non-disabled children over a year. For all reading materials (i.e. meaningful texts, lists of words and single word naming) the effect of treatment over a seven-month period was larger than the performance improvement obtained in non-disabled children passing from third to fourth grade. The proportion varied from 122% in the case of the spelling test to 294% in the case of the *Word and Non-word Reading test* (reading accuracy). The improvement in performance was 107% in the case of the vocal RTs to single words, where only norms from second and third grade were available. Note that using normal reading acquisition as a benchmark produces a conservative estimate (i.e., underestimation) of the treatment effectiveness. In fact, as stated above, in the absence of specific treatment children with dyslexia improve at a much slower rate than non-disabled children (Tressoldi et al., 2001). Therefore, the observed size effects indicate substantial improvements as a function of the rehabilitation treatment. We also graphically compared the change in performance as a function of treatment to the performances of non-disabled readers over a larger age range. These comparisons also indicated that the improvement in children with dyslexia clearly exceeded that shown by non-disabled children.

Overall, these findings indicate that substantial treatment effects can be demonstrated at a relatively early stage of reading acquisition. Of course, intervening at earlier stages is preferable because it helps prevent a gap in scholastic achievement. Furthermore, when necessary, follow-up interventions can be more easily programmed.

One important finding was the generalization to spelling skills of the treatment program aimed at developing the lexical route for reading. The present results are consistent with the literature which reports a significant reading treatment effect generalizing to spelling of untreated target stimuli in children (Lorusso et al., 2004; Lorusso et al., 2006; Torgesen et al., 2001). Moreover, while improved spelling of target stimuli after reading treatment is not surprising (because a child may gradually build up correct orthographic representations after several attempts to read the correctly spelled item), generalization to untrained items could be ascribed to treatment effects on the general process of entering the word recognition system (Weekes, 1996). Overall, these results support the reciprocal relationship between the processing of reading and spelling as proposed by developmental models of literacy acquisition (e.g. Ehri, 1997; Frith, 1985).

A number of limitations of the present study must also be mentioned. Because of the relatively small size of the sample tested (as well as the absence of multiple assessments) over the course of training and of a control group of children without treatment, the present observations must be confirmed in future studies. Furthermore, it should be noted that improvement after treatment was incomplete. The performance of children with dyslexia remained below the means of non-disabled children for all parameters. Also, some evidence of the persistence of sequential analysis in reading was detected in reading lists of short and long words and in the length effect in the case of vocal RTs to single words. Further work is needed to determine whether the partial recovery shown by these children will hold (or further improve) in time or whether they will require further treatment periods. Finally, as stated above, we compared the effects of training with respect to the performance changes observed during normal reading acquisition. To this aim, we referred to established (but different) data sets in the literature. Referring to data from different sources is certainly a limitation of the present study. Therefore, caution must be taken in generalizing the present results. Future research should consider the possibility of gathering data on typical development across tests within a single sample of children.

CONCLUSIONS

Overall, the results indicate the effectiveness of a reading treatment that focused on the global decoding of words by using a tachistoscopic presentation of stimuli; an additional result was the presence of some generalization to spelling. Treatment can be successfully started in third graders who already show a significant lag in the acquisition of written language.

Acknowledgements
We would like to thank teachers and principals of the primary schools in Nettuno for their help in carrying out the intervention project. We would also like to thank Claire Montagna for editing the manuscript and Dr. Chiara Valeria Marinelli for advice in carrying out the statistical analyses.

References

ANGELELLI, P., JUDICA, A., SPINELLI, D., ZOCCOLOTTI, P. & LUZZATTI, C. (2004). Characteristics of writing disorders in Italian dyslexic children. *Cognitive and Behavioral Neurology*, *17*(1),18-31.

ANGELELLI, P., MARINELLI, C.V., IAIA, M., NOTARNICOLA, A., COSTABILE, D., JUDICA, A., ZOCCOLOTTI, P. & LUZZATTI, C. (2016). *DDO-2 Diagnosi dei disturbi ortografici in età evolutiva*. Trento: Erickson.

ANGELELLI, P., MARINELLI, C.V. & ZOCCOLOTTI, P. (2010). Single or dual orthographic representations for reading and spelling? A study on Italian dyslexic and dysgraphic children. *Cognitive Neuropsychology*, *27*, 305-333.

ANGELELLI, P. NOTARNICOLA, A., JUDICA, A., ZOCCOLOTTI, P. & LUZZATTI, C. (2010). Spelling impairment in Italian dyslexic children: Does the phenomenology change with age? *Cortex, 46*, 1299-1311.

BAKKER, D.J. (1992). Neuropsychological classification and treatment of dyslexia. *Journal of Learning Disabilities*, *25*(2), 102-109.

BLOOM, H.S., HILL, C.J., BLACK, A.R. & LIPSEY, M.W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *MDRC Working Papers on Research Methodology*.

BRUNSDON, R., HANNAN, T.J., COLTHEART, M. & NICKELS, L. (2002). Treatment of lexical processing in mixed dyslexia: A case study. *Neuropsychological Rehabilitation*, *12*(5), 385-418.

CORNOLDI, C., COLPO, G. & GRUPPO MT (1998). *Prove di Lettura MT per la scuola elementare - 2. Manuale*. Firenze: Giunti O.S. Organizzazioni Speciali.

DE LUCA, M., BARCA, L., BURANI, C. & ZOCCOLOTTI, P. (2008). The effect of word length and other sublexical, lexical and semantic variables on developmental reading deficit. *Cognitive and Behavioral Neurology*, *21*(4), 227-235.

DE LUCA, M., BORRELLI, M., JUDICA, A., SPINELLI, D. & ZOCCOLOTTI, P. (2002). Reading words and pseudo-words: An eye movement study of developmental dyslexia. *Brain and Language*, *80*(3), 617-626.

DE LUCA, M., DI PACE, E., JUDICA, A., SPINELLI, D. & ZOCCOLOTTI, P. (1999). Eye movement patterns in linguistic and non-linguistic tasks in developmental surface dyslexia. *Neuropsychologia*, *37*(12), 1407-1420.

DE MAURO, T. & MORONI, G.G. (1996). DIB. *Dizionario di base della lingua italiana*. Torino: Paravia.

EHRI, L. (1997). Learning to read and learning to spell are one and the same, almost. In C. Perfetti, L. Rieben & M. Fayol (Eds.), *Learning to spell: Research, theory and practice*. Mahwah, NJ: Erlbaum.

FRITH, U. (1985). Beneath the surface of developmental dyslexia. In K. Patterson, J.L. Marshall & M. Coltheart (Eds.), *Surface Dyslexia*. Hillsdale, NJ: Erlbaum.

GALL, M.D., GALL, J.P. & BORG, W.R. (2011). *Applying Educational Research, 6th ed.* Boston, MA: Pearson.

JITENDRA, A.K., EDWARDS, L.L., STAROSTA, K., SACKS, G., JACOBSON, L.A. & CHOUTKA, C.M. (2004). Early reading instruction for children with reading difficulties: Meeting the needs of diverse learners. *Journal of Learning Disabilities*, *37*(5), 421-439.

JUDICA, A., BALDONI, L., CHIRRI, L., CUCCIAIONI, C. & DEL VENTO, G. (2006). *Parole in corso. Materiali per il recupero delle difficoltà di lettura*. Trento: Erickson.

JUDICA, A., DE LUCA, M., SPINELLI, D. & ZOCCOLOTTI, P. (2002). Training of developmental surface dyslexia modifies performance and eye fixation duration in reading. *Neuropsychological Rehabilitation*, *12*(3), 177-197.

KEEL, M.C., SLATON, D.B. & BLACKHURST, A.E. (2001).

Acquisition of content area vocabulary for student with learning disabilities. *Education and Treatment of Children*, *24*(1), 46-71.

LORUSSO, M.L., FACOETTI, A. & MOLTENI, M. (2004). Hemispheric, attentional, and processing speed factors in the treatment of developmental dyslexia. *Brain and Cognition*, *55*(2), 341-348.

LORUSSO, M.L., FACOETTI, A., PAGANONI, P., PEZZANI, M. & MOLTENI, M. (2006). Effects of visual hemisphere-specific stimulation versus reading-focused training in dyslexic children. *Neuropsychological Rehabilitation*, *16*(2), 194-212.

LORUSSO, M.L., FACOETTI, A., TORALDO, A. & MOLTENI, M. (2005). Tachistoscopic treatment of dyslexia changes the distribution of visual-spatial attention. *Brain and Cognition*, *57*(2), 135-142.

MARCONI, L., OTT, M., PESENTI, E., RATTI, D. & TAVELLA, M. (1994). *Lessico elementare. Dati statistici sull'italiano scritto e letto dai bambini delle elementary*. Bologna: Zanichelli.

MORCHIO, B., OTT, M., PESENTI, E. & TAVELLA, M. (1989). *Tachistoscopio. Un programma per migliorare l'abilità di lettura di parole. II edizione (CNR ITD Genova)*. Bologna: A.S.P.H.I.

PRUNETI, C., FENU, A., FRESCHI, G., ROTA, S., COCCI, D., MARCHIONNI, M., ROSSI, S. & BARACCHINI-MORATORIO, G. (1996). Aggiornamento della standardizzazione italiana del test delle Matrici Progressive Colorate di Raven (CPM). *Bollettino di Psicologia Applicata*, *217*, 51-57.

SCOTT, C. & BYNG, S. (1989). Computer assisted remediation of a homophone comprehension disorder in surface dyslexia. *Aphasiology*, *3*(3), 301-320.

SHARE, D.L. (2008). On the anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. P*sychological Bulletin*, *134*(4), 584-615.

SPINELLI, D., DE LUCA, M., DI FILIPPO, G., MANCINI, M., MARTELLI, M. & ZOCCOLOTTI, P. (2005). Length effect in word naming latencies: Role of reading experience and reading deficit. *Developmental Neuropsychology*, *27*(2), 217-235.

TORGESEN, J.K., ALEXANDER, A.W., WAGNER, R.K., TASHOTTE, C.A., VOELLER, K.K. & CONWAY, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities*, *34*(1), 33-58.

TRESSOLDI, P.E. (2008). I brani della batteria MT si possono leggere tutti con la stessa velocità? Norme trasversali dal secondo anno della primaria al terzo della secondaria di primo grado. *Dislessia*, *5*, 339-345.

TRESSOLDI, P.E., LONCIARI, I. & VIO, C. (2000). Treatment of specific developmental reading disorders, derived from single- and dual-route models. *Journal of Learning Disabilities*, *33*(3), 278-285.

TRESSOLDI, P.E., STELLA, G. & FAGGELLA, M. (2001). The development of reading speed in Italians with dyslexia: A longitudinal study. *Journal of Learning Disabilities*, *34*(5), 414-417.

TRESSOLDI, P.E., VIO, C. & IOZZINO, R. (2007). Efficacy of an intervention to improve fluency in children with developmental dyslexia in a regular orthography. *Journal of Learning Disabilities*, *40*(3), 203-209.

WANZEK, J., VAUGHN, S., WEXLER, J., SWANSON, E.A., EDMONDS, M.E. & KIM, A. (2006). A synthesis of spelling interventions and their effects on the spelling outcomes of student with LD. *Journal of Learning Disabilities*, *39*(6), 528-543.

WEEKES, B. (1996). Surface dyslexia and surface dysgraphia: Treatment studies and their theoretical implications. *Cognitive Neuropsychology*, *13*(2), 277-315.

WILLIAMS, K.J., WALKER, M.A., VAUGHN, S. & WANZEK, J. (2016). A synthesis of reading and spelling interventions and their effects on spelling outcomes for students with learning disabilities. *Journal of Learning Disabilities*, doi: 10.1177/0022219415619753.

WIMMER, H. (1993). Characteristics of developmental dyslexia in a regular writing system. *Applied Psycholinguistics*, *14*(1), 1-33.

WIMMER, H. & GOSWAMI, U. (1994). The influence of orthographic consistency on reading development: Word recognition in English and German children. *Cognition*, *51*(1), 91-103.

ZOCCOLOTTI, P., DE LUCA, M., DI PACE, E., JUDICA, A., ORLANDI, M. & SPINELLI, D. (1999). Markers of developmental surface dyslexia in a language (Italian) with high grapheme-phoneme correspondence. *Applied Psycholinguistics*, *20*(2), 191-216.

ZOCCOLOTTI, P., DE LUCA, M., DI FILIPPO, G., JUDICA, A. & SPINELLI, D. (2005). *Prova di lettura di parole e non parole*. Roma: IRCCS Fondazione Santa Lucia, http://www.hsantalucia.it/modules.php?name=Content&pa=showpage&pid=1032

ZOCCOLOTTI, P., DE LUCA, M., DI PACE, E., GASPERINI, F., JUDICA, A. & SPINELLI, D. (2005). Word length effect in early reading and in developmental dyslexia. *Brain and Language*, *93*(3), 369-373.

ZOCCOLOTTI, P., DE LUCA, M., JUDICA, A. & SPINELLI, D. (2008). Isolating global and specific factors in developmental dyslexia: A study based on the rate and amount model (RAM). *Experimental Brain Research*, *186,* 551-560.

ZOCCOLOTTI, P., DE LUCA, M., DI FILIPPO, G., JUDICA, A. & MARTELLI, M. (2009). Reading development in an orthographically regular language: Effects of length, frequency, lexicality and global processing ability. *Reading and Writing*, *22*(9), 1053-1079.

An Italian Adaptation of the Entrepreneurial Passion Scale

Monica Molino[1], Valentina Dolce[2], Claudio G. Cortese[2], Chiara Ghislieri[2]

[1] Psychology Department, University of Turin & Human Plus Foundation, Turin
[2] Psychology Department, University of Turin

ABSTRACT. Lo studio presenta l'adattamento della versione italiana a 10 item della scala di *Passione Imprenditoriale* proposta da Cardon, Gregoire, Stevens e Patel (2013), che distingue tre dimensioni: *fondare*, *inventare* e *sviluppare*. La ricerca, che ha coinvolto un campione di 535 imprenditori e 638 aspiranti imprenditori, ha confermato la soluzione a tre fattori presente nella versione originale. La scala può avere utili risvolti sia per ricerche, sia per interventi formativi e di sviluppo nel campo dell'imprenditorialità.

SUMMARY. Introduction: Entrepreneurial Passion (EP) is considered a multidimensional construct relevant for entrepreneurial intention and performance. The EP Scale proposed by Cardon, Gregoire, Stevens & Patel (2013) distinguishes passion for founding, for inventing and for developing. The study presents the adaptation of the Italian 10-item version of the scale. Methods: A group of 535 entrepreneurs and 638 aspiring entrepreneurs completed a self-report questionnaire. The factorial validity of the Italian EP Scale has been tested through confirmatory factor analysis (CFA) and multigroup CFA using MPlus 7. Correlations were performed to test the concurrent validity of the scale. Results: CFA confirmed the three-factor structure of the scale. The measures of EP in the three domains showed a good internal consistency and significantly correlated with constructs related to EP in literature. Conclusions: This study contributes to validate the Italian version of the instrument. It could be a useful support for researchers and practitioners in the field of entrepreneurship.

Keywords: *Entrepreneurial passion, Entrepreneurship, Scale adaptation*

INTRODUCTION

Given the economic and social changes in the recent years, entrepreneurship has received an increasing interest in different fields of research and practice. Despite this increasing interest for the topic, there is still a gap in the Work and Organization Psychology literature; in particular, there is not a clear agreement around the psychological characteristics that are considered predictive of entrepreneurial performance (Rauch & Frese, 2007). Moreover, in Italy, the attention for this topic and for start-up phenomenon started later compared with other countries (Oehler, Pukthuanthong, Rummer & Thomas, 2007) and little attention has been given to the study of psychological dimensions involved in entrepreneurial

processes (Cubico, Bortolani, Favretto & Sartori, 2010; Molino, Cortese, Carpaneto, Mercuri & Ghislieri, 2015). For these reasons, the study of entrepreneurship in the Italian context is important and it is crucial to provide researchers and practitioners with tools validated in the Italian language. This study contributes to fill this gap through the validation of the Italian version of the Entrepreneurial Passion Scale (Cardon et al., 2013), involving a group of Italian entrepreneurs and aspiring entrepreneurs.

In recent years, the academic literature has shown a growing interest in different concepts related to entrepreneurship; entrepreneurial passion (EP) is one of them (Cardon et al., 2013). As suggested by scholars, the general notion of passion has a long intellectual history, since it has drawn attention of several great philosophers, such as Aristotle, Spinoza, Descartes, Hegel, Machiavelli (Cardon, Wincent, Singh & Drnovsek, 2009; Vallerand et al., 2003).

Rather, only recently psychologists have been interested in the concept of passion; especially, psychologists usually pay attention to its motivational aspects and identity meaning (Cardon et al., 2009; Vallerand et al., 2003). For instance, according to Vallerand et al. (2003), passion is able to enhance motivation, improve mental activity and provide meaning in everyday work. They argued that passion is "a strong inclination toward an activity that people like, that they find important, and in which they invest time and energy" (Vallerand et al., 2003, p.757).

Baum & Lock (2004) referred to passion for work as "love for work"; in their study passion was measured in terms of emotions of love, longing and attachment. Passion was described also as "enthusiasm, joy, and even zeal that come from the energetic and unflagging pursuit of a worthy, challenging, and uplifting purpose" (Smilor, 1997, p. 342). Cardon et al. (2009) defined entrepreneurial passion as "consciously accessible intense positive feelings experienced by engagement in entrepreneurial activities with roles that are meaningful and salient to the self-identity of the entrepreneur" (p. 517). In other words, according to this definition, two dimensions contributes to define the construct: the intense positive feelings and the identity centrality (Cardon et al., 2009; Cardon et al., 2013). These two dimensions are conceptually and empirically distinct from each other (Cardon et al., 2013). The positive and intense feelings towards entrepreneurial activities occur over time, they are enduring rather than momentary (Cardon et al., 2013).

In Cardon and collegues' theorization, feelings and identity centrality are explored towards three different domains of entrepreneurial tasks and activities: inventing new products and services, founding new organizations and developing organizations beyond their initial survival and success (Cardon et al., 2009; Cardon et al., 2013).

Passion for inventing concerns activities associated with search for new opportunities or solutions to relevant needs and problems; passion for founding is related to wish for identifying the necessary financial, human and social resources in order to create the venture; at the end, passion for developing is related to the enterprise's growth over time (Cardon et al., 2013).

Passion is recognized as a key element of entrepreneurship (Cardon et al., 2009; Cardon et al., 2013), because it can enhance ability to raise funds from venture capitalists, it fuels motivation of employees, further it is associated with characteristics of creativity, persistence and absorption of the entrepreneur.

Despite passion has been said an important aspect of entrepreneurship, there is still a gap in understanding the impact of EP with systematic empirical evidence (Cardon et al., 2013). In order to fill this gap, Cardon et al. (2013) conducted a series of studies to develop and validate a scale for measuring EP. In line with theoretical arguments, the Authors developed items to capture the dimensions of intense positive feelings and identity salience experienced by the entrepreneur, focused on domains of inventing, founding and developing.

In its original version, EP Scale consists of 13 items, 4 of which evaluate intense positive feelings for inventing (e.g. "It is exciting to figure out new ways to solve unmet market needs that can be commercialized"), 3 items measure intense positive feelings for founding (e.g. "Establishing a new company excites me"), 3 items evaluate intense positive feelings for developing (e.g. "I really like finding the right people to market my product/service to"), and finally 3 items measure identity centrality of the three domains to the person (e.g. "Inventing new solutions to problems is an important part of who I am").

The aim of the present study is to explore the psychometric properties of the 10-item Italian version of the EP Scale, which measures the intense positive feelings towards the domains of inventing, founding and developing and do not consider the 3 items that capture the identity centrality of the three domains. Only one item to detect the identity centrality of each domain could be a limit of the instrument, therefore

we decided to focus only on the intense positive feelings as dimension of passion in the three domains.

In the first step, the factorial validity of the 10-item scale was examined through a confirmatory factor analysis (CFA), in order to verify the three factors structure. In the second step, correlations between EP in the three domains and other constructs that are expected to correlate with it were investigated, in order to test the convergent validity of the Italian version of the scale.

METHODS

Participants

The group included 1173 participants (65% male and 35% female). The ISTAT Report (2016) showed that the 23% of the Italian labour force are self-employed persons or entrepreneurs; among them, 70% are male and 30% are female.

The first subgroup included 535 entrepreneurs and startuppers (68% male and 32% female; M_{age} = 35.46, *SD* = 8.65; 19% with high school diploma, 48% with bachelor or postgraduate degree and 28% with PhD or Master degree). The second one consisted of 638 aspiring entrepreneurs and startuppers (62% male, 32% female; M_{age} = 28.53, *SD* = 6.44; 30% with high school diploma, 54% with bachelor or postgraduate degree and 11% with PhD or Master degree).

Measures

The Italian version of the EP Scale consists of 10 items selected by the original 13-item Cardon and colleagues' EP Scale (2013). Items are divided into three domains: inventing (4 items), founding (3 items) and developing (3 items). Items were translated from English to Italian by an Italian native-speaker researcher. Then, an expert English mother tongue made a back-translation from Italian into English. The result was a good correspondence between items. Respondents were asked to indicate their agreement by means of a 7-point Likert scale (1 = very strongly disagree, 7 = very strongly agree), according to Authors who recommended to use 7- or 9-point scales in order to guard against issues of range restriction (Cardon et al., 2013). Table 1 shows the original version and the Italian version of the EP Scale.

The other measures used in this study, all rated on a 7-point Likert scale (1 = very strongly disagree, 7 = very strongly agree), were the following.

- *Self-efficacy*: measured by 10 items (Schuler, Thornton, Frintrup & Mueller-Hanson, 2002); an example item is "I am confident that I will succeed". The CFA performed on the whole group confirmed the 1-factor structure for the Italian version of the scale: $\chi^2_{(35)} = 348.49$, $p<.001$; CFI = .91; TLI = .90; RMSEA = .08 (.05, .08); SRMR = .04. The factor loadings ranged from .48 to .69. Cronbach Alpha in this study was .86.
- *Internal locus of contro*l: measured by 6 items (Argentero & Vidotto, 1994); an example item is "There is a direct relation between a person's skills and her/his role". Cronbach Alpha was .81.
- *Resiliency*: measured by 10 items (Di Fabio & Palazzeschi, 2012); an example item is "I tend to bounce back after illness, injury, or other hardships". Cronbach Alpha was .81.
- *Need for achievement*: measured by 10 items (Borgogni, Petitta & Barbaranelli, 2004); an example item is "Difficult and challenging activities attract me". Cronbach Alpha was .88.
- *Risk propensity*: measured by 10 items (Dahlbäck, 1990; Vecchione & Barbaranelli, 2005); an example item is "I often dare to do risky things which other people are reluctant to do". Cronbach Alpha was .72.
- *Entrepreneurial intention*: measured by 5 ad-hoc items adapted from Linan & Chen work (2009); an example item is "My professional aim is to become an entrepreneur". The CFA confirmed a 1-factor structure for the Italian scale: $\chi^2_{(5)} = 55.98$, $p<.001$; CFI = .95; TLI = .91; RMSEA = .07 (.04, .08); SRMR = .04. The factor loadings ranged from .63 to .76. Cronbach Alpha was .81.

Procedure

Participants completed the online self-report questionnaire STEPS (STartuppers and Entrepreneurs Potential Survey) in the context of a project aimed at evaluating and developing the entrepreneurial potential of entrepreneurs and aspirants, promoted by an Italian no profit organization (Human Plus Foundation). The voluntary and not paid participation to the research, and the anonymity and confidentiality of the data were emphasized. We obtained informed consent by participants.

Table 1 – Entrepreneurial Passion Scale item

	Original items	Italian translations
	Passion for inventing	
1.	It is exciting to figure out new ways to solve unmet market needs that can be commercialized.	*Mi piace trovare nuovi modi per soddisfare esigenze del mercato che possono essere commercializzate.*
2.	Searching for new ideas for products/services to offer is enjoyable to me.	*È divertente cercare nuove idee di prodotti/servizi da offrire.*
3.	I am motivated to figure out how to make existing products/services better.	*Sono motivato a trovare modi per rendere migliori i prodotti/servizi esistenti.*
4.	Scanning the environment for new opportunities really excites me.	*Mi piace analizzare il mercato in cerca di nuove opportunità.*
	Passion for founding	
5.	Establishing a new company excites me.	*Penso che fondare una nuova azienda sia entusiasmante.*
6.	Owning my own company energizes me.	*È stimolante essere il proprietario di un'azienda.*
7.	Nurturing a new business through its emerging success is enjoyable.	*È piacevole coltivare un nuovo business emergente.*
	Passion for developing	
8.	I really like finding the right people to market my product/service to.	*Mi piace molto individuare le persone giuste con le quali commercializzare un prodotto/servizio.*
9.	Assembling the right people to work for my business is exciting.	*È stimolante mettere insieme le persone giuste con le quali lavorare al business che mi interessa.*
10.	Pushing my employees and myself to make our company better motivates me.	*È motivante spingere i miei collaboratori e me stesso a rendere la nostra azienda migliore.*
	Likert frequency scale from 1 = *strongly disagree to* 5 = *strongly agree*	*Scala di risposta Likert da 1 = per niente d'accordo a 7 = del tutto d'accordo*

Statistical Analysis

In order to assess the factorial validity of the Italian EP Scale, a confirmatory factor analysis (CFA) and a multigroup CFA were performed using the software Mplus 7. The CFA method of estimation was maximum likelihood (ML). According to the literature, several goodness-of-fit criteria were considered: the χ^2 goodness-of-fit statistic; the Root Mean Square Error of Approximation (RMSEA); the Comparative Fit Index (CFI); the Tucker Lewis Index (TLI); the Standardized Root Mean Square Residual (SRMR).

As measures of reliability, Cronbach's Alpha and the corrected item-total correlation coefficients were calculated for the three domains. Finally, in order to investigate the concurrent validity, also correlations between EP in the three domains and other constructs indicated in literature as potentially related to EP were tested: entrepreneurial intention, self-efficacy, internal locus of control, resiliency, need for achievement, risk propensity (e.g. Cardon et al., 2013; Molino et al., 2015).

RESULTS

Confirmatory factor analysis

As shown in Table 2, CFA for different factor models were tested in the whole group. The best fitting model resulted for the three-factor structure of the scale, in line with the original version (Cardon et al., 2013).

Figure 1 shows the standardized three-factor solution, with factor loadings ranging from .60 to .73 for passion for inventing, from .64 to .79 for passion for founding, and from .55 to .71 for passion for developing.

A multigroup CFA was conducted to test the measurement model also across the two different subgroups, entrepreneurs and aspiring entrepreneurs. The model showed a good fit to the data: $\chi^2_{(85)} = 369.67$, $p<.001$; CFI = .92; TLI = .91; RMSEA = .06 (.05, .08); SRMR = .04. In entrepreneurs subgroup, factor loadings ranged from .44 to .66 for passion for inventing, from .68 to .79 for passion for founding, and from .42 to .59 for passion for developing. In aspirants subgroup, factor loadings ranged from .49 to .74 for passion for inventing, from .64 to .73 for passion for founding, and from .67 to .70 for passion for developing.

Reliability

Internal consistency of the measures was good, since all values of Cronbach's Alpha met the criterion of .70: as Table 3 shows, Alpha coefficient was .72 for passion for inventing, .80 for passion for founding and .73 for passion for developing. The alpha-if-item-were-deleted values revealed that any item decreased the Alpha value in all the cases. Moreover, corrected item-total correlations ranged from .46 to .56 for passion for inventing, from .62 to .67 for passion for founding and from .52 to .62 for passion for developing. Therefore, all scores were above the cut-off value of .40.

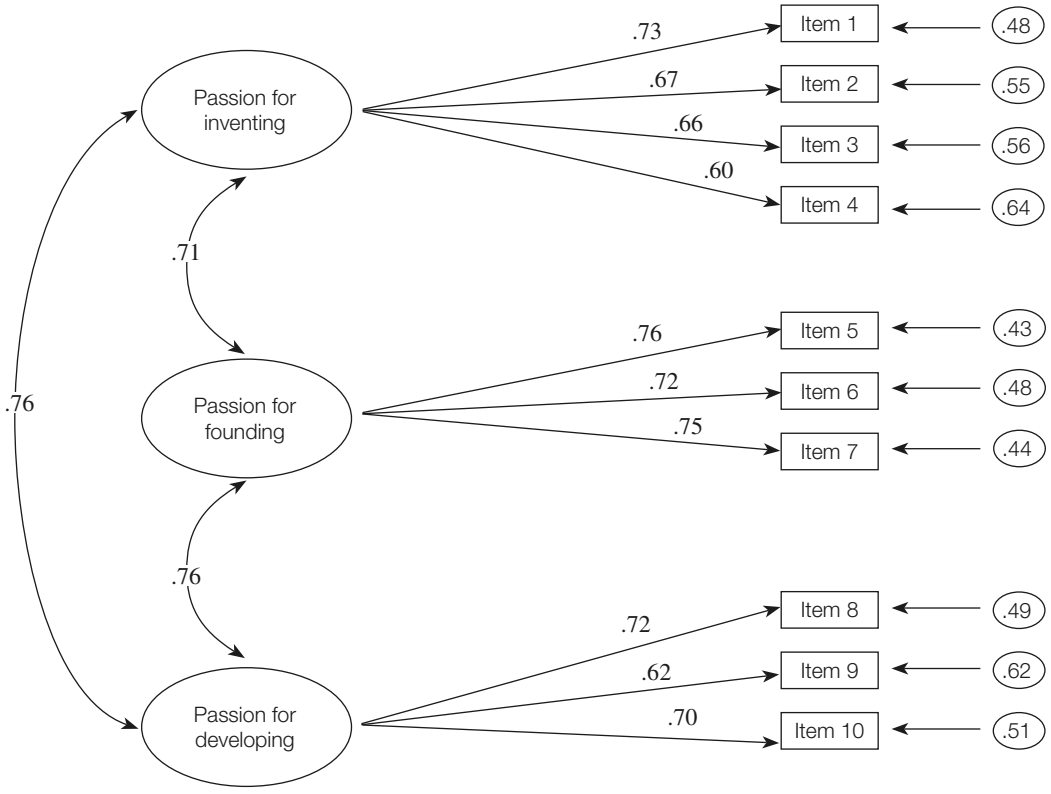
Table 2 – Results of CFA, alternative models (N = 1173)

	χ^2	*df*	*p*	CFI	TLI	RMSEA	SRMR	Comparison	$\Delta\chi^2$	*p*
M_1	345.76	32	<.001	.95	.92	.05 (.05, .07)	.04			
M_2	822.44	35	<.001	.86	.82	.12 (.11, .12)	.06	M_2–M_1	476.68	<.001
M_3	1771.67	35	<.001	.70	.61	.17 (.16, .18)	.26	M_3–M_1	1425.91	<.001

Note. M_1, 3-factor model; M_2, 1-factor model; M_3, 3-factor model, no co-variations.

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

Figure 1 – CFA (ML estimation) standardized solution for the Italian Entrepreneurial Passion Scale (N = 1173)



Correlations with other related dimensions

In order to test the concurrent validity of the EP Scale Italian version, correlation between EP in the three domains and other constructs potentially related to EP were computed (Table 3). All the three domains significantly and positively correlated with entrepreneurial intention, self-efficacy, internal locus of control, resiliency, need for achievement and risk propensity.

DISCUSSION AND CONCLUSION

The purpose of this paper was to validate the Italian version of an instrument designated to evaluate EP. Findings confirmed the factor structure, reliability and validity of the Italian EP Scale. Results from CFA indicated that the Italian EP Scale measures intense positive feelings towards three distinct domains: inventing, developing and founding. Furthermore, the correlation between EP in the three domains and the following constructs, related to EP in literature (e.g. Cardon et al., 2013; Molino et al., 2015), was confirmed: entrepreneurial intention, self-efficacy, internal locus of control, resilience, need for achievement and risk propensity.

The study confirmed that the Italian version of EP Scale can be administrated to entrepreneurs, startuppers and aspirant entrepreneurs with useful implications for future research and practice. As regards the instrument, this study contributed to the further improvement and testing of the Italian version of EP Scale. From the research point of view, the scale could be used to investigate the role of the three EP domains in the dynamics related to entrepreneurial intention and performance, in the Italian context. Moreover, the use of an international scale could support the cross-cultural research and the understanding of differences among countries in entrepreneurial field. Furthermore, concerning the practical implications, this instrument can be adopted by practitioners in the Italian context in order

Table 3 – Correlations (Pearson's r), means, standard deviations and reliabilities of all variables

	1	2	3	4	5	6	7	8	9
1. Passion for inventing	(.72)								
2. Passion for founding	.53**	(.80)							
3. Passion for developing	.52**	.53**	(.73)						
4. Entrepreneurial intention	.20**	.34**	.15**	(.81)					
5. Self-efficacy	.42**	.43**	.39**	.22**	(.86)				
6. Internal locus of control	.23**	.28**	.21**	.13**	.47**	(.81)			
7. Resiliency	.48**	.40**	.43**	.19**	.73**	.34**	(.81)		
8. Need for achievement	.53**	.50**	.48**	.18**	.73**	.35**	.69**	(.88)	
9. Risk propensity	.18**	.19**	.20**	.06*	.28**	.08*	.30**	.32**	(.72)
M	6.32	6.49	6.58	4.70	5.65	4.98	5.84	6.19	3.98
SD	.68	.73	.58	2.35	.75	1.10	.67	.65	.78

Note. Cronbach's Alpha on the diagonal; * $p<.05$; ** $p<.01$.

to evaluate EP and its dimensions, and to support training, mentoring, coaching, vocational and recruitment activities (Cortese et al., 2015; Molino et al., 2015). In conclusion, this study contributed to the Italian context in different ways, providing a valid and simple-to-use tool to evaluate the EP of both entrepreneurs and aspiring entrepreneurs, useful to support researchers and practitioners who want to deepen the psychological dynamics involved in entrepreneurial processes.

In spite of its contribution, this study results should not be interpreted without taking into account some limitations. First, the two convenience subgroups are not representative. Secondly, the study measured only single-source self-report data, which means the possibility of common method bias. Objective data of the enterprises, observation of work behaviours and other-report data could be useful in future studies in order to minimize the potential effects of common method variance. Moreover, this is a cross-section study; it would be necessary to evaluate the Italian version of the scale in more complex and longitudinal research design and the reliability of the measure should be investigated by means of a test-retest procedure. Finally, this version of the original EP Scale considers the intense positive feelings toward the domains of inventing, founding and developing, and not the centrality of these domains to entrepreneurs' self-identity; future studies should work on an exhaustive tool able to measure the identity centrality, as well as the intense positive feelings, in the three domains. Nevertheless, the Italian version of the EP Scale can be considered a useful and complete instrument to evaluate EP in the three domains theorized by Cardon et al. (2013) among Italian entrepreneurs and aspirants.

Acknowledgements

One of the authors acknowledged a grant for the Lagrange Project - CRT Foundation/ISI Foundation.

References

ARGENTERO, P. & VIDOTTO, G. (1994). *LOC-L: Una scala di locus of control lavorativo. Manuale*. Torino: Mediatest.

BAUM, J.R. & LOCK, E.A. (2004). The relationship of entrepreneurial traits, skill, and motivation to subsequent venture growth. *Journal of Applied Psychology*, *89*(4), 587-598. DOI: 10.1037/0021-9010.89.4.587

BORGOGNI, L., PETITTA, L. & BARBARANELLI, C. (2004). *TOM - Test di Orientamento Motivazionale*. Firenze: Giunti O.S. Organizzazioni Speciali.

CARDON, M.S., GREGOIRE, D.A., STEVENS, C.E. & PATEL, P.C. (2013). Measuring entrepreneurial passion: Conceptual foundations and scale validation. *Journal of Business Venturing*, *28*(3), 373-396. DOI: 10.1016/j.jbusvent.2012.03.003

CARDON, M.S., WINCENT, J., SINGH, J. & DRNOVSEK, M. (2009). The nature and experience of entrepreneurial passion. *Academy of Management Review*, *34*(3), 511-532. DOI: 10.5465/AMR.2009.40633190

CORTESE, C.G., GHISLIERI, C., MOLINO, M., MERCURI, A., COLOMBELLI, A., PAOLUCCI, E. & CANTAMESSA, M. (2015). Promuovere lo sviluppo delle start-up. *Sviluppo & Organizzazione*, *265*, 69-78.

CUBICO, S., BORTOLANI, E., FAVRETTO, G. & SARTORI, R. (2010). Describing the entrepreneurial profile: The entrepreneurial aptitude test (TAI). *International Journal of Entrepreneurship and Small Business*, *11*(4), 424-435. DOI: 10.1504/IJESB.2010.036295

DAHLBÄCK, O. (1990). Personality and risk-taking. *Personality and Individual Differences*, *11*(12), 1235-1242. DOI: 10.1016/0191-8869(90)90150-P

DI FABIO, A. & PALAZZESCHI, L. (2012). Connor-Davidson Resilience Scale: Psychometric properties of the Italian version. *Counseling*, *5*(1), 101-109.

ISTAT (2016). *Rapporto annuale 2016. La situazione del Paese*. Roma: Istituto Nazionale di Statistica.

LINAN, F. & CHEN, Y.W. (2009). Development and Cross-Cultural Application of a Specific Instrument to Measure Entrepreneurial Intentions. *Entrepreneurship Theory and Practice*, *33*(3), 593-617. DOI: 10.1111/j.1540-6520.2009.00318.x.

MOLINO, M., CORTESE, C.G., CARPANETO, A., MERCURI, A. & GHISLIERI, C. (2015). Human capital and start-up company success: Which psychological variables "make the difference"? *Counseling*, *8*(1). Retrieved from http://rivistedigitali.erickson.it/counseling/en/archivio/vol-8-n-1/

OEHLER, A., PUKTHUANTHONG, K., RUMMER, M. & THOMAS, W. (2007). Venture Capital in Europe: Closing the Gap to the U.S. In G.N. Gregoriou, M. Kooli & R. Kraeussl (Eds.), *Venture capital in Europe*. London: Elsevier.

RAUCH, A. & FRESE, M. (2007). Let's put the person back into entrepreneurship research: A meta-analysis on the relationship between business owners' personality traits, business creation, and success. *European Journal of Work and Organizational Psychology, 16*(4), 353-385. DOI: 10.1080/13594320701595438

SCHULER, H., THORNTON, G.C. III., FRINTRUP, A. & MUELLER-HANSON, R. (2002). *Achievement Motivation Inventory (AMI)*. Göttingen, Bern, New York: Hans Huber Publishers.

SMILOR, R.W. (1997). Entrepreneurship: Reflections on a subversive activity. *Journal of Business Venturing*, *12*(5), 341-346. DOI: 10.1016/S0883-9026(97)00008-6

VALLERAND, R.J., MAGEAU, G.A., RATELLE, C., LEONARD, M., BLANCHARD, C., KOESTNER, R. & GAGNÉ, M. (2003). Les passions del'ame: On obsessive and harmonious passion. *Journal of Personality and Social Psychology*, *85*(4), 756-767. DOI: 10.1037/0022-3514.85.4.756

VECCHIONE, M. & BARBARANELLI, C. (2005). Risk taking and investment behavior: An Italian version of Declared risk-taking scale and Investment risk-taking scale. *Testing Psicometria Metodologia*, *12*(3), 183-202.

Assessing sequential reasoning skills in typically developing children

Laura Zampini, Paola Zanchi, Chiara Suttora, Maria Spinelli, Mirco Fasolo, Nicoletta Salerni

Department of Psychology, University of Milano-Bicocca

ABSTRACT. Questo studio ha l'obiettivo di descrivere il *Sequential Reasoning Task* (*SRT*), un nuovo strumento sviluppato per valutare la capacità dei bambini di disporre degli eventi in ordine temporale. Hanno partecipato allo studio 200 bambini di età compresa fra 3 e 8 anni, a ciascuno dei quali è stata proposta individualmente una batteria di test cognitivi e linguistici. La prova SRT si rivela un valido strumento per valutare le abilità di ragionamento sequenziale: inoltre, il punteggio ottenuto dai bambini alla prova SRT risulta essere significativamente correlato con le loro competenze verbali e non-verbali.

SUMMARY. Introduction: Since serial ordering has an important role in both language development and learning abilities, the present study aims to describe a new instrument, the Sequential Reasoning Task (SRT), specifically designed to assess children's ability to place events in temporal order. Methods: Participants were 200 typically developing children, ranging from 3 to 8 years of age. Each child was individually administered a battery of cognitive and linguistic tasks. Results: The scores obtained in the SRT by children at different age levels appeared to be significantly different (except for 6- and 7-year-old children). Moreover, the scores obtained in the task were significantly related to the children's non-verbal and linguistic competence. Conclusions: The SRT appeared to be a valid instrument to assess children's sequential reasoning skills. It is engaging for children and easy to be administered also by teachers and therapists.

Keywords: *Sequential reasoning, Picture arrangement, Assessment, Language, Preschool age, School age*

INTRODUCTION

Starting from the preschool period, children have been observed to develop sequential reasoning skills (Catellani, 1991). Children younger than 3 years old have a good ability to comprehend temporal connections and to reproduce event sequences in the order in which they occur (e.g., Bauer & Thal, 1990). The ability to place events in temporal order has an important role in language development, in particular in narrative comprehension, and it is essential to construct a coherent mental representation of the events (Cain, Oakhill & Bryant, 2004; Zampini, Suttora, D'Odorico & Zanchi, 2013). For instance, in their recent study, Zampini et al. (2013) found that sequential reasoning skills, assessed by a picture arrangement task, explained part of the variance in listening text comprehension in 3-year-old children. This result was interpreted considering children's knowledge of scripts, in fact, if a child knows the order of things in real life, he/she may not only better understand how to arrange pictures, but he/she may also anticipate characters' actions in narrative texts.

Some studies found a significant relationship between language development and the ability of ordering events in children with developmental disorders. Johnels, Hagberg, Gillberg & Miniscalco (2013) showed that both temporal sequencing skills and language development were significant predictors of the ability to convey story information in the narratives of children with neurodevelopmental disorders. Moreover, Miranda, McCabe & Bliss (1998) found that children with specific language impairment (SLI) produced more narratives in which events are not ordered chronologically than typically developing children matched for age or syntactic level did.

Since serial ordering has an important role in both language development and learning abilities, the assessment of this competence in children has been frequently considered in standardised intelligence tests, as in the *Sequential Order* subtest of the Leiter-R (Roid & Miller, 1997), in the *Story Completion* subtest of the KABC-II (Kaufman & Kaufman, 2004) and in the *Picture Arrangement* subtest of the WISC-III (Wechsler, 1991), although this subtest has been removed from the most recent versions of the WISC. However, the possibility to objectively assess non-verbal serial ordering could be useful not only for clinicians (during intelligence assessment), but also for teachers, to programme targeted educational programs, and for speech and language therapists, to programme specific rehabilitation treatments. Therefore, a simple and specific instrument designed to assess children's serial order abilities could be helpful in both schools and rehabilitation services.

The aim of the study is to describe a new instrument, the *Sequential Reasoning Task* (SRT), specifically designed to assess children's ability to place events in temporal order. Descriptive data on the performance of children from 3 to 8 years of age are presented. Our hypothesis is that sequential reasoning skills gradually increase with children's chronological age and/or in the transition from kindergarten to primary school. In addition, the relationships between children's sequential reasoning skills and their cognitive and linguistic abilities are investigated. Considering the above-mentioned relationships found between the ability to order events and linguistic skills, we expected to find higher SRT scores in the children with a better linguistic competence.

METHOD

Participants in this study included 200 children, aged 3 to 7;11 years (ages are indicated in years;months), divided into five groups according to their chronological age: 3 years old (from 3 to 3;11); 4 years old (from 4 to 4;11); 5 years old (from 5 to 5;11); 6 years old (from 6 to 6;11); and 7 years old (from 7 to 7;11) (Table 1). Forty children (20 females) were included in each age group. All of the children, recruited from local kindergartens and primary schools, came from monolingual Italian speaking families. None of them were reported to have developmental problems. Parents signed a written consent form.

All of the children participated in a 45-minute test session, at their kindergarten or primary school. Each child was individually administered three different tasks assessing his/her cognitive skills: the *Raven's Coloured Progressive Matrices* (CPM) (Belacchi, Scalisi, Cannoni & Cornoldi, 2008), the *Sequential Order* subtest of the Leiter-R (Roid & Miller, 1997) and our SRT. In addition, the *Test for Receptive Grammar* (TROG-2) (Suraniti, Ferri & Neri, 2009) was administered to assess children's morphosyntactic comprehension abilities: the number of sentence blocks correctly solved (ranging from 0 to 20) was considered.

For the purposes of the present study, CPM raw score (ranging from 0 to 36) was considered as a measure of general non-verbal competence and the raw score of the

Table 1 – Participants' description for each age group

	N (females)	Chronological age (in months)		
		M	***SD***	***Range***
3 year old	40 (20)	43	3.31	36-47
4 year old	40 (20)	54	3.39	48-59
5 year old	40 (20)	66	3.44	60-71
6 year old	40 (20)	78	3.27	72-83
7 year old	40 (20)	89	3.45	84-95

Sequential Order subtest (ranging from 0 to 47) was used to assess children's ability in sequential processing. This task consists of 13 pictorial sequences, with an increasing level of complexity, in which the children are asked to identify the appropriate figures that complete the sequence from an array of options (e.g., the children are requested to order some triangles from smallest to largest in size).

Whilst the *Sequential Order* subtest gave us a measure of the children's ability to process abstract sequences, the SRT aimed to assess children's ability to process complex semantic relationships between events. Therefore, this task required the children to put some narrative sequences in the correct temporal and causal order. A pilot version of this task has been used in a previous study on preschool children (Zampini et al., 2013). Some ambiguous items of the pilot version have been modified (i.e., the items in which the percentage of errors was significantly higher) and some more complex items (i.e., 6-card stories) have been added to better assess sequential reasoning skills in school-age children. The SRT consists of 12 sets of illustrated stories, divided into 4 groups with increasing complexity level, depicted on 3, 4, 5 or 6 picture cards (examples of a 3- and a 6-card story are provided in figure 1_{ab}).

First, a 3-card story example is introduced to the child: he/she is asked to arrange 3 pictures presented in a scrambled order with the aim of creating a story. If the child is not able to arrange the pictures correctly, then the examiner demonstrates the solution. After the example, all of the stories are presented to the child in a fixed order, starting with the shorter ones (depicted on 3 cards) and moving to the longer levels of 4, 5 and 6 pictures. To increase the task complexity, the 6-card stories include a change in the story's scenery (e.g., in the example in figure 1b, the first scenes are located in a garden, whereas the last scenes are located in a bathroom). For each set, the cards are provided in a fixed scrambled order (i.e., the same for each child) and the child is asked to arrange them with the aim of creating a story (the order provided for each set is reported in Appendix). If the child is not able to arrange at least one story in a certain level, task administration is stopped. A child is assigned 3 points for each 3-card set arranged correctly, 4 points for each 4-card set, 5 points for each 5-card set and 6 points for each 6-card set. No points are assigned for picture sets incorrectly arranged. The total raw score (ranging from 0 to 54) is computed by adding the scores of all sets, with the exclusion of the example.

RESULTS

The SRT shows high internal consistency (Cronbach's Alpha = .925). Descriptive statistics for the cognitive and linguistic tasks administered to the children are reported in Table 2.

To analyse the differences in narrative sequential reasoning among children at different ages, univariate ANOVA was conducted. The scores obtained in the SRT by

Figure 1 – Example of a 3-picture (a) and a 6-picture (b) story in the SRT

Note. The original pictures are coloured.

Table 2 – Descriptive statistics for cognitive and linguistic tasks

	CPM			Sequential Order			SRT			TROG-2		
	M	***SD***	***Range***	***M***	***SD***	***Range***	***M***	***SD***	***Range***	***M***	***SD***	***Range***
3 year old	10.30	2.72	6-16	4.05	2.28	0-9	3.45	4.06	0-13	1.40	1.37	0-4
4 year old	14.65	3.45	6-26	5.88	4.65	0-24	12.48	14.51	0-45	4.88	3.04	0-13
5 year old	17.05	5.03	11-31	8.75	5.46	0-26	26.78	16.40	0-54	5.85	3.54	1-16
6 year old	20.80	4.59	14-31	18.65	7.67	4-38	43.23	10.16	6-54	11.38	4.58	2-20
7 year old	24.18	4.60	12-32	23.48	8.09	7-42	43.90	6.32	28-54	12.90	4.50	3-20

children at different age levels appeared to be significantly different ($F_{(4, 195)} = 102.46$; $p<.001$; $\eta^2 = .20$). Bonferroni post-hoc analysis showed that the performance of the 3-year-old children was significantly different from the performance of the children in all of the other age groups (all $p<.01$). The same result was found for the 4-year-old children (all $p<.01$) and the 5-year-old children (all $p<.001$). However, no significant differences were found in the scores obtained by 6- and 7-year-old children ($p>.05$). The 10th, 25th and 50th percentiles are reported for each age group in Table 3.

The children's performance on the SRT was significantly related to both cognitive and linguistics skills. As shown in Table 4, Pearson's *r* partial correlation, controlling for children's age, between the children's general non-verbal intelligence (CPM) and the scores obtained on the SRT was statistically significant. The partial correlation, controlling for age, between the scores obtained on the *Sequential Order* subtest and those obtained on the SRT allows to determine the concurrent validity of the SRT as an instrument to assess children's sequential reasoning competence.

Moreover, partial correlations, controlling for age and CPM, were computed between the scores obtained on the SRT and those obtained on the TROG-2 to verify the assumption of the existence of a specific relationship between narrative sequential reasoning and language development. The analysis showed that children's morphosyntactic comprehension appeared to be significantly related to their ability to put events in the correct order, independent of their age and cognitive level.

Table 3 – Percentiles on the SRT

	10th percentile	**25th percentile**	**50th percentile**
3 year old	0	0	3
4 year old	0	0	6
5 year old	3	10	29
6 year old	34	39	45
7 year old	34	40	44

Table 4 – Pearson's r partial correlations between the scores obtained on the SRT and children's cognitive and linguistic skills

	CPM[1]		**Sequential Order**[1]		**TROG-2**[2]	
	r	*p*	*r*	*p*	*r*	*p*
SRT	.302	<.001	.324	<.001	.345	<.001

Note. [1] Controlling for children's age; [2] Controlling for children's age and CPM scores.

DISCUSSION

The SRT appeared to be a valid instrument to assess children's ability to use sequential reasoning. It is engaging for children and can be administered also by teachers and therapists, because it does not require a qualification level, as intelligence tests do. Therefore, this instrument could be very helpful in both educational and clinical practice. It is possible to administer the task to children from 3 years of age. However, it should be noted that more than 25% of 3- and 4-year-old children were not able to correctly arrange any card set. Therefore, a total task failure (i.e., the inability to solve any item) in these age ranges could not be considered as a marker of a clinical condition. In contrast, the ability to correctly order some sequences in 3- and 4-year-old children can be considered a strength.

No statistically significant differences were found in the scores obtained at the SRT by 6- and 7-year-old children (although ranges are quite different). In school-age children the performance in the SRT does not seem to be related to their chronological age, in fact the percentiles of 6- and 7-year-old children are very similar. Contrary to our hypothesis, the increasing in children's sequential reasoning skills is not gradual. In fact, looking at raw scores, it seems that there is a gap between preschool- and school-age children in this competence. It has to be noted that a similar pattern has been found in the *Sequential Order* scores. Both these results could be related to the beginning of reading and writing learning (i.e., when children are 6-year-old in Italy) that requires and trains children's sequential skills.

Our hypothesis of a relationship between the scores obtained at the SRT and children's linguistic skills has been confirmed: we found a significant correlation between the ability to place events in order and language development, independent of children's age and cognitive level. This result supports the findings of Zampini et al. (2013) who demonstrated the existence of an association between children's sequential reasoning skills and their narrative text comprehension. A possible interpretation of that association was that both the arrangement of a series of pictures and text comprehension require children to infer the entire situation from an analysis of single elements (i.e., pictures or sentences). The same interpretation could explain the association between sequential reasoning skills and morphosyntactic comprehension found in the present study: a child should infer the meaning of a sentence from an analysis of the single words. In addition, we have to consider that processing word order is a fundamental ability to comprehend complex syntactic sentences.

Limits and future directions

The cross-sectional nature of the study does not allow us to establish the direction of the association between linguistic abilities and sequential order: we do not know if morphosyntactic comprehension influences sequential reasoning skills or if the opposite is true. It is also possible that this relationship could be mediated by a third latent variable, such as children's verbal intelligence. Future studies will investigate this relationship with a longitudinal design, to clarify if sequential reasoning skills could be considered as a predictor of language development.

Moreover, future studies will investigate the association between the ability to arrange pictures and morphosyntactic comprehension in children with SLI to verify the role of sequential reasoning skills in language disorders.

Acknowledgments
We gratefully acknowledge Dododesign (http://www.dododesign.it/) that made the drawings of the *Sequential Reasoning* Task and the children who participated in the study.

References

BAUER, P.J. & THAL, D.J. (1990). Scripts or scraps: Reconsidering the development of sequential understanding. *Journal of Experimental Child Psychology*, *50*, 287-304.

BELACCHI, C., SCALISI, T.G., CANNONI, E. & CORNOLDI, C. (2008). T*aratura italiana del test Matrici di Raven Forma Colore (CPM-47)*. Firenze: Giunti O.S. Organizzazioni Speciali.

CAIN, K., OAKHILL, J. & BRYANT, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, *96*(1), 31-42.

CATELLANI, P. (1991). Children's recall of script-based event sequences: The effect of sequencing. *Journal of Experimental Child Psychology*, *52*(1), 99-116.

JOHNELS, J.Å., HAGBERG, B., GILLBERG, C. & MINISCALCO, C. (2013). Narrative retelling in children with neurodevelopmental disorders: Is there a role for nonverbal temporal-sequencing skills? *Scandinavian Journal of Psychology*, *54*(5), 376-385.

KAUFMAN, A.S. & KAUFMAN, N.L. (2004). *KABC-II: Kaufman Assessment Battery for Children*. Circle Pines, MN: AGS.

MIRANDA, A.E., McCABE, A. & BLISS, L.S. (1998). Jumping around and leaving things out: A profile of the narrative abilities of children with specific language impairment. *Applied Psycholinguistics*, *19*(4), 647-667.

ROID, G.H. & MILLER, L.J. (1997). *Leiter International Performance Scale-Revised*. Wood Dale, IL: Stoelting.

SURANITI, S., FERRI, R. & NERI, V. (2009). *Test for reception of grammar - version 2*. Firenze: Giunti O.S. Organizzazioni Speciali.

WECHSLER, D. (1991). *WISC-III: Wechsler intelligence scale for children. Manual*. San Antonio, TX: The Psychological Corporation.

ZAMPINI, L., SUTTORA, C., D'ODORICO, L. & ZANCHI, P. (2013). Sequential reasoning and listening comprehension in preschool children. *European Journal of Developmental Psychology*, *10*(5), 563-579.

APPENDIX

Initial scrambled order provided for each card set

3-card sets	Set 1: 2 1 3	Set 2: 2 1 3	Set 3: 3 1 2
4-card sets	Set 4: 3 2 4 1	Set 5: 4 2 1 3	Set 6: 3 1 2 4
5-card sets	Set 7: 4 2 5 3 1	Set 8: 5 3 4 2 1	Set 9: 4 2 1 5 3
6-card sets	Set 10: 5 4 3 1 6 2	Set 11: 6 1 4 3 2 5	Set 12: 5 1 3 2 6 4

College Satisfaction Scale (CSS): Evaluation of contextual satisfaction in relation to college student life satisfaction and academic performance

Ernesto Lodi[1], Diego Boerchi[2], Paola Magnano[3], Patrizia Patrizi[1]

[1] University of Sassari
[2] Università Cattolica del Sacro Cuore, Milan
[3] Kore University, Enna

ABSTRACT. Molti studi, adottando prospettive legate alla prevenzione e alla promozione della qualità della vita, si sono focalizzati sulla rilevazione dei fattori che promuovono od ostacolano il benessere e i risultati accademici degli studenti universitari. Lo scopo del presente studio è di creare e validare un nuovo questionario per valutare la soddisfazione dominio-specifica nel contesto universitario: la College Satisfaction Scale (CSS). Il questionario è composto da 20 item che valutano la soddisfazione degli studenti in 5 dimensioni: congruenza della scelta del corso di laurea, qualità dei servizi universitari, rapporti con i colleghi e le colleghe, qualità del metodo di studio e utilità del corso di laurea per il futuro professionale. Hanno partecipato alla ricerca 425 studenti universitari italiani, ai quali vanno aggiunti 88 studenti che hanno partecipato al precedente studio pilota. Sia l'affidabilità, valutata con l'indice Omega, sia la validità di costrutto, stimata attraverso l'analisi fattoriale confermativa, offrono buoni risultati per tutte le dimensioni del questionario. I risultati hanno mostrato che i punteggi di efficienza accademica sono fortemente legati alla soddisfazione per la qualità del metodo di studio e leggermente correlati ad alcune delle altre aree di soddisfazione.

SUMMARY. A large number of studies have identified factors that foster or inhibit well-being and academic achievement of college students from the perspective of prevention and quality of life promotion. The aim of the present study was to create and validate a new questionnaire to assess domain-specific satisfaction at University: the College Satisfaction Scale (CSS). It is composed of 20 items to measure 5 dimensions: appropriateness of the student's choice, quality of the University services, relationships with his/her colleagues, quality of his/her study habits and usefulness for his/her future career. Participants were Italian college students: 88 in the pilot study and 425 in the actual one. Both reliability, assessed with Omega index, and construct validity, estimated through confirmatory factor analysis, were good for all the subscales. The results showed that both average grade scores and academic efficiency were strongly related to satisfaction with the efficiency of studying and slightly related with some of the other subscales.

Keywords: *College satisfaction, Life satisfaction, Academic performance*

INTRODUCTION

Students' persistence in college is a problematic issue for many countries. In the United States, around one fifth to one quarter of college students drop out at the end of their freshman year (Ryan, 2004). In Italy, 4 students in 10 leave the University system before obtaining a first-level college degree (Anvur - Biennial report on the status of universities and research, 2014), and the first year at college is a significant period for dropout risk (Cenvus - 11th report on the status of University system, 2011). Regarding this issue, a large number of studies are focused on students' characteristics explained by the "student-centered research tradition" (Smart, Feldman & Ethington, 2006); other studies are focused on how universities can create conditions for student persistence (Berger & Milem, 2000; Chen, 2012). Interest in academic adjustment has recently been increasing along with the evaluation of domain-specific satisfaction, as various studies on students' well-being are related to their academic fit, retention and success. There is also a need for measuring domain-specific satisfaction because, from the perspective of prevention and health promotion, many psychological disciplines are focused on identifying factors that foster or interfere in well-being and quality of life (Seligman, 2002). Some authors have indicated that the concept of satisfaction is also important when applied to the educational context, but there are few studies about this subject, as well as a lack of reliable scales to measure it (Martins, 1998; Sisto et al., 2002). Actually, literature offers several instruments that show some criticism: some of them are multidimensional but they are too long, others are short but one-dimensional. Since domain-specific satisfaction contributes to overall satisfaction in students' lives (Lent et al., 2005), it is necessary to have a reliable, short and multidimensional domain-specific instrument to evaluate satisfaction in the academic context.

LITERATURE REVIEW

Domain-specific satisfaction and relationship with overall life satisfaction

Adopting a social-cognitive view of work and educational satisfaction, Lent (2004) proposed to unify perspectives on subjective and psychological well-being in which cognitive, behavioural, social and personality variables jointly define domain-specific and global life satisfaction. This model, although primarily aimed to explain the assessment of job satisfaction, can be adapted to understand domain-specific satisfaction in the educational context. In fact, consistent with the framework of Social-Cognitive Career Theory (SCCT), education and work are interconnected areas. Consequently, Lent & Brown (2008) believe that the well-being model can be extended to students' satisfaction with the educational context (i.e. satisfaction with their role as a student). Domain satisfaction is one of the antecedents of overall life satisfaction, and it is determined by personality factors and socio-cognitive mechanisms such as self-efficacy and environmental supports (Lent, Taveira, Sheu & Singley, 2009). Education and work are so central to people's lives that satisfaction with social ties and the mood in the workplace (Judge & Ilies, 2004; Rain, Lane & Steiner, 1991) can be extended to all other areas, such as the family environment. In the same way, school or University is equally important in the lives of teenagers and young adults, and satisfaction in this field affects other areas of their lives. In fact, Lent et al. (2007) defined academic satisfaction as the enjoyment of one's role or experiences as a student.

Life satisfaction is influenced by various factors, such as personality traits (positive and negative affect), behavioural characteristics and cognitive processes (Diener Emmons, Larsen & Griffin, 1985; Lent & Brown, 2008). The concept of life satisfaction is associated with social, occupational, mental and physical health outcomes (Pavot & Diener, 2008; Whisman & Judd, 2015). Several authors (Brief, 1998; Heller, Judge & Watson, 2002) have found relationships between domain-specific and general life satisfaction. In particular, they have observed that the two fields are affected by a sort of "osmosis", a relationship between the part and the whole (Judge & Locke, 1993).

Since academic satisfaction is defined as the expected satisfaction with the accomplishment of academic goals or aspirations (Kumar & Dileep, 2006), domain-specific satisfaction can also be affected by variables related to career choice. In fact, academic satisfaction scores have been positively associated with career decision self-efficacy and negatively associated with career choice anxiety and indecision (Nauta, 2007). The congruence between professional interests and the type of academic major choice influences academic performance and satisfaction even more than academic abilities do; students enrolled in academic

programs that are coherent with their professional interests are the most satisfied (Tracey & Robbins 2006).

Finally, educational satisfaction, life satisfaction and satisfaction with one's relationships are significant predictors of positive outcomes in several spheres of life (Wilkins et al., 2014). In fact, adolescents with high life satisfaction have a higher level of physical health, experience enhanced social relationships and show better academic engagement and achievement (Lewis, Huebner, Malone & Valois, 2011; Salmela-Aro & Tuominen-Soini, 2010; Suldo, Huebner, Savage & Thalji, 2011; Suldo & Shaffer, 2008; Suldo, Thalji & Ferron, 2011; Wilkins et al., 2014).

Academic persistence and academic performance: relationship with academic satisfaction

In the literature of the last decades, performance in college assessed by GPA (Grade Point Average) has been the most used indicator to explain the intention of leaving college. Some studies have shown that first-year college GPA is a significant predictor both of early dropout (Ma & Cragg, 2012) and late dropout: a longitudinal study of Spady (1971) underlined that formal academic performance was evidently the main factor related to attrition. Other studies have highlighted that during the freshman year, the combined effect of performance and satisfaction (with courses and college in general) conditions the dropout phenomenon (e.g. Edwards & Waters, 1982).

Several studies have shown that academic performance and satisfaction are related in many ways. From a socio-cognitive point of view, performance can be considered feedback about the progress toward objectives, and the perception of this kind of progress can be considered a predictor of well-being (e.g. Elliot, Sheldon & Church, 1997) that also conditions the motivation of the student (Lent, Brown & Hackett, 1994). Lent & Brown (2008) underlined that domain-specific goals affect students' satisfaction, in particular through their definition, commitment and advancement towards them. Moreover, academic satisfaction is linked to performance and predicts engagement and consequent academic advancement (Huebner & Gilman, 2006; Huebner & McCullough, 2000; Verkuyten & Thijs, 2002).

Nurmi, Aunola, Salmela-Aro & Lindroos (2003) found that students' anticipation of success predicted academic achievement and satisfaction, which in turn improved their subsequent expectation of success. Chambel & Curral (2005) showed that satisfaction has a direct impact on academic performance and that it mediates between academic work control and performance. Strahan & Credé (2015), using a large dataset from 300 institutions, found that college satisfaction exhibits from moderate to strong relationships with retention intentions and weak relationships with academic performance. Kandemir (2014) argued that students' academic unproductiveness may reduce learning motivation, thus increasing academic procrastination. For this author, life satisfaction is a significant variable of academic procrastination (Kandemir, 2014), and other authors have underlined the negative effect of life satisfaction in predicting academic procrastination (e.g. Savithri, 2014). According to these studies, when students are satisfied with their life, their academic procrastination decreases, while their responsibility and motivation to cope with academic duties increase.

Another group of studies underlined the role of context on performance and satisfaction. El-Hilali, Al-Jaber & Hussein (2015), for example, found that students with high GPAs showed a higher level of achievement and satisfaction. Schmitt et al. (2008) showed that academic fit leads to academic satisfaction; in particular, changes in academic fit were related to similar changes in level of satisfaction and GPA and were negatively related to turnover intent. Karemera, Reuben & Sillah (2003) highlighted that satisfaction with the environment and academic services was related with performance, and the appropriateness of library services was linked with positive college results. Finally, some authors have also underlined the effects of elements of the "physical" environment (adequate materials, number of classrooms, campus cleaning, etc.) and the reliability and utility of the University (in terms of programs and acquired knowledge) on college students' level of satisfaction (Negricea, Edu & Avram, 2014). Finally, Silva (2001) argues that academic satisfaction refers not only to the training experience but also to factors such as the student's relationships with teachers, colleagues, administration, facilities and resources (Astin 1993).

Assessment of Academic Satisfaction

The most used scales in previous research on academic satisfaction are the following:

– The *College Student Satisfaction Questionnaire* (CSSQ)

(Starr, Betz & Menne, 1971) consists of 70 items grouped in five scales that are related to various aspects of college life; it evaluates the student's degree of satisfaction on the following five scales: Working Conditions, Compensation, Quality of Education, Social Life and Recognition.

- The *Perceived Quality of Academic Life* (PQAL) (Staats & Partlo, 1990) consists of 10 items, which come from a modified version of the Feelings About College (FAC) scale (Okun et al., 1986).
- The *Student Satisfaction Inventory* (Elliott & Shin, 2002) consists of 116 items that cover a full range of college experiences. Students are asked to rate each college experience with regard to 'importance' and 'satisfaction'. Students are also asked three summary questions dealing with (a) 'overall satisfaction' with their educational experience (b) level of expectations met by their college, and (c) whether they would enrol again at their college.
- The *Academic Major Satisfaction Scale* (AMSS) evaluates global major satisfaction (Nauta, 2007) and is composed of six items. This scale has also been validated on a Korean sample (Sovet, Park & Jung, 2014)
- The *Scale of Academic Satisfaction* (Lent et al., 2007) is a 7-item scale. The authors found academic satisfaction to correlate in the expected direction with self-efficacy, outcome expectations and goal progress, and the scale was found to be strongly reliable.
- The *Academic Satisfaction Scale* (Schmitt et al., 2008) includes five items designed to assess students' academic satisfaction. This scale has also been validated on a Turkish sample (Balkis 2013).

Other scales, as the *Course Experience Questionnaire* (Ramsden, 1991), the *Brief Multidimensional Students' Life Satisfaction Scale* (Huebner, 1994), and the *Escala de Satisfação Acadêmica Universitária* (Sisto et al., 2008), have been designed to assess several facets of students' experience (i.e. quality of instructor, campus environment, friends, etc.), including satisfaction. One limitation of those scales is the large number of items necessary for assessing each facet of college students' life. In the end, some authors use brief scales created ad hoc (Negricea et al., 2014; Nurmi et al., 2003).

Considering the analysis of the literature review, the available instruments seem to have some limitations: some of them are too short and one-dimensional, providing information only on an global satisfaction; others, on the contrary, even though explore different dimensions of the academic satisfaction, are too long: in a perspective of prevention, the use of a battery of instruments with many items makes difficult the realization of a screening for early detection of the academic difficulties. For these reasons, we decided to propose a new questionnaire, the College Satisfaction Scale (CSS), which serves as a more comprehensive measure of academic satisfaction in a multidimensional perspective, more concise, but with good psychometric characteristics. With respect to the previous scales, this new instrument assesses different dimensions of academic satisfaction (choice, study, environment, etc.) using a reduced number of items, despite the multidimensional nature of the information obtained.

AIMS

Considering the analysis of the literature, the study presents the development of a new instrument for the assessment of academic satisfaction and its psychometric properties.

The main aims of this study are:

1) to build a reliable and useful instrument to assess different areas of college students' satisfaction in the academic context;
2) to confirm the validity of the scale scores through the *SWLS - Satisfaction With Life Scale* (Diener et al., 1985) as an external criterion;
3) to explore the relationship of the new instrument with academic performance and efficiency.

METHODOLOGY

Item development

Three independent experts in vocational and educational guidance selected 77 initial items from the analysis of scientific literature and existing tests (Nauta, 2007; Schmitt et al., 2008; Sisto et al. 2008). The initial form covered some macro areas of satisfaction (study, performance, relationship with teachers, relationship with colleagues, academic fit, course choice, physical environment, professional future). Then 36 items were discarded because at least one of three independent judges considered it inadequate to evaluate academic satisfaction or because there were already very similar items. At the end, 41 items remained and they were submitted to a small number of students in order to verify the

items' understandability and no modifications were made. This version was then administered in the pilot study. The final version, after the item analysis consists of 20 items.

Pilot study: scale construction

The aim of the pilot study was to develop the basic structure of the CSS by:

- choosing latent factorial structure through principal component analysis (PCA);
- selecting a small number of items;
- providing evidence regarding the internal consistency through McDonald's omega, which overcomes some weakness of the Cronbach's alpha (Dunn, Baguley & Brunsden, 2014).
- *Participants.* A pilot study was conducted on 88 Italian college students attending a Psychometrics course during the second semester of the second year. Females were more represented than males (90.9% vs 9.1%), and 80.7% were 21 years old or younger.
- *Procedure.* We asked the students to compile the questionnaire anonymously online to build a database for class exercises indicating how satisfied they were with each sentence using a 5-point Likert scale: 1 = not at all, 2 = a little, 3 = somewhat, 4 = very, 5 = completely.
- *Results.* Descriptive statistics showed normal distribution for all the items with skewness and kurtosis from −1 to 1. This allowed us to test a Principal Axes Factoring with Oblimin Rotation with Kaiser Normalization. At a first step, we used an Eigenvalue bigger than 1 as the criterion. It extracted 10 factors explaining 63.14% of variance, some with a large number of items, some with a small number. After, a series of tests was conducted to eliminate those items that were less related to the others. Table 1 shows the final result; the Factorial Model Matrix describes a model composed of 5 factors of 4 items each, explaining 57.02% of variance: appropriateness of the student's choice (CH; i.e. "For choosing this academic path"); quality of the University services (SE; i.e. "Because my University is adequately equipped"); relationships with his/her colleagues (RE; i.e. "Because I have fellow students I'm studying well with"); quality of the his/her study habits (ST; i.e. "About my way of studying"); usefulness for his/her future career (CA; i.e. "Because my studies will be useful for finding future employment").

Subscales were moderately and positively correlated, and internal consistency was good for all of them, with McDonald's omega indexes between .746 and .845 (Table 2).

Main study

The main study sought evidence of the validity and reliability of the CSS scores by:

- Testing a latent factorial structure through Confirmatory Factorial Analysis (CFA);
- Providing evidence regarding the internal consistency through McDonald's omega;
- Testing its concurrent validity with academic performance as the external criterion;
- Testing its concurrent validity by assessing the amount of variance of college satisfaction on general life satisfaction.
- *Participants.* The main study was conducted on two samples: a mixed sample (MS), composed of 278 Italian university freshmen students attending different courses, mostly in humanistic sciences, from two different Italian universities (F = 73.7%; M = 26.3%; 55.7% 20 years old or younger), and a psychology sample (PS), composed of 147 Italian university students attending the third year of a course in psychology (F = 85.0%; M = 15.0%; 65.3% 22 years old or younger).
- *Procedure.* All students were asked to fill out an anonymous questionnaire "about their experience as a college student". To reach all students, including ones who were not attending class, we chose an online survey, well aware that the response rate tends to be low for this option. For MS, the response rate was 8% of all freshmen of the two universities; such a rate is good according to "liberal conditions", but not good enough according to "stringent conditions" (Nulty, 2008) to be considered representative of the population of the two universities.

  PS students were requested to collaborate during their lessons. Students were free to answer the survey, and 82% responded to the questionnaire online at home. Those who did so had the option to indicate their name and surname in order to obtain a report.

  The CSS questionnaire was administered before the *Satisfaction With Life Scale* (SWLS, Diener et al., 1985) and some questions on demographic data (sex, age, average score, number of successfully taken exams, University and course attended).

Table 1 – CSS Pilot: Factorial Model Matrix

	CH	SE	RE	ST	CA
Item 25	**.687**	.007	.130	**.308**	−.017
Item 6	**.644**	.128	.050	.027	.168
Item 13	**.560**	.074	.150	−.096	**.389**
Item 27	**.354**	−.085	.078	.244	.270
Item 7	.014	**.792**	−.087	−.043	−.117
Item 14	−.037	**.680**	−.012	−.045	.131
Item 39	−.050	**.562**	.221	.034	.144
Item 11	.088	**.506**	.028	.152	−.034
Item 33	.170	−.025	**.904**	−.114	−.034
Item 18	.023	−.097	**.870**	−.018	.012
Item 22	−.288	.179	**.459**	.101	.086
Item 17	.202	.179	**.454**	**.308**	−.058
Item 23	−.038	−.044	.010	**.710**	.007
Item 15	−.071	.107	−.053	**.647**	.109
Item 32	.403	.042	.133	**.567**	−.061
Item 30	.299	−.009	.004	**.559**	.167
Item 24	.037	.106	−.026	.034	**.792**
Item 9	.060	.100	−.114	.066	**.708**
Item 31	−.032	−.179	.182	.123	**.691**
Item 2	.331	.237	−.022	−.159	**.337**

Note. Boldface indicates saturation indexes > .300

Legenda. Satisfaction elements: CH = appropriateness of choice; SE = quality of the University services; RE = relationships with colleagues; ST = effectiveness of his/her study habits; CA = usefulness for future career.

Table 2 – CSS Pilot: McDonald's Omega and Correlation Matrix between Subscales

	CH	SE	RE	ST	CA	Mean	SD
CH	**.845**					15.41	2.85
SE	.302**	**.746**				8.38	2.47
RE	.410**	.271*	**.816**			12.97	2.83
ST	.592**	.262*	.421**	**.792**		13.99	2.61
CA	.627**	.361**	.313**	.424**	**.798**	12.27	2.66

Note. Boldface indicates McDonald's Omega index. * $p<.05$; ** $p<.01$

Legenda. Satisfaction elements: CH = appropriateness of choice; SE = quality of the University services; RE = relationships with colleagues; ST = effectiveness of his/her study habits; CA = usefulness for future career.

– *Measures*

CSS-College Satisfaction Scale. The questionnaire was the same as the one used in the pilot study, with the exception of four items which were partially modified to better focus on the topic of the specific subscale they belonged to. Descriptive statistics showed normal distribution for all the items for both samples.

SWLS-Satisfaction With Life Scale (Diener et al., 1985). This is a self-reported one-dimensional scale on general satisfaction with life, originally developed in the United States, that is composed of five items on a 7-point Likert scale (from 'strongly disagree' to 'strongly agree'). It has been translated into and tested in several languages, and some cross-national analysis has been conducted (i.e. Caprara et al., 2012; Whisman & Judd 2015). The Italian version was validated on a college student sample by Di Fabio & Busoni (2009): the questionnaire was mono-factorial, and Cronbach's Alpha was .88.

Performance Index-Average Grade Score (AGS). As underlined previously, Average Grade Score is the most used indicator for academic performance (e.g. El-Hilali et al., 2015; Ma & Cragg, 2012; Rotter, 1988). We asked the students to indicate their average grade score to one decimal. Frequency distribution shows that 77.5% of them indicated "zero" as the decimal, suggesting that these data were not as precise as requested, which may have lowered the strength of the relation to other dimensions. Scores could range from 18 to 30, and data distribution was normal.

Performance Index-Efficiency (EF). We calculated the ratio of the number of passed exams to all the exams each student should have taken. Scores ranged from 0 (no exams passed) to 1 (all exams passed), and data distribution was normal. This index, combined with AGS, is a good indicator of performance, since it has been adapted by Di Nuovo (2009) in a study on the effectiveness of ministerial academic paths.

CSS Structural Validity

We tested the questionnaire structure through CFA using the maximum likelihood method and AMOS software. Goodness-of-fit indexes were examined through the chi-square test, Root Mean Square Error of Approximation (RMSEA) and Comparative Fit Index (CFI). Even if a non-significant chi-square is desired, which would suggest that the observed and reproduced covariance matrix do not significantly differ, models with a large sample can only be evaluated by RMSEA and CFI because this test is sensitive to sample size (Byrne, 2010). Models with acceptable fit also presented RMSEA<.08 and CFI>.90 (Bentler, 1990), whereas models with optimum fit presented RMSEA<.05 and CFI>.95 (Hu & Bentler, 1999; Jackson, Gillaspy & Purc-Stephenson, 2009).

Two different models were tested: a mono-factorial model and a 5 latent correlated variables model. Table 3 shows model

fit indexes: the mono-factorial model was unacceptable, while the 5 latent variables model had acceptable fits, considering covariance between some errors, and it worked better with the PS sample. Factor loadings ranged from .62 to .95 for MS, and from .59 to .94 for PS.

CSS Reliability

Internal consistency was assessed for the five subscales, and all McDonald's Omega indexes were optimum (Table 4).

CSS Concurrent Validity with Performance Indexes

We estimated concurrent validity by correlating the scores of each subscale with the two performance indexes and hypothesising positive relations mostly with study habits and in part with the other subscales. Statistically significant correlations in Table 5 confirm our hypothesis in part. Study habits were highly related with both AGS and EF for PS, while it was highly related with AGS, but less related with EF for MS. PS was related just with study habits, while MS was also related with choice for AGS and with relations for both performance indexes.

Table 3 – Structural Validity: CFA good-of-fit-indexes

Model	χ^2 (p)	df	RMSEA	CFI
Monofactorial				
MS	2097.34***	170	.210	.536
PS	1043.18***	170	.188	.563
5 latents				
MS $_{(5)}$	337.19***	154	**.068**	**.956**
PS $_{(2)}$	223.63***	158	**.053**	**.967**

Note. In brackets the number of covariances related.
Boldface indicates good and optimum fit indexes. *** $p<.001$

Legenda. RMSEA = Root Mean Square Error of Approximation; CFI = Comparative Fit Index; MS = Mixed Sample; PS = Psychology Sample.

Table 4 – Reliability: McDonald's Omega indexes

Scales	MS	PS
CH	.919	.919
SE	.891	.801
RE	.918	.920
ST	.873	.842
CA	.922	.903

Legenda. MS = Mixed Sample; PS = Psychology Sample. Satisfaction elements: CH = appropriateness of choice; SE = quality of the University services; RE = relationships with colleagues; ST = effectiveness of his/her study habits; CA = usefulness for future career.

Table 5 – CSS: Concurrent Validity with Performance Indexes

	MS	PS
Average Grade Score (AGS)		
CH	.272**	.106
SE	.099	.132
RE	.179**	.001
ST	.434**	.469**
CA	.080	.124
Efficiency (EF)		
CH	.037	.110
SE	−.152	−.002
RE	.197*	.104
ST	.179*	.470**
CA	.098	.137

Note. * p<.05; ** p<.01

Legenda. MS = Mixed Sample; PS = Psychology Sample. Satisfaction elements: CH = appropriateness of choice; SE = quality of the University services; RE = relationships with colleagues; ST = effectiveness of his/her study habits; CA = usefulness for future career.

CSS Concurrent Validity with SWLS

We hypothesized that college satisfaction explains part of general life satisfaction and is higher in freshman students because the University experience is more important for them because they are at the beginning of their career.

To be sure about the psychometric quality of the SWLS, we first tested its scores validity and reliability. Structural validity was tested with CFA, and goodness-of-fit-indexes were optimum both for the MS (χ^2 = 6.925, df = 5, n.s.; RMSEA = .039; CFI = .998) and PS (χ^2 = 3.582, df = 5, n.s.; RMSEA = .000; CFI = 1.000). Reliability was tested with McDonald's omega, and the indexes were very good (MS = .901; PS = .858).

Then, we tested the role of college satisfaction on general life satisfaction through regression analysis using SWLS as the dependent variable and CSS subscales as independent variables, inserting each one of them as a block. Considering CSS subscales were correlated, multicollinearity was tested with a variance inflation factor (VIF). Values ranged from 1.281 to 3.153 for MS and from 1.106 to 2.732 for PS, indicating moderate collinearity in both samples. Table 6 shows that, for MS, study habits, relations and choice were able to explain 33.3% of variance, while career and services were not able to explain any of it. Table 7 shows, for PS, that study habits and relations were able to explain 16.5% of variance, which, according to our hypothesis, is smaller than (more or less half of) that explained for MS.

DISCUSSION

The literature, item analysis, and explorative and confirmatory factor analysis conducted on two different samples led us to a new questionnaire able to assess five different areas of satisfaction in college students. Compared

Table 6 – CSS: Linear Regression with SWLS as dependent variable for MS

Blocks	R	R^2	Modification statistics		
			R^2	F (df1, df2)	Sign.
ST	.525	.275	.275	94.953 (1,250)	.000
ST, RE	.560	.313	.038	13.868 (1,249)	.000
ST, RE, CH	.577	.333	.019	7.079 (1,248)	.008
ST, RE, CH, CA	.578	.334	.001	.473 (1,247)	.492
ST, RE, CH, CA, SE	.578	.334	.000	.100 (1,246)	.752

Legenda. Satisfaction elements: ST = effectiveness of his/her study habits; RE = relationships with colleagues; CH = appropriateness of choice; CA = usefulness for future career; SE = quality of the University services.

Table 7 – CSS: Linear Regression with SWLS as dependent variable for PS

Blocks	R	R^2	Modification statistics		
			R^2	F (df1, df2)	Sign.
ST	.329	.108	.108	17.153 (1,141)	.000
ST, RE	.406	.165	.056	9.399 (1,140)	.030
ST, RE, CH	.406	.165	.000	.002 (1,139)	.961
ST, RE, CH, CA	.411	.169	.004	.692 (1,138)	.407
ST, RE, CH, CA, SE	.420	.176	.007	1.239 (1,137)	.268

Legenda. Satisfaction elements: ST = effectiveness of his/her study habits; RE = relationships with colleagues; CH = appropriateness of choice; CA = usefulness for future career; SE = quality of the University services.

to the previous questionnaires, this one was shown to be suitable to assess different dimensions using a reduced number of items despite the multidimensional nature of the information obtained. The CSS scores showed very good psychometric features with both good internal consistency and good structural validity. It seems to confirm that satisfaction involves different areas of academic experience of students and reflect its multidimensional structure (Pike, 1991; Soares, Vasconcelos & Almeida, 2002;), even though not all the areas are strongly related to the performance and the academic success, contrary to our expectations considering previous studies (e.g. Karemera et al., 2003; Negricea et al., 2014; Silva 2001).

The relations with both academic performance and overall life satisfaction proved its concurrent validity. Relations between domain-specific satisfaction and performance

confirmed previous research, such as El-Hilali et al.'s (2015) recent demonstration of the connection between GPA, level of achievement and domain-specific satisfaction. The strong correlation between performance indexes and satisfaction with the effectiveness of studying is congruent with the hypothesis that this factor is mostly related to the evaluation of one's own performance behaviour and results. Differences between freshman and third-year students can be explained in terms of a different level of maturity in managing the role of being a college student.

The close relation between domain-specific satisfaction and overall life satisfaction has also been previously highlighted by the literature according to the socio-cognitive vision of well-being, which sustains a direct relation between the part and the whole (Judge & Locke 1993). In agreement with previous studies (e.g. Heller, Watson & Ilies, 2004) our results confirm the strong relationship between domain-specific and general life satisfaction. The two samples are not completely comparable; however, it is possible and interesting to consider the reduced role of contextual satisfaction in general life satisfaction.

Finally, our results seem to confirm previous literature on the relationship between academic performance and academic satisfaction (Strahan & Credé, 2015): the weak relationship founded in our and previous study, maybe is also influenced by other variables not considered (such as personality traits or the importance level of domain context in a person's life).

CONCLUSIONS

The aim of our research was to measure students' college satisfaction and its relations with academic performance and general life satisfaction. Considering that the literature did not provide a multifactorial questionnaire short enough to be used for research based on a large number of questionnaires, we decided to develop a new instrument, the College Satisfaction Scale, to assess contextual satisfaction for this target. It assesses five different satisfaction areas: appropriateness of choice (CH), quality of the University services (SE), relationships with colleagues (RE), effectiveness of study habits (ST) and usefulness for future career (CA). The questionnaire, composed of 20 items, showed very good psychometric features, and its relations with both academic performance and life satisfaction confirmed its validity.

The findings of our study should be considered in the light of its limitations. Firstly, the research was conducted on two samples that were not representative of the college population or completely comparable. In future studies the psychometric properties of the CSS should be tested with samples more numerous and more heterogeneous at least for sex, years of college attended and types of courses. This could maybe overcome the absence of predictability between some dimensions of the CSS (i.e. CA) and life satisfaction. Moreover, the cross-sectional nature of the study did not allow us to verify both changes over time and the predictive validity of the subscales; for this reason, longitudinal studies should be implemented. Furthermore, future studies should consider additional criterion variables, such as flourishing (Diener et al., 2009), or perceived quality of universities' services.

One more limit is that we did not assess the importance of academic career and academic performance for the students. According to the literature, a stronger relation between domain-specific measures and overall life satisfaction can be found considering the salience of college in the life of persons and the perceived importance of the student role (Lent & Brown, 2008; Rain et al., 1991). Future research should consider these variables as moderators in order to verify the existence of stronger relations with academic performance and life satisfaction. Despite these limitations, the CSS appears to be adequately valid to help clients to identify one or more potential areas of dissatisfaction from the early stages of students' adaptation to the University context.

Study results can be useful for University professionals and career counsellors because, during the first year, programs and practices can probably improve their effects by focusing on student persistence to prevent dropout (Chen, 2012). Researchers and practitioners can use the CSS to better understand the role of each dimension in improving students' satisfaction, considering its relation with academic success. The CSS can be used in College tutoring services for the identification of the dissatisfaction domains that could interfere with the academic success; in fact, a scientific knowledge and analysis of students' opinions about their satisfaction could help educational administrators to improve and to better change contexts and institutions (Silva, 2001). Moreover, college counsellors and tutors can trust in the subscales to recognize strengths and weaknesses of the relationship between students – especially freshmen students – and University, and to plan specific interventions to increase the quality of the services or the career intervention

at the beginning of the degree courses.

It could be useful to implement career counselling interventions based on improving the satisfaction of students, facilitating a higher level of performance and preventing the dropout phenomenon. Given the centrality of the academic domain for college students, working on the creation of a positive academic adjustment from the first year of University could also increase overall life satisfaction (Lent et al., 2005) and perceived well-being (Cantor & Sanderson, 1999; Diener & Fujita, 1995).

References

ASTIN, A.W. (1993). *What matters in College? Four Critical Years revisited*. San Francisco, CA: Jossey-Bass Publishers.

BALKIS, M. (2013). Academic Procrastination, Academic Life Satisfaction and Academic Achievement: The mediation role of rational beliefs about studying. *Journal of Cognitive and Behavioral Psychotherapies*, *13*(1), 57-74.

BENTLER, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238-246.

BERGER, J.B. & MILEM, J.F. (2000). Organizational behavior in higher education and student outcomes. In J.C. Smart (Ed.), *Higher education: Handbook of theory and research. Vol. XV*. New York: Springer.

BRIEF, A.P. (1998). *Attitudes in and around organizations.* Thousand Oaks, CA: Sage.

BYRNE, B.M. (2010). *Structural equation modeling with Amos: Basic concepts, applications, and programming* (*2nd ed.*). New York, NY: Taylor and Francis Group.

CANTOR, N. & SANDERSON, C.A. (1999). Life task participation and well-being: The importance of taking part in daily life. In D. Kahneman, E. Diener & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology*. New York: Russell Sage.

CAPRARA, G.V., ALESSANDRI, G., TROMMSDORFF, G., HEIKAMP, T., YAMAGUCHI, S. & SUZUKI, F. (2012). Positive orientation across three cultures. *Journal of Cross-Cultural Psychology*, doi:10.1177/0022022111422257

CHAMBEL, M.J. & CURRAL, L. (2005). Stress in academic life: Work characteristics as predictors of student well-being and performance. *Applied Psychology,* doi:10.1111/j.1464-0597.2005.00200.x

CHEN, R. (2012). Institutional Characteristics and College Student Dropout Risks: A Multilevel Event History Analysis. *Research in Higher Education*, doi:10.1007/s11162-011-9241-4

DI FABIO, A. & BUSONI, L. (2009). Psychometric properties of the Italian version of the Satisfaction With Life Scale (SWLS) with University students. *Counseling: Giornale Italiano di Ricerca e Applicazioni*, *2*(2), 201-211.

DI NUOVO, S. (Ed.) (2009). *Studio della variabilità dei percorsi studenteschi per la definizione dei vincoli rimuovibili di sistema e di ambiente formative*. Ministero dell'Istruzione, dell'Università e della Ricerca, Comitato Nazionale per la Valutazione del Sistema Universitario.

DIENER, E., EMMONS, R.A., LARSEN, R.J. & GRIFFIN, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, *49*, 71-75.

DIENER, E. & FUJITA, F. (1995). Resources, personal strivings, and subjective well-being: A nomothetic and idiographic approach. *Journal of Personality and Social Psychology*, doi:10.1037/0022-3514.68.5.926

DIENER, E., WIRTZ, D., TOV, W., KIM-PRIETO, C., CHOI, D., OISHI, S. & BISWAS-DIENER, R. (2009). New measures of well-being: Flourishing and positive and negative feelings. *Social Indicators Research*, *39*, 247-266.

DUNN, T.J., BAGULEY, T. & BRUNSDEN, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399-412. doi: 10.1111/bjop.12046

EDWARDS, J.E. & WATERS, L.K. (1982). Involvement, ability, performance, and satisfaction as predictors of college attrition. *Educational and Psychological Measurement*, doi:10.1177/001316448204200421

EL-HILALI, N., AL-JABER, S. & HUSSEIN, L. (2015). Students' Satisfaction and Achievement and Absorption Capacity in Higher Education. *Procedia - Social and Behavioral Sciences*, doi: 10.1016/j.sbspro.2015.02.384

ELLIOT, A.J., SHELDON, K.M. & CHURCH, M.A. (1997). Avoidance personal goals and subjective well-being. *Personality and Social Psychology Bulletin*, *23*, 915-927.

ELLIOTT, K.M. & SHIN, D. (2002). Student Satisfaction: An alternative approach to assessing this important concept. *Journal of Higher Education Policy and Management*, *24*(2), 197-209, doi: 10.1080/1360080022000013518

HELLER, D., JUDGE, T.A. & WATSON, D. (2002). The confounding role of personality and trait affectivity in the relationship between job and life satisfaction. *Journal of Organizational Behavior*, *23*,

815-835.

HELLER, D., WATSON, D. & ILIES, R. (2004). The role of person versus situation in life satisfaction: A critical examination. *Psychological Bulletin*, *130*, 574-600.

HU, L.T. & BENTLER, P.M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: A multidisciplinary journal*, *6*(1), 1-55.

HUEBNER, E.S. (1994). Preliminary development and validation of a multidimensional life scale for children. *Psychological Assessment*, *6*(2), 149-158.

HUEBNER, E.S. & GILMAN, R. (2006). Students who like and dislike school. *Applied Research in Quality of Life*, *1*, 139-150.

HUEBNER, E.S. & McCULLOUGH, G. (2000). Correlates of school satisfaction among adolescents. *Journal of Educational Research, 93*, 331-335.

JACKSON, D.L., GILLASPY Jr, J.A. & PURC-STEPHENSON, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological methods*, *14*(1), 6-23.

JUDGE, T.A. & ILIES, R. (2004). Affect and job satisfaction: A study of their relationship at work and at home. *Journal of Applied Psychology*, doi:10.1037/0021-9010.89.4.661

JUDGE, T.A. & LOCKE, E.A. (1993). Effect of dysfunctional thought processes on subjective wellbeing and job satisfaction. *Journal of Applied Psychology*, *78*, 475-490.

KANDEMIR, M. (2014). Reasons of academic procrastination: Self-regulation, academic self-efficacy, life satisfaction and demographics variables. *Procedia - Social and Behavioral Sciences*, doi: 10.1016/j.sbspro.2014.09.179

KAREMERA, D., REUBEN, L.J. & SILLAH, M.R. (2003). The effects of academic environment and background characteristics on student satisfaction and performance: The case of South Carolina State University's school of business. *College Student Journal, 37*(2), 298-308.

KUMAR, S.P.K. & DILEEP, P. (2006). *Academic life satisfaction scale and its effectiveness in predicting academic success*. ERIC Document Reproduction Service No. ED491869. Retrieved from ERIC database.

LENT, R.W. (2004). Toward a unifying theoretical and practical perspective on well-being and psychosocial adjustment. *Journal of Counseling Psychology*, doi:10.1037/0022-0167.51.4.482

LENT, R.W. & BROWN, S.D. (2008). Social Cognitive Career Theory and Subjective Well-Being in the Context of Work. *Journal of Career Assessment*, doi:10.1177/1069072707305769

LENT, R.W., BROWN, S.D. & HACKETT, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior*, doi:10.1006/jvbe.1994.1027

LENT, R.W., SINGLEY, H.D., SHEU, U.B., GAINOR, K.A., BRENNER, B.R., TREISTMAN, D. & ADES, L. (2005). Social cognitive predictors of domain and life satisfaction: Exploring the theoretical precursors of subjective wellbeing. *Journal of Counseling Psychology*, doi:10.1037/0022-0167.52.3.429

LENT, R.W., SINGLEY, D., SHEU, H., SCHMIDT, J.A. & SCHMIDT, L.C. (2007). Relation of social-cognitive factors to academic satisfaction in engineering students. *Journal of Career Assessment*, doi:10.1177/1069072706294518

LENT, R.W., TAVEIRA, M., SHEU, U.B. & SINGLEY, H.D. (2009). Social cognitive predictors of academic adjustment and life satisfaction in Portuguese college students: A longitudinal analysis. *Journal of Vocational Behavior*, doi:10.1016/j.jvb.2008.12.006

LEWIS, A., HUEBNER, E.S., MALONE, P.S. & VALOIS, R.F. (2011). Life satisfaction and student engagement in adolescents. *Journal of Youth and Adolescence*, doi:10.1007/s10964-010-9517-6

MA, Y. & CRAGG, K.M. (2012). So close, yet so far away: Early vs. late dropouts. *Journal of College Student Retention: Research, Theory and Practice*, doi:10.2190/CS.14.4.f

MARTINS, F. (1998). *A satisfação acadêmica: Construção de uma escala*. In Anais do 4º Congresso Galaico-Português de Psicopedagogia, Universidade do Minho, Braga-Portugal.

NAUTA, M.M. (2007). Assessing college students' satisfaction with their academic majors. *Journal of Career Assessment*, *15*, 446-462. doi: 10.1177/1069072707305762

NEGRICEA, C.I., EDU, T. & AVRAM, E.M. (2014). Establishing Influence of Specific Academic Quality on Student Satisfaction. *Procedia - Social and Behavioral Sciences*, doi:10.1016/j.sbspro.2014.01.961

NULTY, D.D. (2008). The adequacy of response rates to online and paper surveys: What can be done? *Assessment & Evaluation in Higher Education*, doi: 10.1080/02602930701293231

NURMI, J.E., AUNOLA, K., SALMELA-ARO, K. & LINDROOS, M. (2003). The role of success expectation and task-avoidance in academic performance and satisfaction: Three studies on antecedents, consequences and correlates. *Contemporary Educational Psychology*, doi:10.1016/S0361-476X(02)00014-0

OKUN, M.A., KARDASH, C.A., STOCK, W.A., SANDLER, I.N. et al. (1986). Measuring perceptions of the quality of academic life among college students. *Journal of College Student Personnel*, *27*(5), 447-451.

PAVOT, W. & DIENER, E. (2008). The Satisfaction With Life Scale

and the emerging construct of life satisfaction. *The Journal of Positive Psychology*, doi:10.1080/17439760701756946

PIKE, G.R. (1991). The Effects of Background, Coursework, and Involvement on Student's grades and Satisfaction. *Research in Higher Education*, *32*(1), 15-30.

RAIN, J.S., LANE, I.M. & STEINER, D.D. (1991). A current look at the job satisfaction/life satisfaction relationship: Review and future considerations. *Human Resources*, doi:10.1177/001872679104400305

RAMSDEN, P. (1991). A performance indicator of teaching quality in higher education: The Course Experience Questionnaire. *Studies in Higher Education*, *16*(2), 129-150.

RYAN, J.F. (2004). The relationship between institutional expenditures and degree attainment at baccalaureate colleges. *Research in Higher Education*, *45*(2), 97-114. doi:10.1023/B:RIHE.0000015691.02545.61

ROTTER, N.G. (1988). Student attrition in a technological University: Academic lifestyle. *College Student Journal*, *22*(3), 241-248.

SALMELA-ARO, K. & TUOMINEN-SOINI, H. (2010). Adolescents' life satisfaction during the transition to post-comprehensive education: Antecedents and consequences. *Journal of Happiness Studies*, doi:10.1007/s10902-009-9156-3

SAVITHRI, J.J. (2014). Interactive effect of academic procrastination and academic performance on life satisfaction. *International Journal of Science and Research*, *3*(3), 377-381.

SCHMITT, N., OSWALD, F.L., FRIEDE, A., IMUS, A. & MERRITT, S. (2008). Perceived fit with an academic environment: Attitudinal and behavioral outcomes. *Journal of Vocational Behavior,* doi:10.1016/j.jvb.2007.10.007

SELIGMAN, M.E.P. (2002). Positive psychology, positive prevention, and positive therapy. In C.R. Snyder & S.J. Lopez (Eds.), *Handbook of positive psychology*. New York: Oxford University Press.

SILVA, J.C. (2001). Satisfação no trabalho: Percepções dos gestores e gestoras de escolas secundárias públicas no norte de Portugal. *Gestão em Ação*, *4*(1), 31-38.

SISTO, F.F., BATISTA, M.A., FERREIRA, C.E.B., PAVARINI, M.L.C., OLIVEIRA, J.C.S., OLIVEIRA, R., OLIVEIRA, S.M.S.S. & SANTOS, A. (2002). Avaliando a Satisfação escolar no Ensino Fundamental. In E. Boruchovitch & J.A. Bzuneck (Eds.), *A motivação do aluno*. Petrópolis: Vozes.

SMART, J.C., FELDMAN, K.A. & ETHINGTON, C.A. (2006). Holland's theory and patterns of college student success. In *Commissioned report for the national symposium on postsecondary success: Spearheading a dialogue on student success*. Washington, DC: The National Postsecondary Education Cooperative.

SOARES, A.P.C., VASCONCELOS, R.M. & ALMEIDA, L.S. (2002). Adaptação e Satisfação na Universidade: Apresentação e validação do Questionário de Satisfação Acadêmica. In A.S. Pouzada, L.S. Almeida & Vasconcelos, R.M. (Eds.), *Contextos e dinâmica da vida acadêmica*. Guimarães: Universidade do Minho.

SOVET, L., PARK, M.S. & JUNG, S. (2014). Validation and psychometric properties of Academic Major Satisfaction Scale among Korean college students. *Social Indicators Research*, *119*(2), 1121-1131.

SPADY, W.G. (1971). Dropouts from higher education: Toward an empirical model. *Interchange*, doi:10.1007/BF02282469

STARR, A.M., BETZ, E.L. & MENNE, J.W. (1971). *College Student Satisfaction Questionnaire (CSSQ). Manual*. Ames, IA: Central Iowa Associates.

STAATS, S. & PARTLO, C. (1990). Predicting intent to get a college degree. *Journal of College Student Development*, *31*(3), 245-249.

STRAHAN, S. & CREDÉ, M. (2015). Satisfaction with college: Re-examining its structure and its relationships with the intent to remain in college and academic performance. *Journal of College Student Retention: Research, Theory and Practice*, *16*(4), 537-561.

SULDO, S.M., HUEBNER, E.S., SAVAGE, J. & THALJI, A. (2011). Promoting subjective well-being. In M. Bray & T. Kehle (Eds.), *Oxford handbook of school psychology*. New York: Oxford University Press.

SULDO, S.M. & SHAFFER, E.J. (2008). Looking beyond psychopathology: The dual-factor model of mental health in youth. *School Psychology Review*, *37*, 52-68.

SULDO, S.M., THALJI, A. & FERRON, J. (2011). Longitudinal academic outcomes predicted by early adolescents' subjective well-being, psychopathology, and mental health status yielded from a dualfactor model. *Journal of Positive Psychology*, doi:10.1080/17439760.2010.536774

TRACEY, T.J.G. & ROBBINS, S.B. (2006). The interest-major congruence and college success relation: A longitudinal study. *Journal of Vocational Behavior*, doi:10.1016/j.jvb.2005.11.003

VERKUYTEN, M. & THIJS, J. (2002). School satisfaction of elementary school children: The role of performance, peer relations, ethnicity and gender. *Social Indicators Research*, *59*, 203-228.

WHISMAN, M.A. & JUDD, C.M. (2015). A Cross-National Analysis of Measurement Invariance of the Satisfaction With Life Scale. *Psychological Assessment*, doi:10.1037/pas0000181

WILKINS, K.G., SANTILLI, S., FERRARI, L., NOTA, L., TRACEY, T.J.G. & SORESI. S. (2014). The relationship among positive emotional dispositions, career adaptability, and satisfaction in Italian high school students. *Journal of Vocational Behavior*, doi:10.1016/j.jvb.2014.08.004