19.00 €

vol. LXVI. September-December 2018

N° 283

3/3

Four-monthly Journal
ISSN 0006-6761

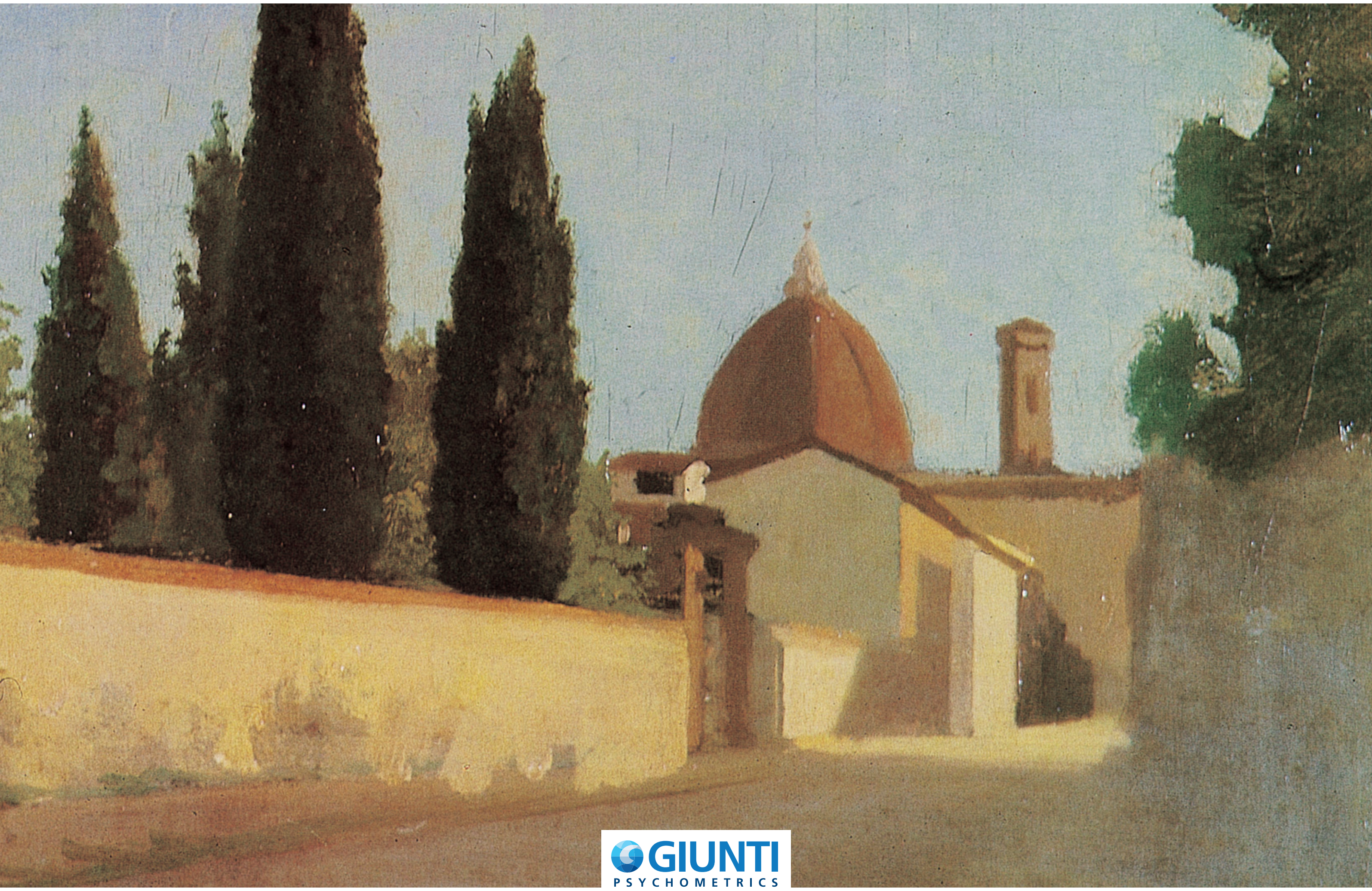
BPA

Bollettino di Psicologia Applicata
APPLIED PSYCHOLOGY BULLETIN

Indexed in PsycINFO® – Scopus Bibliographic Database

Research

Experiences & Tools



GIUNTI PSYCHOMETRICS

Scientific Director Alessandro Zennaro

CONTENTS

Research

Experiences & Tools

DOI: 10.26387/bpa.283.1

The visual perception of volume: Judgment and fixations for objects

Negar Sammaknejad[1], Donald Hoffman[2], Amy Escobar[3], Pete Foley[4], Julie Kwak[5]

[1] Institute for Cognitive and Brain Sciences, Shahid Beheshti University, Tehran, Iran; University of California, Irvine
[2] Department of Cognitive Science, University of California, Irvine
[3] Coastline Community College
[4] Innovation Excellence
[5] University of California, Irvine

ABSTRACT. La presente ricerca ha esplorato, mediante tre esperimenti, quanto i pregiudizi del consumatore e la lunghezza o tipologia di una confezione influenzino le preferenze e l'attenzione nei confronti di un prodotto. I dodici partecipanti sono stati posti di fronte a due immagini di bottiglie posizionate una vicino all'altra sul monitor di un computer Dell Triniton e hanno valutato quale avesse il maggior volume. Sono stati monitorati anche i movimenti oculari, mediante il sistema Eyelink II. I risultati, ottenuti con l'analisi della varianza e il test *t* di Student, hanno confermato l'impatto dei pregiudizi.

SUMMARY. Understanding consumers' perception and judgments of product volume is critical for consumer researchers, package designers, and public health advocates. In this study, in a set of three experiments, observers chose which of two bottle images with different height-to-width ratios depicted greater volume. The elongation bias was replicated and a leftward bias was found. Eye movements were recorded as a measure of attention and pupil dilation was recorded as a measure of cognitive load. Fewer fixations were made to the chosen bottle; the last fixation was more often to the rejected bottle. The top halves of the bottles and the side nearest the alternative bottle receive more attention. There were more fixations, slower responses, and lower confidence for more visually complex bottles. Pupil dilation increased when judging the volume of more complex bottles. The context of a shelf increased confidence in some cases. Implications for packaging design are discussed. .

Keywords: *Eye movements, Decision making, Volume judgment, Left visual field bias, Packaging, Context*

INTRODUCTION

The perceived volume of a bottle or package has important implications for commercial package design, especially in the consumer goods industry. At the most basic level, if two products have similar attributes and price, then shoppers will generally select and purchase the product that is perceived as containing the larger volume. This larger perceived volume implies more product, and hence better perceived value.

There are also potential opportunities associated with influencing perceived volume of packages that go beyond this simple application. For example, there are numerous advantages for creating more concentrated, compact products in many consumer goods categories. Many liquid products, such as detergents, shampoos, and dishwashing liquids have historically contained quite high levels of water. More concentrated products reduce both the financial and environmental cost of packing, shipping and storing these products in a product manufacturing and supply chain. However, consumers can perceive smaller packages a poorer value, even if they contain the same quantity of active ingredient. If this can be lessened by strategic package design for the compacted version, consumers may be more willing to accept compacted products which use less energy and reduce waste, making them legitimate "green" alternatives (Bansal & Roth, 2000).

So there are several potential advantages associated with influencing perceived package volume. An opportunity in this respect lies in our understanding that people are not always accurate in their determination of volume, and that the shape of a package can impact perception of its volume. Many factors can potentially impact perceived volume of two different packages, including three dimensional effects such as body shape, asymmetry, handle shape, curvature, two dimensional effects such as pattern, label shape, geometric complexity, and even the number of displayed packages (Garber, Hyatt & Boya, 2009, 2014). One such factor that is of particular relevance to packaging is the elongation bias, where an increase in the ratio of height versus width creates a perception of greater apparent volume (Been, Braunstein & Piazza, 1964; Frayman & Dawson, 1981; Holmberg, 1975; Kerr, Patterson, Koenen & Greenfield, 2009; Pearson, 1964; Pechey et al., 2015; Raghubir & Krishna, 1999; Wansink & Van Ittersum, 2003; Yang & Raghubir, 2005). This is an effect that has been demonstrated repeatedly in packaging, and also in studies of everyday objects such as drinking glasses, where people repeatedly show a preference for tall, thin glasses over shorter, wider glasses of equal volume, and estimate that the tall, think glasses contain a greater volume (Wansink & Van Ittersum, 2003; Yang & Raghubir, 2005).

The elongation bias is of particular interest in the context of packaging because it appears to robustly and consistently operate across a range of relevant contexts. For example, it is not eliminated by reducing an observer's cognitive load or increasing an observer's motivation to be accurate during volume judgments, suggesting that it is at least in part an automatic process (Raghubir & Krishna, 1999). It is also at least partly robust in the face of expertise. For example, bartenders, when instructed to pour a precise amount into glasses, consistently pour less into elongated, highball glasses. Although the error rate was lower for bartenders than less practiced participants (Wansink & Van Ittersum, 2003), but still persisted. In a related study of purchasing behaviors, Yang & Raghubir (2005) categorized participants as non-drinkers, lighter drinkers and heavier drinkers to reflect their level of experience with buying beer. For all three groups, elongated containers (bottles) were perceived to contain more volume than shorter cans. The effect was strongest for the non-drinkers and weakest for the heavier drinkers. This suggests that this is a tenacious bias that will influence even experienced shoppers, albeit to a potentially lesser degree than less experienced ones.

In the experiments reported here, we have explored the impact of the elongation bias specifically in the context of packages similar to those found in the consumer goods industry. We have tested various prototypes in a context that models to some degree a retail environment such as a supermarket shelf, and evaluated how shape can influence preference as a proxy for shopper purchasing behavior. We expected to replicate the elongation bias, but we have also explored the role of shape, topological properties and holes in volume perception when varied in combination with the elongation bias.

In addition to this, we have also studied eye movements during judgments of relative volume. As mentioned previously, a study by Folkes & Matta (2004) reported that more attention leads to greater judged volume, suggesting that attentional mechanisms may impact volume perception. However, in their study the measure of attention was subjective: where observers reported, using questionnaires, which objects attracted more of their attention. By using eye tracking we expect to explore this hypothesis using a more direct measure of visual attention, and one that encompasses both explicit and implicit attentional effects. Eye tracking is a useful technique to apply in this context, as we know that fixations are often directed to the focus of attention (Deubel & Schneider, 1996; Hoffman & Subramaniam, 1995; Kowler, Anderson, Dosher & Blaser, 1995). Observers more accurately identify simple objects when they are near saccade targets (Deubel & Schneider, 1996; Kowler et al., 1995). Prior to making saccades, observers orient their attention toward the intended target of the saccade (Hoffman & Subramaniam, 1995). Observers find it difficult to orient their attention to one location while making a saccade to a different location (Hoffman & Subramaniam, 1995; Kowler et al., 1995). These tight correlations make fixations a valuable measure of visual attention. When viewing scenes, observers fixate on more informative regions (Loftus & Mackworth, 1978). For instance, when viewing faces, observers fixate more on internal features than the rest of the face (Henderson, Williams & Falk, 2005; Stacey, Walker & Underwood, 2005). Particularly the eye region, which is the most informative region of a face, receives the highest proportion of fixations (Althoff & Cohen, 1999; Barton, Radcliffe, Cherkasova, Edelman & Intriligator, 2006; Walker-Smith, Gale & Findlay, 1977). Given these findings we planned to infer, from fixations, what regions are most informative during judgments of volume.

By exploring the impact of both elongation and topology on preference and attention, we hope to provide a foundation for package design that can increase perceived value between products of equal volume, but also to provide insights that can ultimately be adapted to increase the acceptance of environmentally advantageous 'compact' products.

EXPERIMENT 1

In this experiment, participants viewed two bottles placed side by side and judged which bottle had the greater volume while their fixations were recorded. Participants also made bets to indicate how confident they were in their judgments.

The stimuli were two-dimensional (2D) images of three-dimensional (3D) bottles. Previous studies found that the perceived volume of a 3D object can differ from the perceived volume of a 2D image of that object (Ekman & Junge, 1961; Frayman & Dawson, 1981). However, this is no problem for our experiment since our participants judged relative, not absolute, volume. Using 2D images gave us greater control of our bottle stimuli: we varied their elongation but kept constant their color, shape and area.

EXP. 1 - METHOD

Participants

Twelve participants (six males) with normal or corrected-to-normal vision participated in the study. Participants were paid $10 for their participation. Data from one female participant was excluded because she failed to comprehend the task, and data from one male participant was excluded because he was not naïve to the purpose of the study. The final data set contained data from 10 participants (five males) between the ages of 23 and 29 (M = 25.22, SD = 2.11).

Apparatus

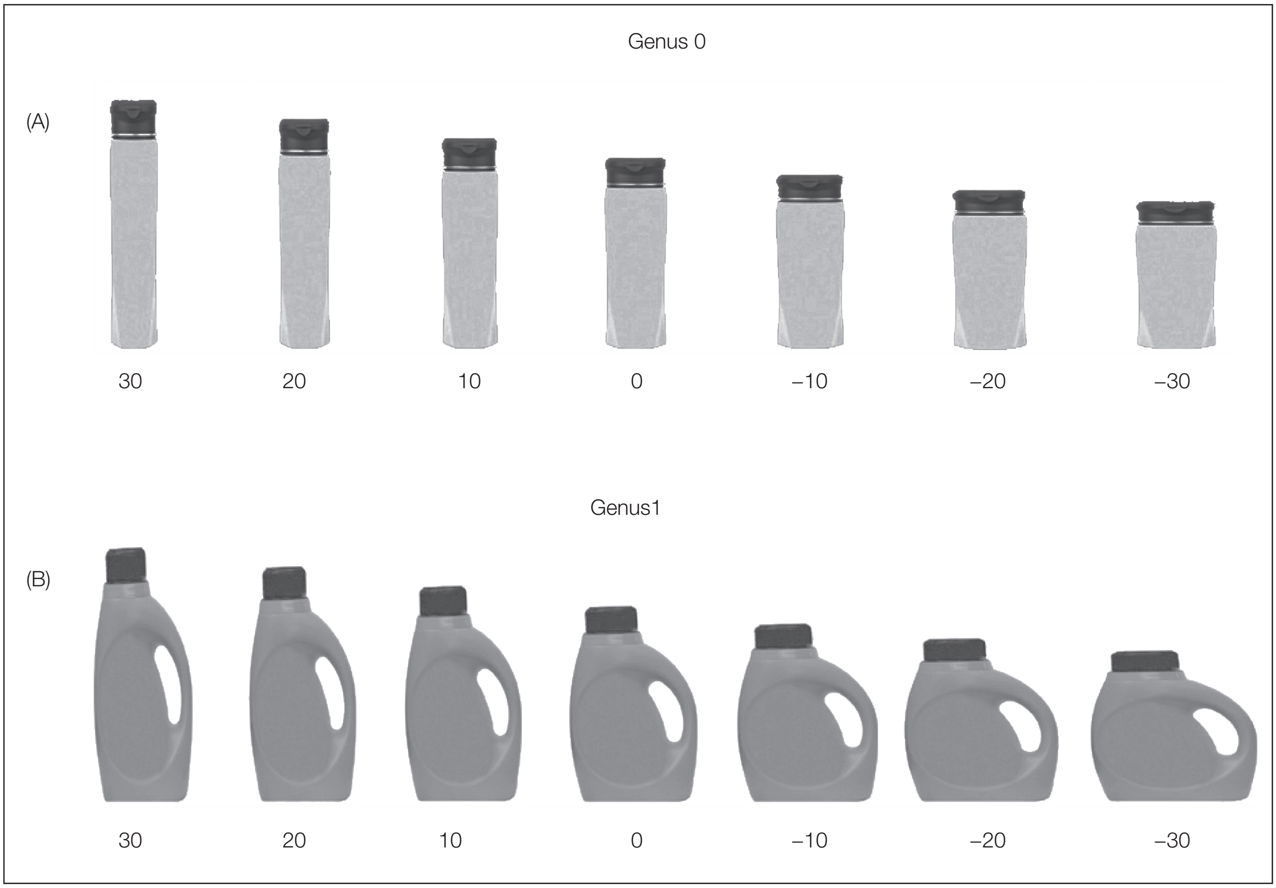
Stimuli were presented in color on a 19 inch Dell Trinitron monitor. Participants sat at about 60 cm from the computer monitor. Eye movements were monitored with the Eyelink II eye tracking system from SR Research.

Materials

The stimuli were created from one original image of a bottle having genus 0 and one original image of a bottle having genus 1, where genus is equivalent to the number of handles an object has. For each original image, three new images were created that had greater elongation: the height was increased by 10, 20 and 30 percent and the widths were decreased so that the surface area was kept constant. Likewise, three new images were created that were less elongated: The height was decreased by 10, 20 and 30 percent and the area kept constant. These six new images are the "altered bottles." Thus there were a total of seven genus 0 images and seven genus 1 images used in the experiment. Figure 1 shows all images, arranged from tallest to shortest.

On each trial, two bottles of the same genus were presented side by side. One was the original and the other was one of the seven bottles of that genus. The original bottle was presented once to the left and once to the right of each of the seven bottles, for a total of 14 pairings for each bottle type.

Figure 1 – Genus 0 bottles used in Experiment 1 (A) and Genus 1 bottles used in Experiment 1 (B)



Note. The percentage of vertical elongation is indicated below each bottle. Genus 0 bottles were off-white, and Genus 1 bottles were red.

To study the effects of context, all 14 pairings were presented twice, once on a shelf and once on a gray background. In total, 28 trials were presented for each genus. The entire experiment consisted of 56 trials presented in random order.

Procedure

Participants sat about 2 feet from the display. They were fitted with the Eyelink II headset, and their fixations were calibrated using Eyelink software. Then a screen appeared with the following instructions: "On each trial, you will see two bottles. Please choose which bottle has the greater volume. If you choose the bottle on the left, press the left arrow key. If you choose the bottle on the right, press the right arrow key. After you press a key, you will be asked to make a bet on how confident you are in your choice. Your bet can be any amount from zero to 100 fake dollars. At the end of the experiment, you will receive real money, up to $10, depending on how well you bet and how accurate your volume choices are".

Participants then pressed any button to begin the experiment. Each trial was self-timed: each pair of bottles was displayed until the participant chose a bottle. A screen then appeared instructing participants to place a bet ranging from zero to 100, where zero indicated no confidence in their choice and 100 indicated the highest confidence. After a participant confirmed the bet amount, the next trial began. A drift-correction dot appeared before each trial to minimize errors in fixation measurements and to center the participant's gaze before the next trial.

After completing the experiment, each participant was told that their performance had earned the full $10 compensation. Participants were not told that there were, in fact, no incorrect answers.

EXP. 1 - RESULTS

Fixations, response times, volume judgments and bets were analyzed for effects of genus, context, elongation, participant gender, and relative bottle location. Fixations and bets for chosen and unchosen bottles were compared, and fixations to different portions of the bottles were analyzed. A four-factor (shelf/no shelf, location, genus, elongation), 2x2x2x7 within-subjects ANOVA was conducted for bets, fixations and response times. Gender was a between-subjects factor in all analysis.

Volume judgments and response times

Results showed that relative location of bottles affected volume judgments ($F_{(1,9)}$ = 8.758, p = .016). Participants more often chose the bottle on the left as having greater volume ($t_{(9)}$ = 2.834, p = .020). There were no other effects on volume judgments or response times.

Fixations

There were more fixations to the unchosen bottle in eight of the ten participants (significant for four participants, two-tailed $t_{(55)}$ = 4.066, 2.469, 2.030, 5.723, all p<.05). On average, 33.5% of fixations were to the chosen bottle, and 38.07% were to the unchosen bottle. Total fixation time was also greater for the unchosen bottle for eight of the ten participants (significant for two, $t_{(55)}$ = 3.592, 4.486, all p<.05). On average, 1975 ms were spent fixating on the chosen bottle, and 2116 ms on the unchosen bottle.

The last fixation was most often to the unchosen bottle ($t_{(9)}$ = 4.628, p = .001). If this bottle was on the right, it received the last fixation on 70% of trials ($t_{(9)}$ = 4.651, p = .0012). If on the left, it received the last fixation on 51% of trials ($t_{(9)}$ = .244, p = .81).

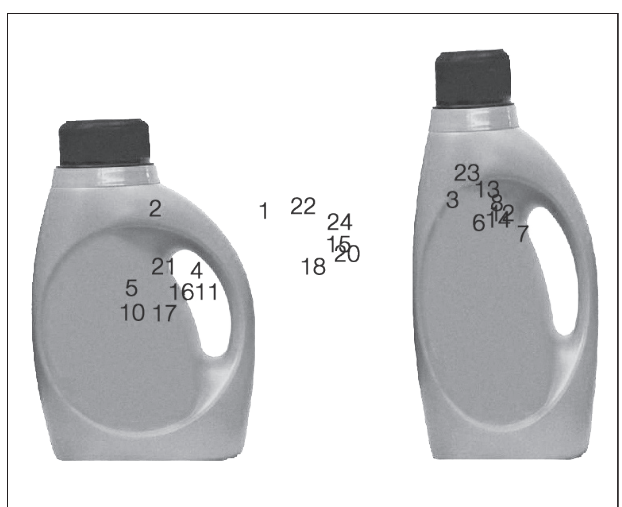
We analyzed the proportion of fixations to the inner side of each bottle, i.e., to the side closest to the other bottle. For all ten participants, two-tailed t-tests showed that the proportion of bottle fixations to the inner side was significantly above 50% ($t_{(55)}$ = 6.929, 9.438, 9.088, 9.365, 13.852, 9.305, 11.253, 13.618, 4.992, 10.951, p<.01 for all participants). The proportion of bottle fixations to the top half of each bottle was also significantly above 50% for all ten participants (p<.01 for all participants). Within-subjects ANOVA showed that this top bias was greater for bottles of genus 0 ($F_{(1,7)}$=5.703, p = .048). Figure 2 shows typical fixations, numbered in sequence.

Bets

A bet was coded with positive sign if an altered bottle was chosen as having greater volume, and with negative sign otherwise. Results showed a significant effect of elongation ($F_{(6,24)}$ = 3.699, p<.01), as shown in Figure 3. For elongations 30, 20 and 10, corresponding to bottles taller and thinner than the original, bets were coded with positive values, indicating that these elongated bottles were chosen. For elongations –10, –20 and –30, corresponding to bottles shorter and wider than the original, bets were coded with negative values, indicating that the original was chosen.

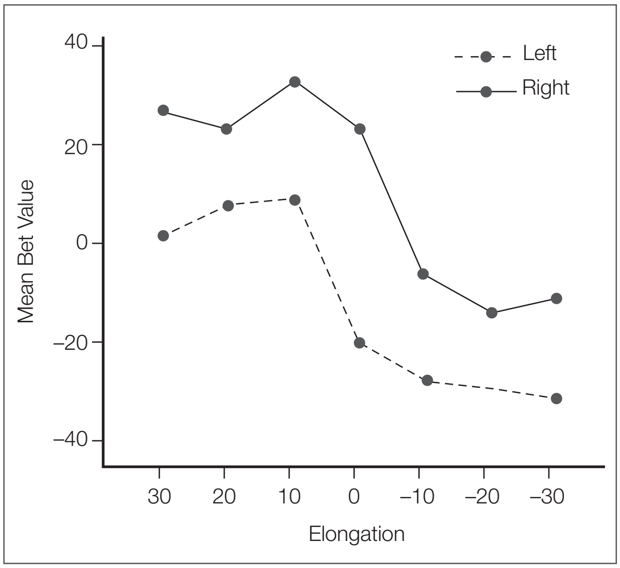
Bets indicate that participants were more confident when they chose the bottle on the left (see Figure 3). A bottle altered to be taller and thinner than the original was likely to be chosen as having greater volume regardless of its location,

Figure 2 – Fixations of one observer during one trial



Note. Fixations are numbered in sequence. There are more fixations to the upper half of the bottle, and to the side of the bottle nearest the other bottle.

Figure 3 – Mean bet values as a function of elongation and location of the original bottle (left or right)



but participants were more confident, as indicated by absolute values of bets, when it appeared on the left. A bottle altered to be shorter and wider was likely not to be chosen regardless of its location, but participants were more confident in rejecting it when it appeared on the right.

The genus of the bottle and the presence or absence of a shelf did not affect bets.

EXP. 1 - DISCUSSION

As expected, elongated bottles were judged to have greater volume. This replicates previous studies, as discussed in the introduction.

Bottle location also influenced judgments of volume. Participants more often chose the bottle on the left as having greater volume, and were more confident when they chose this bottle.

More attention, as measured by number of fixations and total fixation time, was allocated to the bottle that was not chosen as having greater volume.

Regardless of choice or bets, more fixations were made to the top halves of bottles than to the bottom halves. This might be due to the placement of the bottles. As seen in Figure 2, the bottoms of the two bottles are coplanar; however, if the bottles have different elongations then their tops have different heights. Thus the top halves provide more information about the relative heights of bottles, which can be used to estimate relative volumes. This top bias was greater for bottles of genus 0. However, these bottles are relatively cylindrical and have a fairly uniform width, whereas our genus 1 bottles have most of their bulk in the lower half. Thus the upper half of a genus 1 bottle may not be as useful in volume judgments and garners fewer fixations.

More fixations were made to the inner side of each bottle, i.e., to the side closest to the other bottle. This might reflect a strategy for acquiring visual information when making judgments of relative volume. But it might be an artifact of the large separation and spacing between bottles (see Figure 2). The next experiment addresses this issue.

EXPERIMENT 2

As seen in Figure 2, some fixations fell in the empty space between the bottles. Perhaps observers tended to look in the middle of the display and, in consequence, happened to fixate primarily the inner side of each bottle. Experiment 2 studies this issue.

EXP. 2 - METHOD

Participants

Ten observers (five males) with normal or corrected-to-normal vision participated in the study. Observers were paid \$10 for their participation. Observers were between the ages of 20 and 30 (M = 24.2, SD = 3.12).

Apparatus

The apparatus was the same as in Experiment 1.

Materials

Half of the stimuli were those used in Experiment 1. The other half were the same bottle pairs with a decreased distance between bottles. This is illustrated in Figure 4, with

Figure 4 – Genus 0 bottles placed near to each other and in the context of a shelf

two genus 0 bottles placed in the context of a shelf. There were a total of 102 images used: 56 with the original between bottle distance, and 56 with the decreased distance.

Procedure

The procedure was the same as in Experiment 1, except that there were two blocks of trials, one with the original distance between bottles ("far") and one with a decreased distance between bottles ("near"). There were 56 trials in each block, and the order of the two blocks was counterbalanced across participants.

EXP. 2 - RESULTS

A five-factor (placement, shelf/no shelf, location, genus, elongation), 2x2x2x2x7 within-subjects ANOVA was conducted for bets, fixations and response times.

Volume judgments and response times

Bottle location significantly affected bets ($F_{(1,8)} = 9.440$, $p = .015$). On average, the bottle on the left was chosen in 62% of the trials. Response times were longer during trials with genus 1 bottles ($F_{(1,8)} = 9.711$, $p = .014$).

Fixations

As in Experiment 1, observers more often fixated the unchosen bottle, an effect significant for eight of the ten observers ($t_{(111)}$ = 3.912, 7.539, 4.927, 3.477, 3.713, 3.886, 4.136, 2.799, all $p<.01$). 33.1% of fixations were to the chosen bottle, and 40.79% were to the unchosen bottle. The total fixation time to the unchosen bottle was also greater for eight of the ten participants, significantly so for six ($t_{(111)}$ = 3.054, 4.637, 2.303, 3.418, 3.740, 2.840, all $p<.01$). On average, 1310 ms were spent fixating on the chosen bottle, and 1461 ms on the unchosen bottle.

More fixations were made during trials with genus 1 bottles ($F_{(1,8)} = 13.495$, $p = .006$). This was not found in Experiment 1, perhaps because it had half as many trials as Experiment 2, and therefore less power.

As in Experiment 1, we found that more of the last fixations (64%) were made to the unchosen bottle, but only significantly so if it was on the right ($t_{(9)} = 9.239$, $p<.001$). If observers chose the bottle on the left, 72% of the final fixations were to the unchosen bottle; if observers chose the bottle on the right, only 55% of the final fixations were to the unchosen bottle.

All ten observers made more fixations to the inner sides of the bottles ($t_{(112)}$ = 14.212, 18.838, 012.018, 17.816, 11.218, 13.174, 9.825, 15.619, 15.527, 9.783, all $p<.001$). This was affected by placement ($F_{(1,8)} = 9.263$, $p = .016$). The proportion of inner fixations was greater when the bottles were near than when they were far.

Bottle genus affected the proportion of inner fixations ($F_{(1,8)} = 6.031$, $p = .04$), with this proportion being greater for bottles of genus 1 (see Figure 5). This effect was most pronounced for the shorter, wider bottles (elongations –10, –20, and –30).

More fixations (79%) were made to the top halves of bottles. This was significantly greater than 50% for all ten participants (all $p<.05$). There was a significant effect of bottle genus on the proportion of fixations to the top half of the bottle ($p = .03$): the proportion was higher for genus 0 bottles than for genus 1 bottles.

Bets

Elongation had a significant effect on bets ($F_{(6,48)} = 32.950$, $p<.001$). As in Experiment 1, bets were higher when the original bottle appeared on the right ($F_{(1,8)} = 9.440$, $p = .015$).

Figure 5 – Mean difference between number of inner and outer fixations, for Genus 0 and Genus 1 bottles

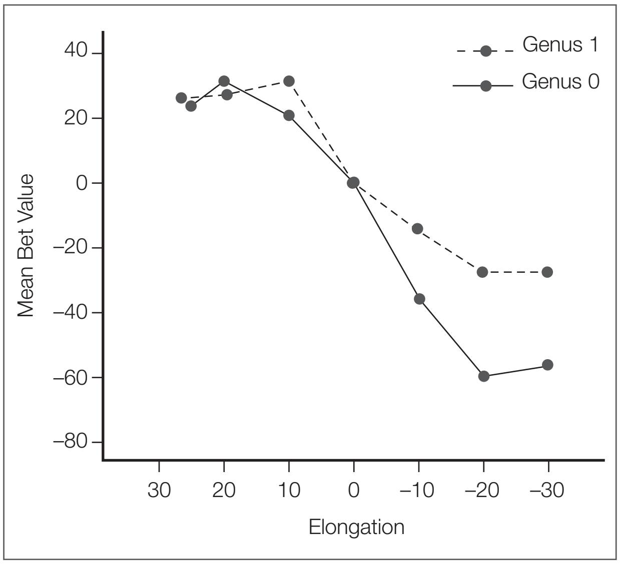
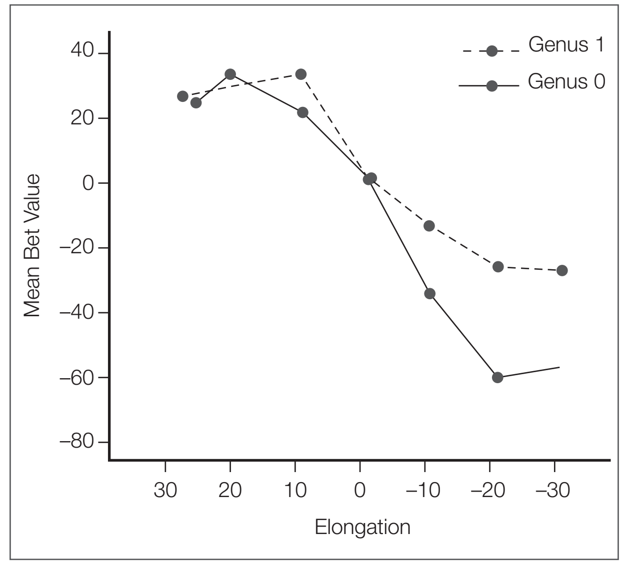


Figure 6 – Mean bets at each elongation for Genus 0 and Genus 1 bottles



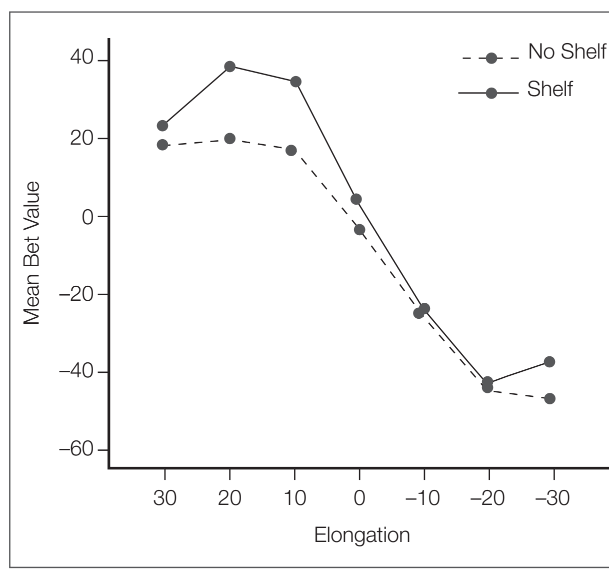
Note. Positive values indicate the altered bottle was chosen; negative values that the original bottle was chosen.

This pattern indicates greater confidence when the bottle on the left is chosen.

Bets were affected by bottle genus ($F_{(1,8)} = 21.037$, $p = .002$). Bets were higher for genus 1 bottles for elongations –10, –20, and –30 (see Figure 6). This effect was not found in Experiment 1, perhaps because Experiment 1 had half as many trials as, and therefore less power.

Context significantly affected bets ($F_{(1,8)} = 12.957$, $p = .007$). For elongations 30, 20, 10, and 0, the bets were higher when there was a shelf present (see Figure 7). These are the taller elongations, and the tops of these bottles were in closer proximity to the shelf above them than were the tops of the shorter bottles.

Figure 7 – Mean bets as a function of elongation and context (shelf or no shelf)



Note. Positive values indicate the altered bottle was chosen.

EXP. 2 - DISCUSSION

Experiment 2 replicated many results found in Experiment 1. Observers showed an elongation bias. They more often chose the bottle on the left, and were more confident when they did. They more often fixated the unchosen bottle, and more often fixated the top halves of bottles.

Observers more often fixated the side of a bottle nearest the other bottle when, as in Experiment 1, the bottles were widely separated. However when the bottles were close together, this effect was even stronger. Thus this fixation pattern cannot be dismissed as an artifact of wide separation between bottles. Instead it reveals an interesting strategy for gathering information when judging relative volume.

Perhaps because Experiment 1 had half as many trials and therefore less power, several effects found in Experiment 2 were not found in Experiment 1.

First, participants made more fixations and were slower to respond during trials with bottles of genus 1. These bottles are more irregularly shaped than the bottles of genus 0, and this extra geometric complexity might require more fixations and computations to judge their volumes. Second, genus affected bets. For the shorter, wider bottles (elongations −10, −20, −30), bets indicate that observers were less confident when choosing bottles of genus 1. This again could be due to the greater geometric complexity of these bottles.

Third, there was a new effect of context. Bets and confidence were higher for bottles with elongations 30, 20, 10 and 0 when they were viewed in the context of a shelf. These elongations correspond to taller bottles, and their height may have made it easier to use the upper shelf as a vertical reference point. Proximity to the upper shelf may have made the bottles look taller than they would without a shelf. The increase in perceived elongation could have increased observers' confidence that the bottles had greater volume.

EXPERIMENT 3

Experiments 1 and 2 studied judgments of relative volume when two objects are visible. However, observers must often judge relative volumes when more than two objects are visible. It is natural to ask whether the patterns of volume judgments found with two objects still holds when more than two objects are visible. Experiment 3 addresses this question, considering the case of four objects. It also investigates the resource demands of volume judgments, using measurements of pupil diameter.

EXP. 3 - METHOD

Participants

Ten observers (five males) with normal or corrected-to-normal vision participated in the study. Observers were paid $10 for their participation. Observers were between the ages of 20 and 34 (M = 23.0, SD = 4.22).

Apparatus

The apparatus was the same as in Experiment 1.

Materials

The stimuli were similar to those used in Experiment 2, with the bottles placed near each other, except that there were four bottles rather than two, and no shelf context was used. The two bottles on the left were identical to each other, as were the two bottles on the right. This is illustrated in Figure 8, with two genus 1 bottles. There were a total of 56 images: 28 with the genus 0 bottles and 28 with genus 1 bottles.

Procedure

The procedure and instructions were the same as the instruction in Experiment 1, except that observers were instructed to judge the relative volumes of the two middle bottles, and the block of 56 trials was presented twice. Trials were presented at random within each block. Two blocks allowed us to study practice effects on volume judgments and pupil diameter.

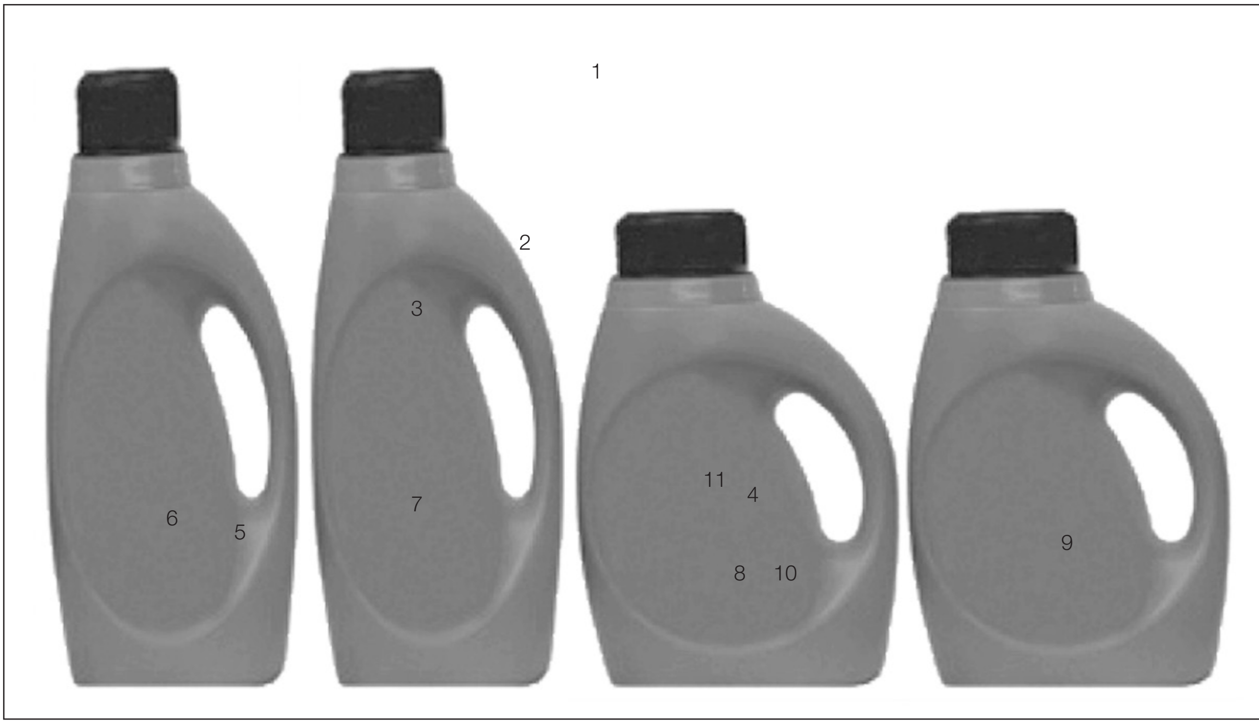
EXP. 3 – RESULTS

A four-factor (block, bottle location, bottle genus, elongation), 2x2x2x7 within-subjects ANOVA was conducted for bets, fixations and response times.

Volume judgments and response times

There was a main effect of block on bets ($F_{(1,8)}$ = 8.390, p = .020) and response times ($F_{(1,8)}$ = 8.815, p = .018); observers bet more confidently in the first block and responded more quickly in the second block. There was a marginal main effect of genus on bets ($F_{(1,8)}$ = 5.241, p = .051); observers bet more confidently on bottles of genus 0. There was a main effect of elongation on bets ($F_{(6,48)}$ = 4.650, p = .007); observers rated more elongated bottles as having greater volume. Notably, unlike Experiments 1 and 2, there was not a main effect of

Figure 8 – A sample stimulus used in Experiment 3, overlaid with fixations from one observer



bottle location on bets; observers no longer demonstrated a left field bias.

Fixations

As in Experiments 1 and 2, observers more often fixated the unchosen bottle; six of the ten observers showed this pattern, significantly so for three ($t_{(55)}$ = 2.422, 3.496, 3.196, all p<.02). On average, 25.95% of fixations were to the chosen bottle, and 28.60% were to the unchosen bottle. Also as in Experiments 1 and 2, observers more often fixated last on the unchosen bottle ($t_{(9)}$ = -3.074, p = .013); this effect was greater if the unchosen bottle was on the left. The total fixation time to the unchosen bottle was also greater for eight of the ten participants, significantly so for three ($t_{(55)}$ = 2.911, 3.613, 2.682, all p<.01). On average, 1580 ms were spent fixating on the chosen bottle, and 1760 ms on the unchosen bottle.

All ten observers made more fixations to the inner sides of the bottles than to the outer sides, significantly so for nine ($t_{(55)}$ = 3.202, 4.176, 5.022, 6.809, 6.148, 7.192, 5.216, 8.817, 6.862, all p<.002). Eight of ten observers made more fixations to the top halves of the bottles, significantly so for five ($t_{(55)}$ = 2.302, 3.488, 2.121, 2.763, 3.834, all p<.04). There was an interaction between top fixations and elongation ($F_{(6,48)}$ = 40.135, p<.001); observers did not preferentially fixate the tops in trials where the two bottles had precisely the same height.

Pupillometry

The mean pupil diameter was larger in the first block of trials than in the second ($F_{(1,8)}$ = 7.142, p = .028), as shown in Figure 9a; so also was the maximum pupil diameter ($F_{(1,8)}$ = 5.822, p = .042). The mean pupil diameter was larger for genus 1 bottles than for genus 0 bottles ($F_{(1,8)}$ = 27.424, p = .001), as shown in Figure 9b.

EXP. 3 - DISCUSSION

Experiment 3 replicated the elongation bias, and the bias found in Experiments 1 and 2 for more inner fixations and top fixations, and for more last fixations to the unchosen bottle.

Experiment 3, unlike Experiments 1 and 2, did not find a left field bias in volume judgments. This might be due to the presence of two extra bottles in each trial of Experiment 3. These extra bottles typically attracted a few fixations, as

Figure 9a – Mean pupil diameter in mm as a function of elongation and block

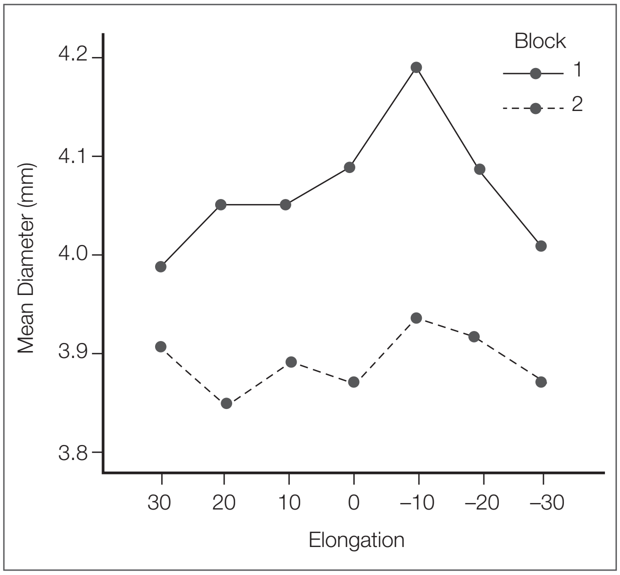
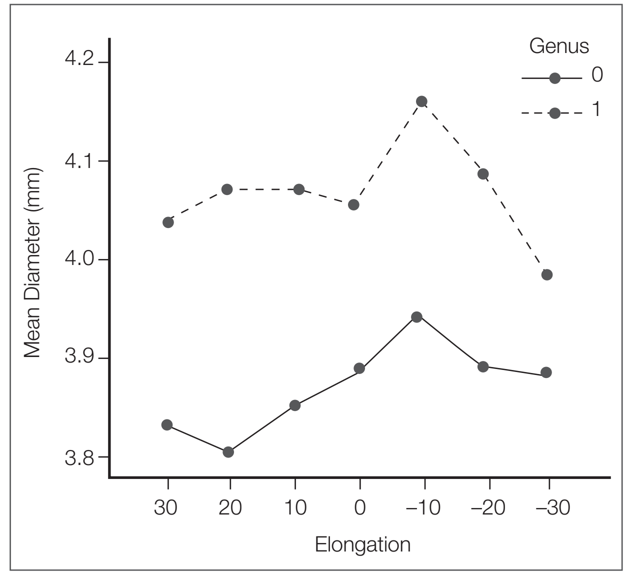


Figure 9b – Mean pupil diameter in mm as a function of elongation and genus of bottle



is seen in Figure 8. When an observer fixates the far left bottle, the middle left bottle is no longer in the left visual field. Similarly, when an observer fixates the far right bottle, the middle right bottle is no longer in the right visual field. This switching of visual fields could smear out the left field biakangs found in Experiments 1 and 2.

Pupil dilation is correlated with increases in attention and cognitive load (e.g., Beatty, 1982; Kang, Huffer & Wheatley, 2014; Peavler, 1974; Siegle, Ichikawa & Steinhauer, 2008). The greater pupil diameter in the first block of trials suggests that the volume judgment task became easier with practice. The greater pupil diameter for genus 1 bottles suggests that volume judgments were more difficult for the more complex bottles.

GENERAL DISCUSSION

Elongation

We replicated the well-known elongation bias: elongated bottles were seen to have greater volume. In addition, we found in Experiment 2 that the elongation effect can be enhanced by placing bottles in the context of a shelf. The shelf above the bottles may act as a vertical reference frame, improving the visual measurement of relative heights. This finding has obvious practical application in stores which display products on shelves.

Location

In Experiments 1 and 2, observers more often chose the bottle on the left as having more volume, and were more confident when they did. A left field bias has been found for other visual capacities, such as face perception (Barton et al., 2006; Gilbert & Bakan, 1973; Mertens, Siegmund & Grüsser, 1993; Phillips & David, 1997) and consumers' judgments of products price (Valenzuela & Raghubir, 2015). Our experiments are the first to suggest a left field bias in judgments of volume.

This bias might reflect hemispheric asymmetries in processing spatial relationships. Judgments of volume rely, one would expect, not just on categorical judgments such as "left of" or "above" but also on estimates of coordinates and distances. Kosslyn et al. (1989) found a left field advantage for processing such coordinate relationships.

The left bias might result from how we match objects to representations in memory. The right hemisphere appears to have an advantage for processing objects with the same basic features as a familiar object, but with an unfamiliar overall shape (Koivisto & Revonsuo, 2003). The altered bottles in our experiments have the same features as the original bottle, but are more or less elongated. Perhaps comparing the altered bottles to a representation in memory of a standard bottle could be done more quickly when the altered bottle is in the left visual field, thus leading to faster and more confident judgments of volume.

Or the left bias might be due to the functioning of two subsystems of visual working memory. One subsystem deals with specific exemplars and the other deals with abstract categories. Marsolek & Burgund (2008) found hemispheric differences in accuracy of judgments based on these subsystems. They presented a cue object followed by a probe object to participants and either asked "Is the probe object the same as the cue object?" or "Is the probe object in the same category as the cue object?". The first question taps into the specific subsystem and the second into the abstract subsystem. When participants were asked the first question, they were more accurate when the probe object was presented in the left visual field. Volume judgments also require a specific metrical comparison, not just a categorical classification. Thus a left field advantage for comparing specific exemplars might facilitate the computation of volume.

The left bias was eliminated in Experiment 3, which had four bottles on each trial rather than just two. The extra bottles attracted some eye fixations. When the far left bottle was fixated, the middle left bottle appeared, momentarily, in the right visual field; when the far right bottle was fixated, the middle right bottle appeared, momentarily, in the left visual field. This switching of visual fields might be responsible for the elimination of the left field bias. Thus, considering that in a real packaging environment, it is rare for two choices to be presented in isolation and shelves are usually crowded in super premium categories, the left visual field bias that was found in experiments 1 and 2 may have more mechanistic and technical value rather than real world potential applications.

Last fixation

More of the last fixations were made to the unchosen bottle. This result raise the question whether a last fixation on a rejected bottle suggest a deselection visual search mechanism. The answer to this question has implications for commercial application, and package design in the context of a shelf and could be explored further in additional studies, where a whole shelf is displayed, and designed to facilitate deselection of competitive or rival products.

Attention to unchosen bottle

Prior studies have found a correlation between greater attention and greater perceived size or volume (Folkes & Matta, 2004). We found the opposite: less attention was correlated with greater perceived volume. This is a surprising result deserving comment.

Folkes & Matta (2004) found that containers which attracted more attention were judged to have greater volume. Their study differs from ours in that their assessment of attention was subjective, based on the self reports of their observers, whereas ours was objective, based on measurements of fixations. Subjective reports might reflect how interesting an object is, rather than how long it holds attention. This is likely in the study by Folkes & Matta (2004), since their containers differed, intentionally, not only in elongation but also in other visual features that affect visual interest.

Anton-Erxleben, Henrich & Treue (2007) presented two moving patterns of random dots. A cue drew attention to one of the patterns, and the observer judged which pattern was larger. They found that the attended pattern was judged to be larger. Their study differs from ours in that their observers judged 2D sizes of dot patterns whereas ours judged volumes of bottles. Moreover, their stimulus presentation was too brief for observers to make a saccade. Our trials were self-timed so observers could fixate as they wished. Thus, Anton-Erxleben et al. (2007) found a correlation between *brief covert* attention and increased perceived size, whereas we found a correlation between *extended overt* attention and decreased perceived volume. This difference in types of attention and their impacts on perceived size or volume deserves further empirical study.

It also deserves further theoretical investigation. Extended overt attention might allow the observer to adopt more sophisticated computational and information-gathering strategies than are possible with brief covert attention. Observers might, for instance, tentatively select one bottle as having greater volume, and then recheck their assessment of the rejected bottle, leading to more fixations of that bottle. This result opens up doors for leveraging the balance between system 1 and 2 decision pathways. For example, could increasing overt attention increase opens to newer more innovative products, whereas decreased overt attention favor more familiar products which require less cognitive engagement, or are more prone to habit derived selection?

In summary, the observation that the chosen bottle has few fixations is surprising and counter intuitive finding for the packaging industry, as heat maps are often used as a proxy for preference. It has important real world implications, and worth further study.

Attention to regions of bottles

In both experiments, more fixations were made to the top halves of bottles than to the bottom halves. The tops of objects have been found to be more salient (Schiano, McBeath & Chambers, 2008). In a matching task, observers were more likely to match objects that had similarly shaped tops (Chambers, McBeath, Schiano & Metz, 1999). For many naturally occurring objects the more informative regions, such as the heads of animals, are at the top.

An alternative explanation is that observers base their judgments of volume on the most salient or reliable information they can gather. The bottles in our experiments stood side by side, with their bottoms coplanar, and with the tops varying in height. Thus the tops were the most informative regions for volume judgments.

Future experiments can test these two hypotheses. For instance, the bottles could be placed one above the other, rather than side by side. According to the first hypothesis, observers should still fixate the tops of bottles. According to the second, observers would fixate the top of the bottom bottle, and the bottom of the top bottle, where the geometry of the two bottles can most easily be compared. The second hypothesis also predicts that in this case there might not be an elongation bias. The widths, not the heights, are the most salient differences when the bottles are placed one above the other. Thus wider, not taller, bottles might be judged as having greater volume. If this were found, it would indicate that the elongation bias is not a fundamental principle in volume perception, but simply an artifact of the side-by-side presentation of the objects to be compared.

Attention to bottles was skewed not only in the vertical dimension. There was also a difference horizontally: most fixations were to the inner side of each bottle, i.e., to the side nearest the other bottle. We wondered if this was due to the large distance between bottles in Experiment 1. However, in Experiment 2, when the bottles were closer, inner fixations actually increased. So, rather than being an artifact of the distance between bottles, this fixation pattern appears to be a strategy that observers use to gather information when judging the relative volumes of two objects placed side by side.

This has important implications for asymmetrical objects. For example, our genus 1 bottles have a handle on one side and the bulk of the volume on the other. If the handle is placed on the side nearest the other bottle, then a strategy of inner fixations might bias observers to sample less from the portion of the bottle that contains most of the volume information. Placement could be a key factor in how volume is perceived for asymmetrical objects.

Bottle genus

More fixations were made to the top halves of genus 0 bottles than to the top halves of genus 1 bottles. The bulk of the genus 1 bottle, and thus most of its volume, is in the bottom half. This could draw the observer's attention downward in an attempt to get information necessary for a volume judgment. The genus 0 bottle is cylindrical, with no extra bulk at the bottom to draw attention.

The bias to inner fixations was greater for genus 1 bottles. These bottles are wider than the genus 0 bottles (for any given elongation), and they have a handle. Future experiments, using different combinations of widths and handles, could determine whether these features influence the bias to inner fixations.

Experiment 2 suggests that volume judgments are more difficult for irregularly-shaped bottles. Observers made more fixations to the irregularly-shaped genus 1 bottles, and took longer to respond to trials with genus 1 bottles. For more complex shapes, observers may need to gather more information to estimate volume. Response times may increase due to longer sampling and calculation times. Experiment 3 supports this interpretation. Pupil diameters were greater for genus 1 bottles, indicating greater cognitive load.

In Experiment 2, genus also affected bets. For the shorter and wider bottles, observers were less confident when choosing bottles of genus 1. This could be due to the greater geometric complexity of these bottles or an asymmetry effect with the handled bottles. In asymmetric bottles, the direct comparison point between the two bottles is a little different in terms of height and slope at the adjacent left/right edges of the bottles where they are directly compared. In this case, the right hand edge of the left bottle is lower and has greater slope than the left hand edge of the right bottle, while the negative space between them can also create an illusion of slope, and potentially height, similar to the Tower of Pisa illusion (Kingdom, Yoonessi & Gheorghiu, 2007). The differences between the symmetric and asymmetric legs would indicate that this might be an effect that is at play, and could open up an interesting direction for additional study with real world benefits.

CONCLUSION

The experiments presented here find that observers, when judging the volumes of bottles placed side-by-side, attend more to the top halves and inner halves of the bottles. This suggests that variations in shape in the top half of a bottle influence apparent volume more than the same variations in the bottom half; similarly, *mutatis mutandis*, for the inner half. This knowledge can be used to design products that optimize the perceived volume of a package.

Moreover, the insights from this research might have potential application in influencing relative choice between products in a category (share of the market), and to offset some of the challenges associated with 1) Compaction and 2) Direct relative comparison at the shelf.

1) *Compaction*. There is a potential to leverage the elongation bias effect in the service of compaction. This is important conceptually, as it has the potential to improve the ecological footprint of a product and package combination. Perceived value issues associated with reduced pack size is one of the biggest barriers to compaction. They are therefore a barrier to the environmental benefits it can bring in terms of reduced fuel, transportation, and storage. The experiments we propose infer this potential. Using the principles we have uncovered via these experiments, the compact package can be designed to maximize the perceived volume of the compacted product. Hence, the volume discrepancy can be lessened for the compacted version and consumers may be more willing to purchase compacted products (green alternatives), which foster a culture of environmental responsibility.
2) *Direct relative comparison at the shelf*. While compaction may have been a conceptual goal, because of the way the experiment is designed, it has even more relevance in influencing simple, relative choice between other similar competing products at the point of purchase. Shoppers can be quite sensitive to small differences when they compare competing packs at the shelf, where direct paired comparisons of relative value are made, and any differences magnified by direct side to side comparisons. All other attributes being equal, relative perceived size, and the perception that "I am getting more for my dollar" will influence value perception (big is better, more is better), and will likely drive choice and purchase towards the pack that is perceived as bigger in a consistent and relatively universal way, and drive market share. Because of its' simplicity, this will also likely be a decision metric that operates even in relatively time constrained, low engagement decisions that are common in a supermarket. The direct, real time choice is also what we are measuring, or at least modeling, in the research.

At the end, it is important to note that in all the experiments, participants were university students. It would be interesting to recruit a wider sample of participants, possibly with a variety in ages, educational background, job, shopping habits, and also to keep track of who is responsible for shopping, either when living with their families or independently.

References

ALTHOFF, R.R. & COHEN, N.J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25* (4), 997-1010. http://dx.doi.org/10.1037//0278-7393.25.4.997

ANTON-ERXLEBEN, K., HENRICH, C. & TREUE, T. (2007). Attention changes perceived size of moving visual patterns. *Journal of Vision*, *7* (11), 1-9. http://dx.doi.org/10.1167/7.11.5

BANSAL, P. & ROTH, K. (2000). Why companies go green: A model of ecological responsiveness. *The Academy of Management Journal*, *43* (4), 717-736. http://dx.doi.org/10.2307/1556363

BARTON, J.J.S., RADCLIFFE, N., CHERKASOVA, M.V., EDELMAN, J. & INTRILIGATOR, J.M. (2006). Information processing during face recognition: The effects of familiarity, inversion, and morphing on scanning fixations. *Perception*, *35* (8), 1089-1105. http://dx.doi.org/10.1068/p5547

BEATTY, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91* (2), 276-292. http://dx.doi.org/10.1037/0033-2909.91.2.276

BEEN, R.T., BRAUNSTEIN, M.L. & PIAZZA, M.H. (1964). Judgment of volume reduction in distorted metal containers. *Journal of Engineering Psychology*, *3* (1), 23-27.

CHAMBERS, K.W., McBEATH, M.K., SCHIANO, D.J. & METZ, E.G. (1999). Tops are more salient than bottoms. *Perception & Psychophysics*, *61* (4), 625-635. http://dx.doi.org/10.3758/BF03205535

DEUBEL, H. & SCHNEIDER, W.X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36* (12), 1827-1837. http://dx.doi.org/10.1016/0042-6989(95)00294-4

EKMAN, G. & JUNGE, K. (1961). Psychophysical relations in visual perception of length, area and volume. *Scandinavian Journal of Psychology*, *2* (1), 1-10. http://dx.doi.org/10.1111/j.1467-9450.1961.tb01215.x

FOLKES, V. & MATTA, Sh. (2004). The effect of package shape on consumers' judgments of product volume: Attention as a mental contaminant. *Journal of Consumer Research*, *31* (2), 390-401. http://dx.doi.org/10.1086/422117

FRAYMAN, B.J. & DAWSON, W.E. (1981). The effect of object shape and mode of presentation on judgments of apparent volume. *Perception & Psychophysics*, *29* (1), 56-62. http://dx.doi.org/10.3758/BF03198840

GARBER, L.L., HYATT, E.M. & BOYA, Ü.Ö. (2009). The effect of package shape on apparent volume: and exploratory study with implications for package design. *Journal of Marketing Theory and Practice*, *17* (3), 215- 34. http://dx.doi.org/10.2753/MTP1069-6679170302

GARBER, L.L., HYATT, E.M. & BOYA, Ü.Ö. (2014). The perceived size of packages of complex vs. simple shape depends upon the number of packages presented. *International Journal of Management Practice*, *7* (2). http://dx.doi.org/10.1504/IJMP.2014.061475

GILBERT, C. & BAKAN, P. (1973). Visual asymmetry in perception of faces. *Neuropsychologia*, *11* (3), 355-362. http://dx.doi.org/10.1016/0028-3932(73)90049-3

HENDERSON, J.M., WILLIAMS, C.C. & FALK, R.J. (2005). Eye movements are functional during face learning. *Memory & Cognition*, *33* (1), 98-106. http://dx.doi.org/10.3758/BF03195300

HOFFMAN, J.E. & SUBRAMANIAM, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics*, *57* (6), 787-795. http://dx.doi.org/10.3758/BF03206794

HOLMBERG, L. (1975). The influence of elongation on the perception of volume of geometrically simple objects. *Psychological Research Bulletin*, *15* (2), 1-18.

KANG, O.E., HUFFER, K.E. & WHEATLEY, T.P. (2014). Pupil dilation dynamics track attention to high-level information. *PLoS ONE*, *9* (8), e102463. http://dx.doi.org/10.1371/journal.pone.0102463

KERR, W.C., PATTERSON, D., KOENEN, M.A. & GREENFIELD, T.K. (2009). Large drinks are no mistake: Glass size, not shape, affects alcoholic beverage drink pours. *Drug Alcohol Review*, *28* (4), 360-365. http://dx.doi.org/10.1111/j.1465-3362.2009.00056.x

KINGDOM, F.A., YOONESSI, A. & GHEORGHIU, E. (2007). The leaning tower illusion: A new illusion of perspective. *Perception*, *36* (3), 475-477. http://dx.doi.org/10.1068/p5722a

KOIVISTO, M. & REVONSUO, A. (2003). Object recognition in the cerebral hemispheres as revealed by visual field experiments. *Laterality: Asymmetries of Body, Brain, and Cognition*, *8*, 135-153. http://dx.doi.org/10.1080/713754482

KOSSLYN, S.M., KOENIG, O., BARRETT, A., CAVE, C.B., TANG, J. & GABRIELI, J.D. (1989). Evidence for two types of spatial representations: Hemispheric specialization for categorical and coordinate relations. *Journal of Experimental Psychology: Human Perception and Performance*, *15* (4), 723-735. http://dx.doi.org/10.1037/0096-1523.15.4.723

KOWLER, E., ANDERSON, E., DOSHER, B. & BLASER, E. (1995). The role of attention in the programming of saccades. *Vision Research*, *35* (13), 1897-1916. http://dx.doi.org/10.1016/0042-

6989(94)00279-U

LOFTUS, G.R. & MACKWORTH, N.H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *4* (4), 565-572. http://dx.doi.org/10.1037/0096-1523.4.4.565

MARSOLEK, C.J. & BURGUND, E.D. (2008). Dissociable neural subsystems underlie visual working memory for abstract categories and specific exemplars. *Cognitive, Affective & Behavioral Neuroscience*, *8* (1), 17-24. http://dx.doi.org/10.3758/CABN.8.1.17

MERTENS, I., SIEGMUND, H. & GRÜSSER, O.J. (1993). Gaze motor asymmetries in the perception of faces during a memory task. *Neuropsychologia*, *31* (9), 989-998. http://dx.doi.org/10.1016/0028-3932(93)90154-R

PEARSON, R.G. (1964). Judgment of volume from photographs of complex shapes. *Perceptual and Motor Skills*, 18 (3), 889-900. http://dx.doi.org/10.2466/pms.1964.18.3.889

PEAVLER, W.S. (1974). Pupil size, information overload, and performance differences. *Psychophysiology*, *11* (5), 559-566. http://dx.doi.org/10.1111/j.1469-8986.1974.tb01114.x

PECHEY, R., ATTWOOD, A.S., COUTURIER, D.L., MUNAFÒ, M.R., SCOTT-SAMUEL, N.E., WOORD, A. & MARTEAU, Th. (2015). *PLoS ONE*, *10* (12), e0144536. http://dx.doi.org/10.1371/journal.pone.0144536

PHILLIPS, M.L. & DAVID, A.S. (1997). Viewing strategies for simple and chimeric faces: An investigation of perceptual bias in normal and schizophrenic patients using scan paths. *Brain and Cognition*, *35* (2), 225-238. http://dx.doi.org/10.1006/brcg.1997.0939

RAGHUBIR, P. & KRISHNA, A. (1999). Vital dimensions in volume perception: Can the eye fool the stomach? *Journal of Marketing Research*, *36* (3), 313-326. http://dx.doi.org/10.2307/3152079

SCHIANO, D.J., McBEATH, M.K. & CHAMBERS, K.W. (2008). Regularity of symmetry vertically guides perceptual judgments of objects. *The American Journal of Psychology*, *121* (2), 209-227. http://dx.doi.org/10.2307/20445457

SIEGLE, G.J., ICHIKAWA, N. & STEINHAUER, S. (2008). Blink before and after you think: Blinks occur prior to and following cognitive load indexed by pupillary responses. *Psychophysiology*, *45* (5), 679-687. http://dx.doi.org/10.1111/j.1469-8986.2008.00681.x

STACEY, P.C., WALKER, S. & UNDERWOOD, J.D. (2005). Face processing and familiarity: Evidence from eye-movement data. *British Journal of Psychology*, *96* (4), 407-422. http://dx.doi.org/10.1348/000712605X47422

VALENZUELA, A. & RAGHUBIR, P. (2015). Are consumers aware of top-bottom but not of left-right inferences? Implications for shelf space positions. *Journal of Experimental Psychology*, *21* (3), 224-241. http://dx.doi.org/10.1037/xap0000055

WALKER-SMITH, G.J., GALE, A.G. & FINDLAY, J.M. (1977). Eye movement strategies involved in face perception. *Perception*, *6* (3), 313-326. http://dx.doi.org/10.1068/p060313

WANSINK, B. (1996). Can package size accelerate usage volume? *Journal of Marketing*, 60 (3), 1-14. http://dx.doi.org/10.2307/1251838

WANSINK, B. & van ITTERSUM, K. (2003). Bottoms up! The influence of elongation on pouring and consumption volume. *Journal of Consumer Research*, *30* (3), 455-463. http://dx.doi.org/10.1086/378621

YANG, Sh. & RAGHUBIR, P. (2005). Can bottles speak volumes? The effect of package shape on how much to buy. *Journal of Retailing*, *81* (4), 269-281.

DOI: 10.26387/bpa.283.2

Validation of the Italian version of the Need for Cognition Scale – Short Version

Antonio Aquino[1], Laura Picconi[2], Francesca Romana Alparone[1]

[1] Department of Neuroscience, Imaging and Clinical Science, Chieti-Pescara University
[2] Department of Psychological, Health and Territory Sciences, Chieti-Pescara University

ABSTRACT. Il contributo è finalizzato a fornire una validazione italiana della scala del Need for Cognition – Short Version (Cacioppo, Petty & Kao, 1984), uno strumento che misura le differenze individuali nella motivazione ad impegnarsi e apprezzare attività che richiedono uno sforzo cognitivo. L'interesse verso questa scala nasce dal suo utilizzo in diversi campi di ricerca, quali la persuasione, la percezione sociale, la psicologia politica. Sia l'analisi esplorativa che l'analisi confermativa, condotte su un campione di 508 partecipanti, hanno evidenziato l'emergere di due fattori legati a due differenti motivazioni: l'approccio alle situazioni cognitive e l'evitamento di situazioni che richiedono uno sforzo cognitivo. Le buone proprietà psicometriche della scala consentono di usare questo strumento in differenti ambiti sia di ricerca che scolastici.

SUMMARY. This research aims at providing an Italian validation of the Need for Cognition Scale – Short Version (NCS). This instrument measures individual differences in the motivation to enjoy effortful cognitive activities. NCS was administered to 508 students, equally distributed by gender (Mean age = 20.78; SD = 1.75). The Italian version of the scale, translated and adapted from the original version, is composed of 18 items on a 7-point Likert scale. An exploratory factorial analysis (Random split sample = 254) and a confirmatory factorial analysis (Random split sample = 254) proved that the scale had two correlated factors measuring two different kind of motivations (approach and avoidance of effortful cognitive activities). Results also indicated that the NCS had good reliability indices and satisfactory discriminant and convergent validity. Thanks to its good psychometric properties, the NCS has been proven to be a reliable tool in both educational and research areas.

Keywords: *Need for cognition, Approach, Avoidance*

INTRODUCTION

The *Need for Cognition* (*NC*; Cacioppo & Petty, 1982; Cacioppo, Petty, Feinstein & Jarvis, 1996) refers to individual differences in the motivation to enjoy and engage in effortful cognitive activities. Individuals with lower intrinsic motivation to think are characterized as cognitive misers, whereas individuals possessing higher intrinsic motivation to reasoning are thought to be chronic thinkers. Extensive research has showed that the NC affects different cognitive processes, including decision making, information processing, evaluating and recalling. In relation to the decision making, those high in NC tend to overthink available options prior to making a final decision. Furthermore, those who are high in NC have more positive attitudes toward tasks that require reasoning (e.g., recalling an information) and make more frequent or more extensive experiences using technologies that require effortful thinking (e.g., computer-aided instruction). Similarly, individuals high in NC are more influenced by the quality of arguments concerning a persuasive message processing compared to individuals low in NC. According to the Elaboration Likelihood Model of Persuasion (Petty & Cacioppo, 1986), in fact, people that are relatively unmotivated or unable to carefully/thoroughly process a persuasive message appear to be influenced by heuristic cues in the persuasion setting (e.g., how attractive the message source is). Research supported the idea that NC acts as a motivational drive to thinking and has shown that individuals low in NC are more influenced by heuristic cues than individuals high in NC (see Cacioppo et al., 1996, for a review). An alternative model offers a single-route reconceptualization that treats the dual routes to persuasion as involving functionally equivalent types of evidence from which persuasive conclusions may be drawn (Kruglanski & Thompson, 1999). However, also in the single-route model the NC is recognized as a motivation in determining the extent to which available evidence gets processed.

Since Cacioppo and Petty (1982) described the NC as a stable individual difference, they developed a 34-items scale for its assessment (*Need for Cognition Scale*, *NCS*), characterized by a single dominant factor as resulted from the Principal Component Analysis (PCA). Cacioppo et al. (1984) subsequently reduced the NCS to 18 items, based on those items with the highest factor loadings. Half of the items reflect a preference for effortful cognitive endeavours (e.g., "I really enjoy a task that involves coming up with new solutions to problems"), whereas the remaining items reflect the absence of such preference (e.g., "Thinking is not my idea of fun"). PCA on these 18 items extracted a single dominant factor that explained the 37% of the variance, with a high level of internal consistency (Cronbach's alpha = .90). Other authors have previously supported such one-dimension structure (e.g., Furlong, 1993), based on the PCA and the reliability index. However, the fact that all items of a scale load positively on a first unrotated factor, and that factor accounts for a moderate proportion of the total variance, does not preclude the emergence of two or more interpretable factors, after rotation (Fabrigar, Wegener, MacCallum, & Strahan, 1999). Specifically, Stark, Bentley, Lowther and Shaw (1991) proposed a bi-factorial solution with a differentiation between the items reflecting an approach to cognitive effortful activities and those reflecting an avoidance of the cognitive activities. This solution has become predominant in last years. Relevant for the present paper, Forsterlee and Ho (1999) performed PCA followed by oblique rotation on the 18-item NFC and they reported a two-factor solution with the differentiation between the approach and the avoidance dimension. The 2 factors resulted highly correlated (r = .52). Similarly, Bors, Vigneau and Lalande (2006) reported a two-factor model for the French version of the scale with the differentiation between the approach and the avoidance dimension. Interestingly, the authors found out that only the avoidance dimension of NC was predictive of the academic success, supporting the idea that the approach and the avoidance are separate constructs of the NC. Recently, Zhang, Noor and Savalei (2016) performed a parallel analysis on NCS and the plot clearly indicated the bi-dimensional solution. In psychological research, however, the differentiation between the approach and the avoidance dimensions has already been widely accepted. A long-standing tradition of psychological theory and research suggests that these two motivations are at least somewhat distinct and, therefore, both motivations should be addressed separately (see Maio, Haddock & Verplanken, 2018, for a review).

To sum up, despite the one-dimensional solution has long been considered the best solution for the NCS, in the last decades the bi-factorial solution with the differentiation between the approach and the avoidance dimension of the NC predominates. Although several studies have used the NCS, to the best of our knowledge, researchers have not directly tested the NCS structure in the Italian context. The present study, therefore, aims to provide a NC scale for the Italian context and to test its structure and validity.

AIMS AND OBJECTIVES

In the present research we addressed the study of the validation of the NCS (Short Version) in the Italian context. In particular, we aimed: 1) to test the NCS factor structure in an Italian sample; 2) to test the reliability of NCS in terms of internal consistency; 3) to investigate the relationship between the NC and other measures of cognition. More precisely, we explored the relationship between the NC and the cognitive dimension of the *Motivated Consumer Innovativeness* (*MCI*; Vandecasteele & Geuens, 2010), that is the extent to which an individual is oriented to buy new products for the desire to be mentally stimulated. We expected the CCI to correlate only with the approach dimension of the NC, given that both these dimensions reflect an approach to objects requiring effortful cognitive activities.

Furthermore, we explored the relationships among the dimensions of the NC and the *Need for Cognitive Closure* (*NCC*; Krusglanski, 1990), that is a cognitive-motivational content independent construct, defined as preference for definitive order and structure, a desire for firm or stable knowledge and a desire to figure out quick-fix solutions. Antecedents of this epistemic motivation are to be found in certain specific conditions that highlight the cost of openness and the benefits of closure (e.g. time pressure, ambient noise, mental fatigue). Past studies showed that NCC is negatively related to NC (Cacioppo et al., 1996), but a possible different relation with the approach and the avoidance dimension of NC has not been investigated yet. It could be reasonable to expect that this relationship is mainly driven by the avoidance dimension, given that this dimension reflects a tendency to avoid situations requiring long reasoning and a preference for a fast solution. We expected low or no correlation between the NCC and the approach dimension of NC.

The differentiation between the approach and the avoidance is not confined to the NC but it is present in other psychological constructs, as, for instance, the *Need for Affect* (*NA*; Maio & Esses, 2001), that is a motivation to approach emotional situations. Literature in this field showed a positive relationship between the total score of NA and NC, suggesting that NC also involves openness to emotional experience (Maio et al., 2018). To the best of our knowledge, nobody investigated the relationship between the approach and the avoidance dimensions of NA and NC. We expected the approach dimensions of these two scales to correlate with each other. Similarly, we expected the avoidance dimensions in the two scales to correlate (with each other) as well.

METHOD

Participants and procedure

The sample included 508 participants, with a mean age of 20.78 years (*SD* = 1.75, *range* = 19-36). Of these participants, 302 were females (59.40%). All participants had a high-school diploma, (4.5% of the sample further achieved the BA-degree).

The Italian version of the NCS was assessed both via an online procedure and a pen-pencil procedure. The students attending the University of Chieti and the University of Caserta were invited to enrol in research regarding attitudes and to complete an online (or a pen-pencil) questionnaire. In the first page, participants were informed that participation was voluntary, and that data were collected anonymously and used for research purposes only. The first section of the questionnaire aimed to assess demographic characteristics (i.e., age, gender, instruction). Then, all participants completed the Italian translation version of 18-items NCS (Cacioppo et al., 1984). In order to translate the items of the scale, a back-translation method was used. The original items of the scale and the translated ones are presented in Table 1.

Additionally, a sub-sample of 70 participants also completed the scales necessary to assess the convergent and divergent validity of the NCS. At the end of the questionnaire participants were thanked and debriefed.

Measures

– *Need for Cognition*. Participants rated the extent to which they agreed with the translated items of the approach (e.g., "I really enjoy a task that involves coming up with new solutions to problems", α = .79) and the avoidance dimension (e.g., "Thinking is not my idea of fun", reverse scored, α = .77). Participants responded to these statements on a 7-point scale from 1 = totally disagree to 7

Table 1 – Translated items of the NCS

Translated (and original) items of the NCS
NC1 - Preferisco i problemi complessi a quelli semplici (I prefer complex to simple problems)
NC2 - Mi piace avere la responsabilità di occuparmi di una situazione che richiede lunghi ragionamenti (I like to have the responsibility of handling a situation that requires a lot of thinking)
NC6 - Provo soddisfazione a riflettere lungamente ed intensamente per ore (I find satisfaction in deliberating hard and for long hours)
NC10 - Mi piace l'idea di fare strada facendo affidamento sul mio pensiero per raggiungere il massimo (The idea of relying on thought to make my way to the top appeals to me)
NC11 - Mi piacciono veramente i compiti che richiedono di escogitare nuove soluzioni ai problemi (I really enjoy a task that involves coming up with new solutions to problems)
N*C13* - Preferisco che la mia vita sia piena di problemi da risolvere (I prefer my life to be filled with puzzles I must solve)
NC14 - Mi attira l'idea di pensare in modo astratto (The notion of thinking abstractly is appealing to me)
NC15 - Preferirei un compito intellettuale, difficile ed importante, piuttosto che uno che sebbene importante non richieda molte riflessioni (I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought)
N*C18* - Di solito finisco col riflettere sui problemi anche quando non mi riguardano personalmente (I usually end up deliberating about issues even when they do not affect me personally)
NC3re - Pensare non corrisponde all'idea che ho del divertimento (Thinking is not my idea of fun)
NC4re - Preferirei fare qualcosa che richieda poche riflessioni piuttosto che qualcosa che sicuramente rappresenti una sfida alle mie capacità cognitive (I would rather do something that requires little thought 1than something that is sure to challenge my thinking abilities)
NC5re - Cerco di prevenire ed evitare situazioni in cui ci sia un'elevata probabilità di dover riflettere a fondo su qualche argomento (I try to anticipate and avoid situations where there is a likely chance I will have to think in depth about something)
NC7re - Penso solo tanto quanto basta (I only think as hard as I have to)
NC8re - Preferisco pensare a piccoli progetti quotidiani piuttosto che a progetti a lungo termine (I prefer to think about small daily projects to long term ones)
NC9re - Mi piacciono quei compiti che richiedono poca riflessione dopo avere imparato a svolgerli (I like tasks that require little thought once I've learned them)
NC12re - Non mi eccita granché imparare nuovi modi di pensare (Learning new ways to think doesn't excite me very much)
N*C16re* - Mi sento più sollevato che soddisfatto dopo aver terminato un lavoro che mi ha richiesto un grande sforzo mentale (I feel relief rather than satisfaction after completing a task that requires a lot of mental effort)
NC17re - Mi basta sapere che qualcosa abbia permesso di concludere il lavoro; non mi interessa come o perché funzioni (It's enough for me that something gets the job done; I don't care how or why it works)

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

= totally agree. For the approach dimension, we computed a score as the mean of the items, reflecting the approach dimension, so that higher scores indicate higher tendency to approach cognitive tasks. Similarly, we computed a score for the avoidance dimension by reverse scoring the avoidance items and calculating their mean. Consequently, for the avoidance dimension, a higher score indicated a minor tendency to avoid cognitive situations.

– *Cognitive Consumer Innovativeness*. The Italian version of the CCI Scale (Caricati & Raimondi, 2015; α = .85) comprises 6 items which measure the consumer innovativeness motivated by the desire to engage in mentally stimulating activities (e.g., "I often buy new products that make me think logically"). Participants responded to these statements on a 7-point scale from 1 = totally disagree to 7 = totally agree. A final score was computed as the mean of the items.
– *Need for Affect*. Participants' NFA was assessed with the short version of the NFA Scale (Appel, Gnambs & Maio, 2012). This scale comprises 10 items: five items measure the motivation to approach emotions (e.g., "Emotions help people to get along in life", α = .79), and five assess the motivation to avoid emotions (e.g., "I do not know how to handle my emotions, so I avoid them", α = .74). Participants responded to these statements on a 7-point scale from 1 = totally disagree to 7 = totally agree. Similarly to the NCS, we computed a mean score for the approach dimension and a mean score for the avoidance dimension by reversing the avoidance items. We selected the 10 items from the Italian version of the NFA Scale (Leone & Presaghi, 2012).
– *Need for Cognitive Closure*. The Italian version of the Revised NCC Scale (Pierro & Kruglanski, 2005; α = .81) comprises 14 items measuring a desire to look for a fast solution. Participants rated each item on a 7-point scale, with a higher value representing a higher NCC. A final score was computed as the mean of the items.

Data analysis

A preliminary analysis of the NCS scale was performed with the support of IBS SPSS Statistics for Windows, Version 22.0 (2012), in order to check the normal distribution by calculating mean, standard deviation, and indices of skewness and kurtosis. Inspection of skewness and kurtosis indicated that departures from normality were not severe (the indices were between −1.20 and 1.56), so no variable transformations were deemed necessary except for item 10 (see next section for more detailed information about this item). The sample was randomly divided into two samples of similar size. Random sample I (N = 254) was used to conduct an exploratory factor analysis (EFA) and data from the second split sample (N = 254) were used to conduct a confirmatory factor analysis (CFA). Through this methodology, the first sample can be used to develop a good fitting solution, and the final model is then fitted in the second sample to determine its replicability with independent data. The investigation of the factorial structure of the NCS (EFA) was performed through a Maximum Likelihood Factor Analysis (MLFA), with an Oblimin rotation to test whether the factors were related to each other. Confirmatory Factor Analysis (CFA) was conducted with EQS 6.0, allowing for correlation among error terms. To evaluate the CFA models, goodness of fit was estimated by Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI), Standardized Root Mean Square Residual (SRMR) and the Non-Normed Fit Index (NNFI). A Maximum Likelihood (ML) method of estimation was applied to test the hypothesized model. The Akaike Information Criterion (AIC) was used to compare the relative fit of models, with lower AIC values indicating superior model fit. Competing models were compared with regard to their model fit by performing a χ^2 difference test. If this difference is significant, the model with lower χ^2 is the best fit model, otherwise, if the difference in χ^2 is not significant, the more parsimonious model (i.e. the model with less parameters) is preferred (Bollen, 1989; Kline, 1998; Schermelleh-Engel, Moosbrugger & Muller, 2003). To compare the competitive models, we also used the difference in CFI (difference ≥.001 indicates better fit to data; Wang, 2015). Internal consistency was estimated by Cronbach's alpha and mean total correlations corrected item.

RESULTS

Factorial structure of the NCS

An exploratory factor analysis was performed on the NCS items in the Random Sample 1. To determine the appropriateness of factor analysis, we examined the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and the Bartlett's test of sphericity. According to Tabachnick and Fidell (2007), KMO should be >.80, and the chi-square value of Bartlett's test should be significant. Both indices confirmed the adequacy of the sample: KMO = .80; χ^2 Bartlett (153) = 1184.33, *p*<.001. To determine the optimal number of factors to retain (i.e., the best trade-off between under- and over-factoring; see Fabrigar et al., 1999) we used the parallel analysis (Horn, 1965), as well as the theoretical basis of the different solutions. In the parallel analysis a set of eigenvalues is computed from randomly generated correlation matrices. These values can then be compared to eigenvalues extracted from the researcher's dataset. The number of factors to retain will be the number of eigenvalues (generated from the researcher's dataset) that are larger than the corresponding random eigenvalues (Horn, 1965). The EFA showed that the bi-factorial solution was more suitable to the data that the mono-factorial solution. However, given that item 10 and item 16 had not adequate loadings on any factor, we decided to run again the factorial analysis without these two items. The parallel analysis without these two items confirmed that the bi-factorial solution was the best solution for the data: only the first two eigenvalues obtained from real data (respectively 4.44 and 2.11) were greater than randomly generated eigenvalues. The rotated bi-factorial solution accounted for the 40% of the total variance (the first factor explained the 18% of the variance, the second one explained the 22% of the post-rotation variance). All items had loadings greater than .30 on the intended factors and negligible loadings on the other factor. Table 2 (in particular 2a) shows the items' factor loadings after the rotation. The loadings in the two factors were substantially identical to those emerged in the approach-avoidance differentiation (Stark et al., 1991) and were thus accordingly labelled in the same way. The two factors resulted correlated each other, *r* =.38.

To sum up, the exploratory analysis suggested the two-factor solution for the Italian short form version of the NCS with a distinction between the approach and the avoidance of effortful cognitive activities. CFA was conducted on the second split sample (Random Sample 2) to test the two-factor structure obtained with EFA. We used the maximum likelihood estimation method. The examined model was a two-related factor model in which the items were predicted to load onto the two factors derived from the EFA. A model can be said to have a good fit when the chi-square test is non-significant. However, given that for models with many cases, the chi-square is almost always statistically significant, other model fit indices are considered. Specifically, a model is considered to have acceptable fit when CFI and NNFI are higher than .90 and the SRMR and RMSEA values are smaller than .08 (smaller than .05 for excellent fit; Hu & Bentler, 1999). Modification indices were also inspected to assess the extent to which the hypothesized model was appropriately described. Correlated errors are specified when the items share a part of the variance.

CFA showed that the uni-dimensional model had bad fit indices: RMSEA = .113, 90% CI [.10;.12], CFI = .67; NNFI = .62; SRMR = .10. On the contrary, the bi-dimensional solution showed good fit, RMSEA = .058: 90% CI [.04;.07], CFI = .91; NNFI = .90; SRMR = .06. The modification indices analysis suggested to add covariance between the errors of item 1 and item 2 and the errors of item 2 and item 11 (freeing up errors covariances was allowed because they are part of the same latent variable). The covariance between the errors of item 1 and item 2 could reflect an approach to situation require long and complex reasoning. The covariance between the errors of item 2 and item 11 could reflect the pleasantness towards situation requiring reasoning and new solution to problems. In the final solution with these covariances, the fit indices for the bi-dimensional solution further improved (RMSEA = .051, 90% CI [.04;.06], CFI = .93; NNFI = .92; SRMR = .06) and demonstrated better fit compared to the unidimensional model [Chi-square difference = 182.06; *df* = 1; *p*<.001, Difference CFI >.001) that continued to show bad fit to data (RMSEA = .097, 90% CI [.09;.11], CFI = .76; NNFI = .71; SRMR = .10).

The AIC index confirmed that the bifactorial solution (AIC = -34.52) better fitted the data compared to the mono-factorial solution (AIC = 349.53). All factor loadings were statistically significant and ranged from .35-.79, with an average standardized factor loading of .57. Squared multiple correlations ranged from .12-.61, with an average SMC of .33 indicating that, on average, 33% of the variance in observed variables was accounted for by latent factors. CFA upheld that the factors were related with each other, *r* = .46.

To sum up, both the EFA and CFA supported a bi-dimensional solution for the Italian versions of the NCS, with a differentiation between the approach and the avoidance dimension of the cognition. Figure 1 depicts the bi-factorial solution of Italian NCS.

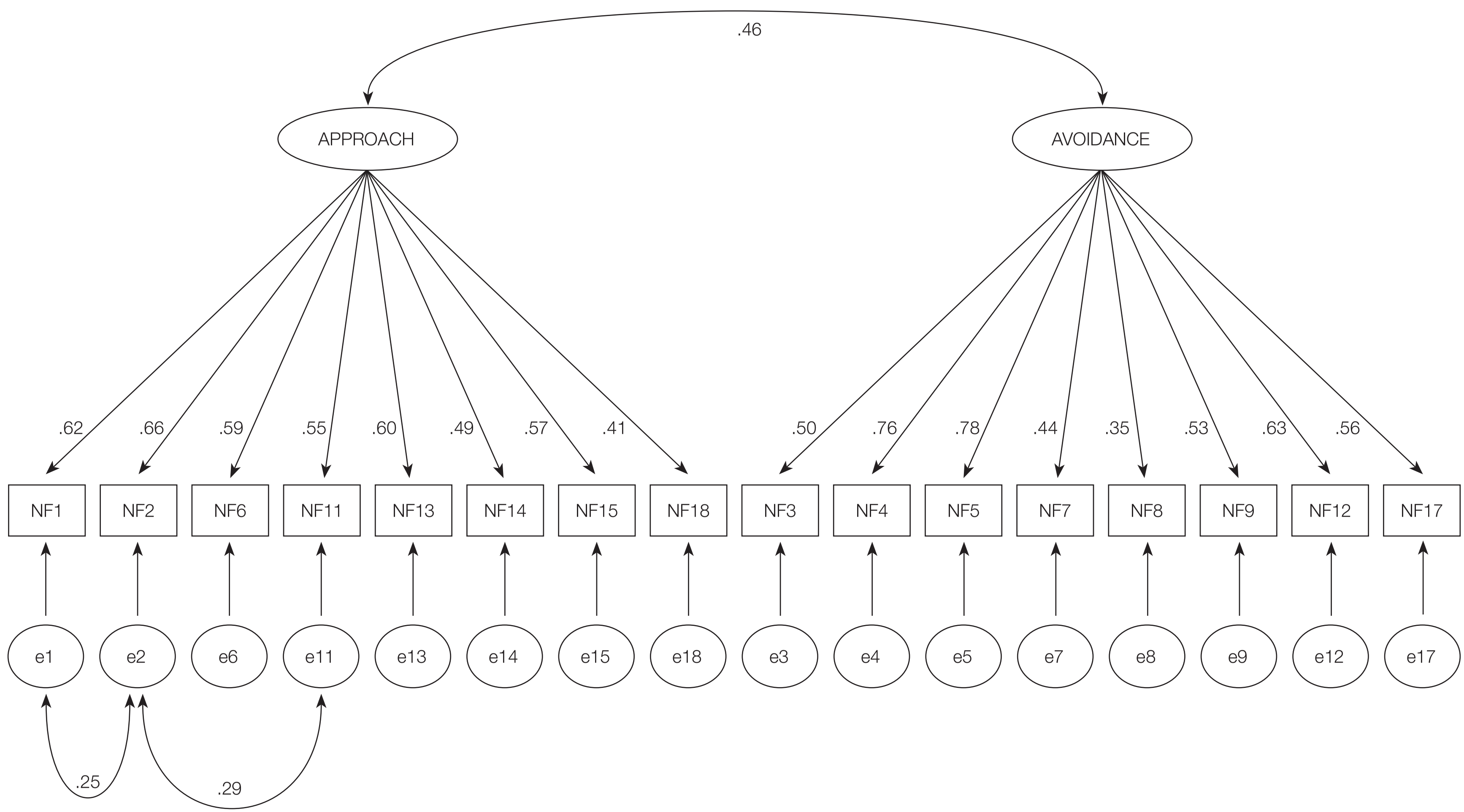
However, it sometimes happens that a genuinely unidimensional scale results as bi-dimensional due to the distortion stemming from the acquiescent response

Table 2 – Factor loadings (a), Percentage of variance explained by the factors (b), Mean items - Total correlations (c), Cronbach's alpha (d)

(a)* Factor loadings (Method of extraction: Principal Axis Factoring, Oblimin Rotation)	**Approach**	**Avoidance**
NC1	**.776**	–.103
NC2	**.662**	.122
NC6	**.575**	.117
NC11	**.493**	.101
NC13	**.531**	–.061
NC14	**.334**	.023
NC15	**.688**	–.122
NC18	**.320**	.092
NC3re	.060	**.375**
NC4re	.202	**.656**
NC5re	.048	**.723**
NC7re	.016	**.347**
NC8re	–.013	**.440**
NC9re	.141	**.509**
NC12re	–.057	**.619**
NC17re	–.130	**.635**
(b)* Percentage of variance explained	18%	22%
(c) Mean item - Total correlations**	.49	.49
(d) Cronbach's alpha**	.79	.77

Note. * Random sample 1 (N = 254), ** Total sample (N = 508). The factor loading in bold is significant.

Figure 1 - Confirmative Factorial Analysis of the Italian Need for Cognitive Scale with standardized regression weights



Note. All the standardized regression weights are significant at p<.001.

set (Marsch, 1989). Schriesheim and Hill (1981) reported that negatively phrased items are less reliable, especially when they are mixed with positively phrased ones: such poor reliability may increase overall measurement error in the total scores. Responses to positively worded items may be more straightforward than responses to negatively worded items because of differences in semantic complexity, which may result in greater measurement error among the negative phrased items (Hankins, 2008). Method effects are systematic variance that is attributable to the measurement method rather than to the constructs the measures represent (Podsakoff, MacKenzie, Lee & Podsakoff, 2003). To ascertain that the bi-factorial solution emerged from our data was not due to a method errors, we have compared the bi-factorial solution with an alternative model, by resorting to the correlated uniqueness approach (CCA; Marsch, 1989). The CAA allows the researcher to test the degree of distortion due to the response set and to correct for this distortion, by correlating the errors of the negative phrased items. Although this alternative model showed acceptable fit, except for the NNFI (RMSEA = .064, 90% CI [.05;.08], CFI = .92; NNFI = .87; SRMR = .05), the bi-factorial solution continued to fit better the data (Chi-square differences = 12.32, df = 24, p = .098, Difference CFI = .001). The AIC index confirmed that the bifactorial solution (AIC = −34.52) better fitted the data compared to the mono-factorial solution with correlated errors among the negatively worded items (AIC = 3.15).

Given the equivalence of the solution emerged from the EFA and CFA, we estimated the reliability and internal consistency of the NCS on the total sample of 508 participants.

Cronbach's alpha for the approach and the avoidance dimension were .79 and .77, respectively, thus confirming a good reliability. According to Nunnally and Bernstein (1994), an item is considered to have an acceptable level of internal consistency if its corrected item-total correlation is equal or greater than .33. All items satisfied this criterion, the mean of the item-total correlation was .49 for both the approach and the avoidance dimension (see Table 2, in particular 2c).

Convergent and divergent validity

Table 3 shows the correlations of the approach and avoidance dimension of NCS with other measures.

In line with our hypotheses, the approach dimension of NC correlated positively only with CCI, r (70) = .56, p<.001, and with the approach dimension of NFA, r (70) = .29, p = .01. The avoidance dimension of NC correlated instead with the avoidance dimension of NFA, r (70) = .26, p = .03, and with NCC, r (70) = −.44, p<.001. As expected, Table 3 also shows that the approach dimension was not related neither with NCC nor with the avoidance dimension of NFA. On the other hand, the avoidance dimension of NC did not correlate neither with CCI nor with the approach dimension of NFA. Taken together, these findings confirmed the convergent and divergent validity of NCS. Further, none of the correlation coefficients was equal to or greater than .70, thus indicating that the NCS dit not overlap with other constructs associated with the cognition and the psychology of the attitudes.

Table 3 – Zero-order correlation coefficients between the NCS and measured constructs

Variables	Cognitive Consumer Innovativeness	Need for Affect (Approach dimension)	Need for Affect (Avoidance dimension)	Need for Cognitive Closure
Factor 1 (Approach)	.56***	.29**	−.13	−.08
Factor 2 (Avoidance)	.08	.14	.26*	−.44***

Note. * p<.05; ** p<.01; *** p<.001; N = 70.

DISCUSSION

The aim of the present study was to provide a scale for the NC in the Italian context and to test its structure and its validity. The results confirm the reliability and validity of the Italian version of the NCS. Both the exploratory factor analysis and the confirmatory factor analysis suggested a bi-factorial solution for the Italian version of the NCS, with a differentiation between the approach to cognitive effortful activities and the avoidance of situations requiring a lot of thinking. Both the approach (α = .78) and the avoidance dimensions (α = .77) showed good internal consistency.

A separate examination of cognition approach and cognition avoidance is a valuable goal because these motivations might have distinct correlates, as confirmed from convergent and divergent validity. In fact, results showed that only the approach dimension is related to the cognitive desire to acquire new stimulating objects, whereas only the avoidance dimension is related to a desire to arrive fast at a solution, by avoiding uncertainty. Furthermore, the approach dimensions of NFA and NC were correlated with each other. Similarly, the avoidance dimensions of the two scale were related with each other, supporting the differentiation between the approach and avoidance in psychology research. The differentiation between the approach and the avoidance dimensions of NC could also differently predict other outcomes and future studies could explore these relationships.

The NCS may turn out a useful tool in both research and educational areas. For instance, in the research field, NCS could be used by scholars interested in the persuasion, given the extended literature showing that people who like reflection are more persuaded by a message which describes the details of the product, whereas people who avoid reflection are more persuaded by a message which does not require longer information processing. NCS could be used also in the social perception field, recent research suggests, in fact, that people with high level of NC more strongly appreciate competent people compared to incompetent people (Aquino, Haddock, Maio, Wolf & Alparone, 2016). Furthermore, NCS could be used in studies about the motivations underlying the use of technologies, given that people who like reflection usually enjoy stimulating technologies (Amichai-Hamburger, Kaynar & Fine, 2007). In the educational field, NCS could be used to have an indication about the teaching strategies, given that an efficient teacher should stimulate the reflection and thinking in the learners.

However, some limitations of this research need to be taken into account when interpreting its findings. First, the sample mainly consisted of young students, and this suggests caution regarding the generalizability of results. This problem does not affect the psychometric properties of the scale, but rather the demographic differences in the scores. Another limitation of this research is the limited number of participants used for the convergent and divergent validity. Given the low number of participants, we have tested the construct validity by performing a correlation approach rather than a SEM approach. However, the aim of the present research was the exploration of the factorial structure of NCS in the Italian context, thus future studies could purposely explore the convergent and divergent validity of the scale with a more adequate sample.

Overall, we provide evidence for the good psychometric properties of the NCS, a useful instrument for researchers and practitioners in several domains of the psychological field.

Acknowledgments
We thank Rossella Citro for her contribution in the data collection.

References

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

APPEL, M., GNAMBS, T. & MAIO, G.R. (2012). A short measure of the need for affect. *Journal of Personality Assessment, 94*, 418-426. https://doi.org/10.1080/00223891.2012.666921

AQUINO, A., HADDOCK, G., MAIO, G.R., WOLF, L.J. & ALPARONE, F.R. (2016). The role of affective and cognitive individual differences in social perception. *Personality and Social Psychology Bulletin*, *42*, 798-810. https://doi.org/10.1177/0146167216643936

BOLLEN, K.A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

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

CACIOPPO, J.T. & PETTY, R.E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*, 116-131. https://doi.org/10.1037/0022-3514.42.1.116

CACIOPPO, J.T., PETTY, R.E., FEINSTEIN, J.A. & JARVIS, W.B.G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, *119*, 197-253. https://doi.org/10.1037//0033-2909.119.2.197

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

CARICATI, L. & RAIMONDI, M. (2015). The Motivated Consumer Innovativeness Scale: Initial Italian validation. *TPM: Testing, Psychometrics, Methodology in Applied Psychology*, *22*, 363-383. https://dx.doi.org/10.4473/TPM22.3.4

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

FORSTERLEE, R. & HO, R. (1999). An examination of the short form of the need for cognition scale applied in an Australian sample. *Educational and Psychological Measurement*, *59*, 471-480. https://doi.org/10.1177/00131649921969983

FURLONG, P.R. (1993). Personal factors influencing informal reasoning of economic issues and the effect of specific instructions. *Journal of Educational Psychology*, *85*, 171-181. https://doi.org/10.1037/0022-0663.85.1.171

HANKINS, M. (2008). The factor structure of the twelve item General Health Questionnaire (GHQ-12): The result of negative phrasing? *Clinical Practice and Epidemiology in Mental Health*, *4*, 10. https://doi.org/10.1186/1745-0179-4-10

HORN, J.L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179-185. https://doi.org/10.1007/BF02289447

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

KLINE, R.B. (1998). *Principles and practice of structural equation modeling*. New York, NY: Guilford.

KRUGLANSKI, A.W. (1990). Motivations for judging and knowing: Implications for social attributions. In E.T. Higgins & R.M. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior*. New York, NY: Guilford.

KRUGLANSKI, A.W. & THOMPSON, E.P. (1999). Persuasion by a single route: A view from the unimodel. *Psychological Inquiry*, *10*, 83-109. https://doi.org/10.1207/s15327965pl100201

LEONE, L. & PRESAGHI, F. (2012). Validity of the Need for Affect Scales: Factorial structure, invariance and validity in the Italian context. *TPM: Testing, Psychometrics, Methodology in Applied Psychology, 14*, 1-18.

MAIO, G.R. & ESSES, V.M. (2001). The need for affect: Individual differences in the motivation to approach or avoid emotions. *Journal of personality*, *69*, 583-614. https://doi.org/10.1111/1467-6494.694156

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

MARSCH, H.W. (1989). Confirmatory factor analyses of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement, 13*, 335-361. https://doi.org/10.1177/014662168901300402

NUNNALLY, J.C. & BERNTSTEIN, I.H. (1994). *Psychometric theory* (*3rd ed.*). New York, NY: McGrawHill.

PETTY, R.E. & CACIOPPO, J.T. (1986). The elaboration likelihood model of persuasion. *Advances in experimental social psychology*, *19*, 123-205. https://doi.org/10.1007/978-1-4612-4964-1_1

PIERRO, A. & KRUGLANSKI, A.W. (2005). *Revised Need for Cognitive Closure Scale* (*NfCS*). Unpublished manuscript, Sapienza University of Rome, Italy.

PODSAKOFF, P.M., MacKENZIE, S.B., LEE, J.Y. & PODSAKOFF, N.P.

(2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*, 879-903. https://doi.org/10.1037/0021-9010.88.5.879

SCHERMELLEH-ENGEL, K., MOOSBRUGGER, H. & MULLER, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*, 23-74. https://doi.org/10.3389/fpsyg.2014.00181

SCHRIESHEIM, C.A. & HILL, H.D. (1981). Controlling acquiescence response bias by item reversals: The effect on questionnaire validity. *Educational and Psychological Measurement*, *41*, 1101-1114. https://doi.org/10.1177/001316448104100420

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

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

VANDECASTEELE, B. & GEUENS, M. (2010). Motivated consumer innovativeness: Concept and measurement. *International Journal of Research in Marketing*, *27*, 308-318. https://doi.org/10.1016/j.ijresmar.2010.08.004

WANG, V.C.X. (Ed.) (2015). *Handbook of research on scholarly publishing and research methods: Advances in knowledge acquisition, transfer, and management.* IGI Global.

ZHANG, X., NOOR, R. & SAVALEI, V. (2016). Examining the effect of reverse worded items on the factor structure of the need for cognition scale. *PloS ONE, 11*, e0157795. https://doi.org/10.1371/journal.pone.0157795

DOI: 10.26387/bpa.283.3

Clinical characteristics of the subtypes of trichotillomania: The Italian Milwaukee Inventory for the Subtypes of Trichotillomania – Adult Version (MIST-A)

Andrea Pozza[1], Douglas W. Woods[2], Davide Dèttore[1]

[1] Department of Health Sciences, University of Florence
[2] Department of Psychology, Marquette University, Milwaukee

ABSTRACT. Il *Milwaukee Inventory for the Subtypes of Trichotillomania – Adult Version* (MIST-A) indaga due sottotipi di tricotillomania. Il primo, definito Automatico, è lo strappamento di peli o capelli messo in atto in modo inconsapevole nel corso di attività sedentarie, quali la lettura. Il secondo, definito Focalizzato, viene messo in atto come strategia di regolazione emotiva. Il presente studio ha indagato la struttura fattoriale della versione italiana, la consistenza interna e validità concorrente/divergente con misure di dermatillomania, sensibilità all'ansia, esperienze dissociative, difficoltà di regolazione emotiva, evitamento esperienziale su 1142 adulti della popolazione generale (età media = 38,41, 60% femmine). La versione italiana del MIST-A dimostra solide proprietà psicometriche: il sottotipo Focalizzato sembra associato a sensibilità all'ansia ed evitamento esperienziale in misura maggiore di quello Automatico.

SUMMARY. Trichotillomania (TTM), is characterized by recurrent pulling out of hair. Current evidence suggests that it has different subtypes with distinct characteristics. Automatic pulling occurs out of awareness and includes situations where pulling is engaged in sedentary activities. Focused pulling occurs in response to negative emotions. The Milwaukee Inventory for the Subtypes of Trichotillomania – Adult Version (MIST-A) measures the TTM subtypes. In Italy, TTM is still under-recognized by clinicians and researchers. The current study investigated the factor structure, the reliability of the Italian MIST-A and its concurrent/divergent validity with measures of skin picking, anxiety sensitivity, dissociative experiences, difficulties in emotion regulation, psychological inflexibility and experiential avoidance in Italian community individuals. A large group of 1142 adults from the general population (Mean age = 38,41, SD = 14.67, females 60%) completed the MIST-A. Theoretical explanations, implications for practice and research are discussed.

Keywords: *Air pulling, Body-focused behaviours, Assessment, Emotion regulation, Factor structure*

INTRODUCTION

Trichotillomania (TTM), also referred to as hair pulling, is a psychiatric condition characterized by recurrent pulling out of hair, resulting in noticeable hair loss (American Psychiatric Association, 2012). Individuals typically experience an increasing sense of tension prior to or when attempting to resist pulling and relief or pleasure when pulling. This repetitive behaviour results in clinically significant distress or impairment and is not better accounted for by another mental health or medical condition (American Psychiatric Association, 2012). Hair pullers typically tend to report increased levels of shame, self-blame, and frustration from pre- to post-pulling, and lower levels of calmness after hair pulling episodes, while experiencing higher relief across the pulling cycle (Bottesi, Cerea, Ouimet, Sica & Ghisi, 2016).

Current evidence and various models of TTM suggest that the disorder may have different subtypes with distinct phenomenological and functional characteristics (e.g., Diefenbach, Mouton-Odum & Stanley, 2002). In a group of 60 adults diagnosed with TTM, Christenson, Mackenzie e Mitchell (1991) reported that 5% of participants endorsed hair pulling completely out of awareness, 15% reported pulling in which the focus of attention was directly on hair pulling, but the majority of participants (80%) reported pulling that ranged from complete to incomplete awareness of the behaviour. This research led to the identification of two pulling subtypes referred to as *automatic* and *focused* pulling (Christenson & Mackenzie, 1994). Automatic pulling is characterized by pulling episodes that occur primarily out of an individual's awareness and may include situations in which he/she pulls hair while engaging in sedentary activities (e.g., watching television, or reading a book), but he/she is unaware of pulling until after the pulling episode is complete. Focused pulling is characterized by pulling with an almost compulsive quality and includes situations in which the individual pulls in response to negative cognitive emotional states (e.g., anxiety, sadness, anger or boredom, an intense thought or urge, or in an attempt to establish symmetry). Research suggested that focused pulling may represent an attempt to decrease levels of unpleasant private experiences (Woods, Wetterneck & Flessner, 2006).

This conceptualization of the TTM phenomenology led to the suggestion that different clinical presentations may warrant different treatment strategies (Franklin, Tolin & Diefenbach, 2006; Woods et al., 2006). Accordingly, assessment instruments that evaluate the severity of different pulling subtypes may enhance treatment tailoring and the optimization of clinical care.

The *Milwaukee Inventory for Subtypes of Trichotillomania – Adult Version* (*MIST-A*; Flessner et al., 2008) is an instrument designed to assess subtypes of hair pulling. Overall, it is composed by 15 items, that measure focused (10 items) and automatic (5 items) pulling. In the original validation study, exploratory and confirmatory factor analyses provided evidence for a structure including the two uncorrelated factors. Both the focused pulling (Cronbach's alpha = .77) and automatic pulling (Cronbach's alpha = .73) scales demonstrated acceptable internal consistency (Flessner et al., 2008).

Despite this preliminary evidence in support of the MIST-A's psychometric properties, more recent research suggests that the original two-factor structure may not optimally capture TTM phenomenology. In a replication sample of 193 clinically characterized hair pullers, Keuthen and colleagues (2015) evaluated the MIST-A using exploratory factor analysis. Results suggested a two-factor solution composed by 13 items overall, divided into an 8-item *Intention* scale and a 5-item *Emotion* scale. More recently, the notions of automatic and focused pulling were challenged and new dimensions of TTM were proposed to study the factor structure of the MIST-A. In a treatment-seeking sample with TTM, Alexander, Houghton, Bauer, Lench & Woods (2018) reported a different two-factor solution for the MIST-A. The first factor, defined as *Awareness of pulling*, consisted of 5 items that measured the degree to which pulling is done with awareness. The second factor consisted of 8 items and was defined as an *Internal-regulated pulling* factor that measured the degree to which pulling is done to regulate internal stimuli (e.g., emotions, cognitions and urges). A limitation of the current studies was that two out of three studies used only exploratory factor analysis to test the structure. A systematic overview of the previous studies on the MIST-A is provided in Table 1.

Despite multiple studies on the MIST-A, several issues require further study. First, there is concern about low internal consistency. If one considers the Nunnally and Bernstein criteria (1994), in which Cronbach's alpha values higher than .80 and .90 suggest good and excellent reliability, then the MIST-A internal consistency is only acceptable, but not good to excellent (see Table 1). Second, the available evidence about the relations between each TTM

Table 1 – Comparison of the factor structures and Cronbach's alphas of the MIST-A in previous studies, data analysis, populations, and concurrent/divergent validity measures

Alexander et al. (2016)					Flessner et al. (2008)					Keuthen et al. (2015)				
Items, factor structure, and Cronbach's alphas		**Data analysis strategy**	**Study population**	**Concurrent/ divergent validity measures**	**Items, factor structure, and Cronbach's alphas**		**Data analysis strategy**	**Study population**	**Concurrent/ divergent validity measures**	**Items, factor structure, and Cronbach's alpha**		**Data analysis strategy**	**Study population**	**Concurrent/ divergent validity measures**
Awareness of pulling	Internal-regulated pulling	Exploratory Factor Analysis	Clinical sample with TTM ($n = 91$)	MGH-HS NIMHT-SSS BAI BDI-II	Focused pulling	Automatic pulling	Exploratory Factor Analysis ($n = 848$) + Confirmatory Factor Analysis ($n = 849$)	Clinical sample with TTM	TIS DASS-21	Intention pulling	Emotion	Exploratory Factor Analysis	Clinical sample with TTM ($n = 193$)	NIMHT-SSS TIS DASS-21
4	1				4	1				1	8			
8	2				5	2				2	10			
9	5®				6	3				4®	11			
10	7				8	7				5®	13			
11	12				9	12				6®	14			
13					10					7	15			
14					11					9®				
15					13					12				
					14									
					15									
Alpha .74	Alpha .80				Alpha .77	Alpha .73				Alpha .78	Alpha .75			

Legenda. ® = reverse item. BAI = Beck Anxiety Inventory; BDI-II = Beck Depression Inventory-II; DASS-21 = Depression Anxiety Stress Scales-21; MGH-HS = Massachusetts General Hospital - Hairpulling Scale; NIMHT-SSS = National Institute of Mental Health Trichotillomania Severity; TTM = Trichotillomania; TIS = Trichotillomania Impact Survey.

subtype and other clinical variables remains inconclusive. Several relationships may be worthy of further exploration, including the relationships between pulling style, comorbid body-focused repetitive behaviours, experiential avoidance, emotional dysregulation, dissociation, and anxiety sensitivity. Although co-occurrence of hair pulling and skin picking behaviours is quite frequent (Snorrason, Belleau & Woods, 2012), no study examined this relationship using measures of subtypes. Only one study investigated the relationship between the MIST-A subscales and psychological inflexibility/experiential avoidance (Alexander et al., 2018), which showed a moderate association between psychological inflexibility/experiential avoidance and MIST-A Internal-Regulated pulling, but no relation with MIST-A Awareness of pulling (Alexander et al., 2018). Since hair pulling, particularly the focused subtype, is hypothesized to be done in response to negative emotions (Woods et al., 2006), further evidence about this relationship is required. In addition, although emotion regulation deficits are hypothesized to be associated with TTM, particularly the focused subtype (Woods et al., 2006) and similarly also with focused skin picking (Pozza, Giaquinta & Dèttore, 2017), construct validity studies on the MIST-A did not investigate the relation between the two subtypes and emotion dysregulation. In addition, no study investigated concurrent validity of the MIST-A with dissociative experiences, despite the fact that nearly 20% of adults with TTM experience significant dissociative experiences (Carlson & Putnam, 1993), which may relate to pulling without awareness. Moreover, dissociative experiences could be related also to focused pulling since it could be used by the individual as a strategy of emotional regulation or control. Another clinical construct in need of further investigation is anxiety sensitivity, a cognitive dimension consisting of physical concerns (e.g., the belief that normal body sensations, such as an increase in heartbeat, lead to death), social concerns (e.g., the belief that publicly observable anxiety reactions will elicit social refusal), and cognitive concerns (e.g., the belief that cognitive difficulties lead to insanity). Anxiety sensitivity is believed to be central to the development of anxiety disorders (Taylor et al., 2007), and given the high comorbidity between TTM and anxiety (Flessner et al., 2008), one might expect correlations to emerge between this factor and focused pulling. A systematic comparison of the measures used to investigate concurrent and divergent validity in previous studies is provided in Table 1.

Finally, it seems important to look at how focused and automatic pulling may or may not occur across different cultures, since the questionnaire has not been validated in other languages than English. In the Italian context, hair pulling is still an under-recognized condition by clinicians and a very small number of studies was conducted. Recently, in the first epidemiologic contribution Ghisi, Bottesi, Sica, Ouimet & Sanavio (2013) reported prevalence rates ranging from about 2.1% to 16.5%, depending on the stringency of the TTM criteria used. However, to date no instrument is available with well-established psychometric properties to assess the subtypes of TTM during clinical practice or for research purposes. Starting from these considerations, the aims of the current study were to provide further evidence about the clinical characteristics of TTM subtypes, specifically:

1. investigating the psychometric properties, particularly the factor structure and the reliability of the MIST-A, in the Italian community (Study 1);
2. examining its concurrent and divergent validity with measures of skin picking, anxiety sensitivity, anxiety, dissociative experiences, difficulties in emotion regulation, psychological inflexibility and experiential avoidance (Study 2).

STUDY 1: FACTOR STRUCTURE AND RELIABILITY OF THE MIST-A

Participants

A large group of 1142 adults were recruited from the general population [Mean age (years) = 38,41, *SD* = 14.67, *range* = 18-75, percentage of females = 60%]. Data were collected from October 2012 to July 2017. Through convenience sampling, participants were recruited in a variety of public settings in several cities located in the Northern, Mid or Southern Italy. Psychologists approached participants in public settings, including high schools, universities, railway stations, libraries, malls, sports or volunteering associations. When approached, each participant was provided with a brief overview of the study, provided a description of hair pulling behaviours, and specifically asked whether they reported doing this behaviour, they were invited to participate. The fact that pulling behaviours tend to occur also in the general population was highlighted. If interested, each participant was taken aside to complete the questionnaires individually. In accordance with the *Ethical Principles of Psychologists and Code of Conduct* (American Psychological Association, 2012), all the participants who were recruited, provided written informed consent to participate after having received a detailed description of the study aims. Participants' identities remained anonymous and participation was entirely volunteer and uncompensated. Contact information of the study coordinator (AP) was provided if participants had further questions or concerns regarding their participation. Participants were considered eligible for the study if they stated that they engaged in hair pulling behaviours to some degree and if they provided written informed consent to participate. An overview of demographics of the participants in study 1 is provided in Table 2.

Measures: MIST-A

Participants completed the Italian translation of the MIST-A (Flessner et al., 2008). The measure, consisting of 15 items on a 10-point Likert-type response format (Not true of any of my hair pulling = 0, True for all of my hair pulling = 9), showed acceptable internal consistency (Cronbach's alpha = .77 for the *Focused* subscale, alpha = .73 for the *Automatic* one). Higher scores indicate more intense hair pulling behaviours. The process of translation in Italian followed a protocol according to international standards (Behling & Law, 2000). This process includes a translation from English into Italian, and a subsequent independent translation from Italian into English. The first translation was conducted independently by two native Italian clinical psychologists with excellent knowledge of English and double-checked by an Italian professional translator. Later, this version was translated back in English by a bilingual professional translator, who was unfamiliar with the original items. A final comparison between the latter translation and the original English version conducted to the generation of the Italian version of the scale. This Italian pilot version of the MIST-A was administered to ten Italian individuals in the community, and interviews were conducted by a psychologist in order to verify the semantic equivalence, the comprehensibility and content validity. Since this version was found to be valid in terms of comprehensibility, it was used for the current study.

Statistical analyses

In order to test the factor structure, a confirmatory factor analysis (CFA) was carried out using structural equations modelling on the whole community group (n = 1142). The distributional properties of the items were examined by conducting the Kolmogorov-Smirnov test and the inspection of the ratio between kurtosis and skewness and their standard errors.

First, a model with two uncorrelated factors was tested, as showed in Flessner and colleagues (2008). A second model including two correlated factors and a third model with a higher-order factor and two lower-order factors were also tested. Finally, a two-factor model was also tested without item 3 ("I am in an almost 'trance-like' state when I pull my hair") or item 11 ("I have a strange sensation just before I pull my hair"). This model was tested following the results reported in Keuthen and colleagues (2015), where items 3 and 11 were removed. Of added note, item 3 had subthreshold loadings values on both the factors (the *Intention* and the *Emotion* factors), and alpha values for the *Emotion* scale increased to .78 without item 3.

In order to evaluate the model's goodness of fit to the data, the following indices were used (Hu & Bentler, 1999): the Bentler-Bonett Normed Fit Index (NFI), the Bollen's Relative Fit Index (RFI), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI). For these indices, values between .95 and 1 represent a good fit, values between .90 and .95 acceptable fit. In addition, the Root Mean Square Error of Approximation (RMSEA) was used as index of fit. For the RMSEA, values less than .08 represent acceptable fit, and values less than .05 represent good fit. Reliability was evaluated as internal consistency using Cronbach's alpha coefficients calculated on the total sample. Reliability coefficients were evaluated according to Nunnally and Bernstein (1994) (alpha>.70 = acceptable; alpha>.80 = good; alpha>.90 = excellent). Further, reliability was verified as three-month-temporal stability using Pearson's correlation test-retest coefficients in a subsample ($n = 97$) of the total sample.

In order to collect three-month temporal stability data, all the participants were asked to give their e-mail addresses, so investigators could contact them. A subset ($n = 30$) agreed to participate in a second administration when contacted by study personnel. A psychologist reached each of the participants who accepted to complete the MIST-A in this second phase of the study. One-way ANOVA analyses have been performed, in order to compare scores between males

Table 2 – Demographics of study 1 (n = 1142), study 2 (n = 355)

	Study 1 group (*n* = 1142)	**Study 2 group (*n* = 355)**
	M (SD; range)/ n (%)	M (SD; range) n (%)
Age (years)	38.41 (14.67; 18-75)	40.83 (15.54; 18-75)
Female gender	685 (60)	202 (57)
Marital status		
Single	604 (53)	169 (47.6)
Married	374 (32.7)	159 (44.8)
Separated/Divorced	151 (13.3)	22 (6.2)
Widowed	13 (1)	5 (1.4)
School license		
Elementary school license	17 (1.5)	7 (2)
Middle school license	201 (17.6)	53 (15)
High school license	463 (40.5)	176 (49.6)
Degree	324 (28.4)	104 (29.3)
Post-graduate specialization	128 (11.2)	14 (4)
Ph.D.	9 (0.8)	1 (0.3)
Employment status		
Students	263 (22.8)	59 (16.6)
Employed	550 (48.2)	233 (65.6)
Unemployed	269 (25.6)	33 (9.3)
Retired	44 (3.9)	10 (2.8)

and females in the total scores and in the subscales scores. Reliability calculations were conducted with SPSS software version 21.00. Reliability was assessed as internal consistency and temporal stability. Internal consistency was calculated through Cronbach's alpha coefficients and assessed according to the criteria provided by Nunnally and Bernstein (1994) (alpha>.70 = acceptable; alpha>.80 = good; alpha>.90 = excellent). The CFA and internal consistency analyses were conducted through the Amos and the SPSS 21.00 version software.

Results

The scores on all the MIST-A items did not fit a normal distribution, since the ratio between skewness and kurtosis and the corresponding standard errors resulted out of the chosen range between -1 and +1. To correct for skewness, a logarithmic transformation to the scores on all the MIST-A items was applied. Results of CFA indicated that a model including two correlated factors had the best fit as compared with a model including two uncorrelated factors, a model with two uncorrelated factors without items 3 and 11 as in Keuthen et al. (2015), or a model with a higher-order and two lower-order factors. An overview of the fit indices across the models is presented in Table 3. The final items used for the Italian version of the MIST-A, their distribution and standardized coefficients in the two factors are presented in Table 4.

Internal consistency was excellent for the total scores of the MIST-A (Cronbach's alpha = .92) according to Nunnally and Bernstein (1994). Corrected item-total correlations were ranged from .10 and .85 and alpha values ranged between .90 and .92, when each of the items was deleted. Cronbach alpha values were .88 and .90 for the Focused and the *Automatic* subscales, respectively, suggesting good and excellent internal consistency (Nunnally & Bernstein, 1994). For the Focused subscale, alpha values ranged between .83 and .87, when each of the items was deleted; for the *Automatic* subscale, alpha values ranged between .86 and .89, when each of the items was deleted.

Temporal stability was good for the total MIST-A scores (Pearson's r = .85) and for scores on both the *Automatic* (Pearson's r = .86) and the *Focused* subscale (Pearson's r = .79), as strong values in the bivariate correlation coefficients were observed.

STUDY 2: CONCURRENT AND DIVERGENT VALIDITY OF THE MIST-A

Participants

A subgroup of 355 individuals from the total community group, who gave their consent to participate in a further assessment, completed also other self-report measures in addition to the MIST-A, in order to investigate its concurrent and divergent validity. All the participants in the subgroup stated that they engaged in hair pulling behaviours to some degree. This subgroup was created on the basis of the participant's consent to complete further questionnaires. An overview of socio-demographic characteristics of this subgroup is shown in Table 1.

Table 3 – Fit indices of tested models of the Italian MIST-A (n = 1142)

Tested models	χ^2	df	p-value	TLI	CFI	NFI	RFI	RMSEA
2 uncorrelated factors as in Flessner et al. (2008)	3878.92	90	.0001	.62	.68	.61	.61	.192
2 correlated factors as in Flessner et al. (2008)	2772.73	89	.0001	.73	.77	.77	.72	.163
2 correlated factors and modification indices as in Flessner et al. (2008)	552.75	78	.001	.95	.95	.96	.94	.075
2 uncorrelated factors without items 3 and 11 as in Keuthen et al. (2015)	3449.46	65	.0001	.59	.65	.65	.58	.214
2 correlated factors as in Alexander et al. (2016)	1737.28	64	.0001	.79	.82	.82	.78	.155
1 higher-order factor and 2 lower-order factors	3549.01	91	.0001	.66	.70	.70	.65	.182

Legenda. df = degree of freedom; TLI = Tucker-Lewis Indez; CFI = Comparative Fit Index; NFI = Bentler-Bonnett Normed Fit Index; RFI = Bollen's Relative Fit Index; RMSEA = Root Mean Square Error of Approximation.

Table 4 – Distribution and standardized coefficients of the items in the Italian MIST-A across the two factors

	Focused factor	Automatic factor
1. I pull my hair when I am concentrating on another activity.		.79
2. I pull my hair when I am thinking about something unrelated to hair pulling.		.80
3. I am in an almost "trance-like" state when I pull my hair.		.82
4. I have thoughts about wanting to pull my hair before I actually pull.	.50	
5. I use tweezers or some other device other than my fingers to pull my hair.	.22	
6. I pull my hair while I am looking in the mirror.	.27	
7. I am usually not aware of pulling my hair during a pulling episode.		.71
8. I pull my hair when I am anxious or upset.	.85	
9. I intentionally start pulling my hair.	.54	
10. I pull my hair when I am experiencing a negative emotion, such as stress, anger, frustration, or sadness.	.87	
11. I have a "strange" sensation just before I pull my hair.	.62	
12. I don't notice that I have pulled my hair until after it's happened.		.73
13. I pull my hair because of something that has happened to me during the day.	.82	
14. I pull my hair to get rid of an unpleasant urge, feeling, or thought.	.76	
15. I pull my hair to control how I feel.	.62	

Note. Scoring: the score on each of the two subscales is calculated by summing the raw scores on all the corresponding items.

Measures

The *Milwaukee Inventory for the Dimensions of Adult Skin Picking* (*MIDAS*; Walther, Flessner, Conelea & Woods, 2009) is a 12-item self-report measure that assesses pathological skin picking. Each item of the MIDAS is rated from 1 = not true for any of my behaviours of skin picking, to 5 = true for all my behaviours of skin picking. The MIDAS is the only instrument designed to evaluate subtypes of skin picking: a focused subtype, which typically concerns specific areas of the body and occurs in response to negative emotions (such as anger or anxiety), or bodily sensations, and an automatic subtype, which occurs without awareness during activities not related to the picking behaviours. The MIDAS items were modelled based on those of the MIST-A. The validation study (Walther et al., 2009) was conducted through an online survey on a sample of 92 participants, who reported repetitive body-focused behaviours, including skin picking and trichotillomania. The validation study of the Italian version of the measure suggested three factors, assessing a focused, automatic, and mixed subtype, respectively (Pozza, Mazzoni et al., 2016).

The *Beck Anxiety Inventory* (*BAI*; Beck, Epstein, Brown & Steer, 1988), a 21-item questionnaire, was used to assess anxious symptoms. Items are rated from 0 to 3 scores, which can range from 0 to 63 with higher scores indicating greater anxiety severity. The measure showed very good internal consistency (Beck et al., 1988). The Italian translation (Sica & Ghisi, 2007) showed excellent internal consistency for student and clinical samples.

The *Anxiety Sensitivity Index-3* (*ASI-3*; Taylor et al., 2007) is an 18-item self-report questionnaire on a 5-point Likert-type scale (0 = very little; 4 = very much). The ASI-3 is used to measure the three dimensions of anxiety sensitivity: Physical concerns (e.g., "When I feel pain in my chest, I worry that I'm going to have a heart attack"), Cognitive concerns (e.g., "When I cannot keep my mind on a task, I worry that I might be going crazy"), and Social concerns (e.g., "I worry that other people will notice my anxiety"). The measure showed to have good to excellent internal consistency in both clinical and non-clinical samples from different countries (Taylor et al., 2007). The Italian version (Pozza & Dèttore, 2015) showed good to excellent internal consistency in both non-clinical and clinical samples.

The *Dissociative Experiences Scale – II* (*DES-II*; Carlson & Putnam, 1993), a 28-item questionnaire on a 11-point scale, was used as a measure of dissociative experiences. Scores are calculated by dividing the total by 28, leaving a potential *range* from 0 to 100. Two subscales are calculated: *Compartmentalization and Detachment*. Higher scores on the two subscales indicate a higher degree of dissociative experiences.

The *Difficulties in Emotion Regulation Scale* (*DERS*; Gratz & Roemer, 2004) was used to assess self-reported emotion regulation difficulties. Six subscale scores can be computed from the 36 items, namely Non-acceptance of emotions (6 items; e.g., "When I'm upset, I feel guilty for feeling that way"), Difficulties engaging in goal-directed behaviour when distressed (5 items; e.g., "When I'm upset, I have difficulty concentrating"), Impulse control difficulties (6 items; e.g., "When I'm upset, I become out of control"), Lack of emotional awareness (6 items; e.g., "I pay attention to how I feel" [reversed]), Limited access to emotion regulation strategies (8 items; "When I'm upset, it takes me a long time to feel better") and Lack of emotional clarity (5 items; "I am confused about how I feel"). Participants rate each item on a scale from 1 (almost never, 0- 10%) to 5 (almost always, 91-100%). The authors describe good psychometric properties for all subscales, e.g., adequate to good internal consistencies (Cronbach's alpha values >.80). The Italian version showed acceptable to excellent internal consistency across all the subscales (Sighinolfi et al., 2010).

The AAQ-II (Bond et al., 2011) was used to assess psychological inflexibility and experiential avoidance. The AAQ-II is a revised version of the original AAQ (Hayes et al., 2004). The AAQ-II is a seven-item self-report measure that uses a 7-point Likert scale (1 = never true; 7 = always true). Sample items include "I am afraid of my feelings" and "Emotions cause problems in my life".

The AAQ-II exhibits a single-factor structure, good internal consistency and good test-retest reliability. The Italian version showed good internal consistency (Pennato, Berrocal, Bernini & Rivas, 2013).

Statistical analyses

To evaluate convergent validity, Pearson's bivariate correlation coefficients were calculated between scores on the MIST-A and scores on the MIDAS, BAI, ASI-3, AAQ-II, DES-II, and DERS. Values of correlation coefficients were interpreted as follows: $0<r<|.30|$ = weak, $|.30|<r<|.50|$ = moderate, $|.50|<r<\pm|.70|$ = strong, $r<\pm|.70|$ = very strong (Cohen, Cohen, West & Aiken, 1998). Power calculations were run for this analysis. For a medium effect size, 80% power, and significance set at the level described above, the required sample size for bivariate correlations was at least 82. To compare the magnitude of Pearson' correlation coefficients between scores on the MIST-A with scores on the measures used to assess concurrent and divergent validity, Fisher's z coefficients for dependent samples were calculated. The bivariate correlations were conducted with SPSS software version 21.00. Power calculations were performed using the GPower 3.1.7 software.

Results

Scores on the MIST-A Focused subscale strongly and positively correlated with scores on the MIST-A *Automatic* subscale. Pearson's bivariate correlations with Fisher's *z* coefficients between scores on the MIST-A and scores on the MIDAS are presented in Table 5.

Scores on the MIST-A *Focused* subscale weakly and positively correlated with scores on ASI-3 Physical concerns, DES-II subscales and AAQ-II and moderately with scores on ASI-3 Cognitive and Social concerns and BAI. Scores

Table 5 – Pearson's bivariate correlations (Fisher's z coefficients) between scores on the MIST-A and the MIDAS (n = 355)

	1.	2.	3.	4.	5.
1. MIST-A Focused	1	.71**	.40** (1.95*)	.43** (.82)	.44** (3.24**)
2. MIST-A Automatic			.47**	.40**	.32**
3. MIDAS Automatic				.72**	.51**
4. MIDAS Focused					.51**
5. MIDAS Mixed					1
Mean	8.42	5.32	7.53	4.89	5.70
SD	13.13	9.26	4.01	2.99	2.69

Legenda. MIDAS = Milwaukee Inventory for the Dimensions of Adult Skin picking.
Note. **p*<.05**, *p*<.001.

on the MIST-A *Automatic* subscale weakly and positively correlated with scores on all the ASI-3, BAI, DES-II and AAQ-II subscales. Significant differences between scores on the MIST-A *Focused* and *Automatic* subscales were found in the magnitudes of the correlations with scores on the ASI-3 Physical (Fisher's z = 2.49, p<.01) and Cognitive concerns (Fisher's z = 3.40, p<.01), BAI (Fisher's z = 2.38, p<.01) and AAQ-II (Fisher's z = 2.02, p<.05): scores on the MIST-A *Focused* subscale more strongly correlated with scores on these subscales than those on the MIST-A *Automatic*. Pearson's bivariate correlations with Fisher's z coefficients between scores on the MIST-A and scores on the ASI-3, BAI, DES-II and AAQ-II are presented in Table 6.

Scores on the MIST-A *Focused* and *Automatic* subscales weakly and positively correlated with scores on all the DERS subscales, except for those on the DERS Lack of emotional awareness, with which correlations were negative. No difference between scores on the MIST-A *Focused* and *Automatic* subscales was found in the magnitudes of the correlations with scores on the DERS subscales, as indicated by Fisher's z coefficients. Pearson's bivariate correlations with Fisher's z coefficients between scores on the MIST-A and scores on the DERS are presented in Table 7.

GENERAL DISCUSSION

Recent clinical models of TTM have conceptualized the disorder as a multidimensional condition composed of different subtypes. However, inconclusive and inconsistent evidence has been produced about its dimensionality. The current study expanded the present knowledge on the clinical characteristics of TTM subtypes investigating further the psychometric properties of the MIST-A in a large group of individuals recruited from the community, who stated that they engaged in hair pulling. In comparison with previous studies, a strength of the current one was the use of confirmatory factor analysis. The study investigated concurrent and divergent validity with unexplored clinical variables, such as measures of skin picking, anxiety sensitivity, dissociative experiences, and difficulties in emotion regulation. An original element of the study was the calculation of Fisher' s z coefficients that allowed comparing the differential magnitudes of the intercorrelations between the two TTM subtypes and the clinical variables.

Different from the initial validation study (Flessner et al., 2008), where the two subtypes were uncorrelated (Pearson's r = .01), in the current study a model in which the focused and

Table 6 – Pearson's bivariate correlations (Fisher's z coefficients) between scores on the MIST-A, ASI-3, BAI, DES-II and AAQ-II (n = 355)

	3.	4.	5.	6.	7.	8.	9.
1. MIST-A Focused	.19* (2.49**)	.37* (3.40**)	.35* (1.31)	.38* (2.38**)	.27* (1.27)	.25* (1.01)	.25* (2.02***)
2. MIST-A Automatic	.09	.24*	.30*	.29*	.22*	.21*	.17*
3. ASI-3 Physical concerns		.63*	.47*	.42*	.16*	.09	.26*
4. ASI-3 Cognitive concerns			.58*	.52*	.28*	.29*	.37*
5. ASI-3 Social concerns				.37*	.22*	.21*	.32*
6. BAI					.36*	.31*	.29*
7. DES-II Compartmentalization						.73*	.25*
8. DES-II Detachment							.24*
9. AAQ-II							1
Mean	5.32	3.61	7.03	11.06	318.44*	54.37	20.72
SD	4.61	4.01	4.65	8.82	240.62*	73.49	6.14

Legenda. ASI-3 = Anxiety Sensitivity Index-3; BAI = Beck Anxiety Inventory; DES-II = Dissociative Experiences Scale-II; AAQ-II = Acceptance and Action Questionnaire-II version.

Note. *p<.001, **p<.01, ***p<.05.

the automatic subtypes were strongly intercorrelated, yielded a better fit (Pearson's r = .71). The inclusion of covariances between residuals of some items was necessary to improve the model fit. When the covariances were introduced, values on all the fit indices were acceptable on the TLI, CFI, NFI, as they were equal or higher than 0.95. The RMSEA value also became acceptable. Thus, the factor structure reported in Flessner and colleagues (2008) was preferred also since it was supported by more robust methods, such as both exploratory and confirmatory factor analyses conducted on two large independent samples (n = 848 and n = 849, for exploratory and confirmatory factor analysis, respectively). The factor structures reported in Keuthen et al. (2015) and Alexander et al. (2018) were based only on exploratory analyses and were tested in relatively small samples: n = 193 for Keuthen et al. (2015), n = 91 for Alexander et al. (2018). However, it should be noted that a limitation of the current data was that the RMSEA value resulted lower than the threshold of .08, indicating acceptable fit, but it was higher than .06, that is the threshold for good fit. The current evidence about the intercorrelation between the two TTM subtypes was not consistent with the original theoretical model of the TTM subtypes proposed by Flessner and colleagues (2008), where the two subtypes were hypothesized being uncorrelated. From a clinical point of view, however, a model with two intercorrelated subtypes may be more consistent with clinical research and practice with individuals reporting body-focused repetitive behaviours (Arnold, Auchenbach & McElroy, 2001; Pozza, 2018). The subtypes of body-focused repetitive behaviours often present with common clinical characteristics related to personality and emotion regulation (Pozza, Giaquinta & Dèttore, 2016). Indeed, in clinical practice individuals with TTM typically

Table 7 – Pearson's bivariate correlations (Fisher's z coefficients) between scores on the MIST-A and the DERS (n = 355)

	3.	4.	5.	6.	7.	8.
1. MIST-A Focused	.12* (1.48)	.27** (.76)	.16** 1.24)	.27** (1.53)	.01 (.24)	−.20** (−1)
2. MIST-A Automatic	.06	.24**	.11*	.21**	.02	−.16**
3. DERS Non-acceptance of emotional responses		.55**	.50**	.47**	.35**	−.06
4. DERS Difficulties engaging in goal directed behaviour			.54**	.64**	.29**	−.15**
5. DERS Limited access to emotion regulation strategies				.61**	.44**	.06
6. DERS Impulse control difficulties					.29**	−.14**
7. DERS Lack of emotional clarity						.33**
8. DERS Lack of emotional awareness						1
Mean	11.10	12.04	16.53	11.01	10.35	6.71
SD	4.61	4.44	5.13	4.39	3.52	2.82

Legenda. DERS = Difficulties in Emotion Regulation Scale.
Note. *p<.001, **p<.01.

report both the subtypes when they get in contact with clinicians; alternatively, at the time of the clinical evaluation, they show a specific subtype while having suffered from the other subtype in the past, before seeking help from a clinician. Therefore, a model with two intercorrelated subtypes may confirm that the subtypes belong to a TTM syndrome and they are not just distinct symptoms. In addition, this model may have clinical implications and prognostic utility, since it may suggest that clinicians should be aware about the possibility that hair-pullers have the characteristics of both the subtypes despite apparently showing only one subtype or may develop also the other subtype in the future. In clinical practice, individuals with TTM frequently show the characteristics of a subtype for a certain period and the characteristics of the other subtype for another period.

The Italian MIST-A had excellent internal consistency for the total scale and the *Automatic* subscale. Internal consistency was good for the Focused subscale. These values were substantially higher than those reported in the initial validation study (Flessner et al., 2008), where the *Focused* and the *Automatic* subscale showed Cronbach's alpha values of .77 and .78, respectively.

Evidence of convergent and divergent validity supported that the two TTM subtypes were strongly correlated each other. In addition, both the TTM subtypes measured by the MIST-A were moderately associated with all the subtypes of skin picking assessed by the MIDAS. Significant differences between scores on the MIST-A *Focused* and *Automatic* subscales were found in the magnitudes of the correlations with scores on the MIDAS *Automatic* and *Mixed* subscales: scores on the MIST-A *Automatic* subscale correlated more strongly with scores on the MIDAS *Automatic* subscale than those on the MIST-A *Focused* subscale; scores on the MIST-A Focused correlated more strongly with scores on the MIDAS *Mixed* than those on the MIST-A *Automatic*. This is the first study investigating the intercorrelations between TTM subtypes and skin picking subtypes. The current findings suggested that subtypes of TTM and skin picking can be highly intercorrelated, have largely overlapping clinical characteristics and similarities in the clinical presentation in accordance with some recent reviews (Snorrason et al., 2012). Overall, these data showed that the co-occurrence of hair-pulling and skin picking is quite frequent. While the focused subtype of TTM was more strongly associated with focused and mixed skin picking than the automatic subtype of TTM; the latter was more strongly associated with the automatic subtype of skin picking than the first one. These results supported good concurrent and divergent validity of the MIST-A, as it demonstrated to be able to discriminate the two specific subtypes of body focused behaviours, regardless the type of body focused behaviours, TTM or skin picking.

The moderate and weak correlations of the focused and the automatic subtypes respectively with anxiety were consistent with the results reported in the initial validation study by Flessner and colleagues (2008). In addition, the focused subtype was associated more strongly with anxiety than the automatic one. This result could support that body focused behaviours are associated with negative emotions; this could be viewed as consistent with previous data, which suggested that focused pulling may represent an attempt to decrease levels of negative affect or regulate aversive feelings, particularly anxiety, resulting in a paradoxical increase of negative feelings (Woods et al., 2006). Consistently, Diefenbach and colleagues (2002) showed that anxiety and tension may serve as triggers and pulling behaviours as negative reinforcers. Conversely, extant research has failed to report a relationship between automatic pulling and negative affect (Diefenbach et al. 2002). A main limitation of the current study was the cross-sectional design, which prevented to draw reliable conclusions about causality.

Focused pulling was moderately and more strongly associated with experiential avoidance and psychological inflexibility than the automatic subtype, which was only weakly related to it. This result appeared consistent with the general observation, obtained also from other measures of TTM than the MIST-A, that focused hair pulling is more strongly associated with experiential avoidance and psychological inflexibility than the automatic one (Norberg, Wetterneck, Woods & Conelea, 2007; Shusterman, Feld, Baer & Keuthen, 2009). The relationship between focused hair pulling and experiential avoidance and psychological inflexibility appeared consistent with previous data, which indicated that those individuals who tended to engage in experiential avoidance also experienced greater frequency and intensity of urges, greater struggle with urges to pull, and increased distress associated with pulling, in comparison to those who tended to be more experientially accepting (Begotka, Woods, & Wetterneck, 2004). In addition, the paradoxical effect of experiencing more frequent and intense urges in relation to high experiential avoidance is consistent with the literature on thought suppression (Purdon & Clark, 2000), which has found that attempts to suppress unwanted thoughts often results in an increased frequency of those thoughts. The evidence that focused hair pulling was more strongly related to experiential avoidance than the automatic pulling one was also consistent with the data on the clinical characteristics of subtypes of skin picking (Walther et al., 2009).

On one hand, focused pulling was moderately correlated with Anxiety sensitivity cognitive and Social concerns, while it was weakly associated with Physical concerns; on the other hand, automatic pulling was weakly associated with all the Anxiety sensitivity dimensions. Focused pulling was more closely associated with Physical and Cognitive concerns than automatic pulling, but this difference did not emerge for Social concerns. The weak association between hair pulling and Social concerns appeared in contrast with previous research indicating a significant association between TTM and social anxiety (Flessner et al., 2008).

Both pulling subtypes were weakly associated with dissociative experiences and no difference emerged in the magnitude of the association between each subtype and dissociative compartmentalization and detachment. This evidence could question the notion that automatic pulling is engaged specifically without awareness and during trance-like states of alteration of consciousness.

Surprisingly, only weak associations were found between both the TTM subtypes and emotion dysregulation dimensions and no difference emerged in the magnitude of the associations between the two subtypes. This outcome appeared in contrast with previous literature reporting that some specific emotion dysregulation dimensions are significant predictors of body focused behaviours, including skin picking (Alexander et al., 2018; Pozza, Giaquinta et al., 2016)

Finally, some limitations should be pointed out. A first one was the lack of a clinical group with a diagnosis of TTM. Future research on the MIST-A should investigate its dimensionality and also concurrent validity in an Italian clinical group diagnosed with TTM. Through ROC analysis, future studies should clarify whether the tool is able to detect patients with primary hair pulling as compared with other kinds of patients. In addition, a measure of self-reported hair pulling severity was not used, since in the Italian context such a self-report measure has not been yet validated. Moreover, it could be interesting in the future to investigate the relationship between hair pulling behaviours and other emotional feelings than anxiety, such as anger, boredom, guilt, and shame. Another point which needs to be addressed is responsiveness, that is the capacity of the tool to measure changes in pulling behaviours after a specific psychotherapeutic intervention for TTM. In addition, the failure to evidence a relation between the two subtypes and emotion dysregulation could be in part due to the use of self-report measures. Additional research should use observational instruments or experimental tasks to assess more comprehensively this aspect. Another point which requires further investigation regards which clinical characteristics are more specifically associated with the automatic subtype than the focused one, since in the current study no difference in the magnitudes of the correlations was found favouring the first subtype. Thus, more knowledge about the specific features of the automatic pulling is needed.

CONCLUSIONS

Focused hair pulling seems to be a subtype which, different from the automatic one, is characterized by more intense anxiety, stronger anxiety sensitivity (particularly Cognitive and Physical concerns), higher experiential avoidance and psychological inflexibility. Interestingly and different from the literature, the focused pulling subtype and the automatic one seemed to be equally correlated with emotion regulation deficits and dissociative experiences. Therefore, the focused subtype could rely on emotion regulation deficits only related to avoidance of negative emotions, rather than on other kinds of deficits.

In conclusion, the current study provided further evidence about the clinical characteristics of TTM subtypes, supporting a two-factor structure of the MIST-A as a valid and reliable measure, which could be clinically useful to identify different types of clients with TTM, needing for specific tailored interventions.

References

ALEXANDER, J.R., HOUGHTON, D.C., BAUER, C.C., LENCH, H.C. & WOODS, D.W. (2018). Emotion regulation deficits in persons with body-focused repetitive behavior disorders. *Journal of Affective Disorders*, *227*, 463-470.

AMERICAN PSYCHOLOGICAL ASSOCIATION (2012). *Ethical principles of psychologist and code of conduct*. Washington, DC: American Psychological Association.

ARNOLD, L.M., AUCHENBACH, M.B. & McELROY, S.L. (2001). Psychogenic excoriation. Clinical features, proposed diagnostic criteria, epidemiology and approaches to treatment. *CNS Drugs*, *15*, 351-359.

BATEMAN, A. & FONAGY, P. (2006). *Mentalization-based treatment for borderline personality disorder: A practical guide*. Oxford: Oxford University Press.

BECK, A.T., EPSTEIN, N.B., BROWN, G. & STEER, R.A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology*, *56*, 893-897.

BEGOTKA, A.M., WOODS, D.W. & WETTERNECK, C.T. (2004). The relationship between experiential avoidance and the severity of trichotillomania in a nonreferred sample. *Journal of Behavior Therapy and Experimental Psychiatry*, *35*, 17-24.

BEHLING, O. & LAW, K.S. (Eds.) (2000). *Translating questionnaires and other research instruments: Problems and solutions*. Thousand Oaks: Sage.

BOTTESI, G., CEREA, S., OUIMET, A.J., SICA, C. & GHISI, M. (2016). Affective correlates of trichotillomania across the pulling cycle: Findings from an Italian sample of self-identified hair pullers. *Psychiatry Research*, *246*, 606-611.

CARLSON, B.E. & PUTNAM, F.W. (1993). An update on the Dissociative Experiences Scale. *Dissociation: Progress in the Dissociative Disorders*, *6*, 16-27.

CHRISTENSON, G.A.H. & MACKENZIE, T.B. (1994). Trichotillomania. In M. Hersen & R.T. Ammerman (Eds.), *Handbook of prescriptive treatments for adults*. New York, NY: Springer.

CHRISTENSON, G.A., MACKENZIE, T.B. & MITCHELL, J.E. (1991). Characteristics of 60 adult chronic hair pullers. *American Journal of Psychiatry*, *148*, 365-370.

COHEN, J. (1988). *Statistical power analysis for the behavioural sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

COHEN, J., COHEN, P., WEST, S.G. & AIKEN, L.S. (1998). *Applied multiple regression/correlation analysis for the behavioral sciences*. New York, NY: Routledge.

DIEFENBACH, G.J., MOUTON-ODUM, S. & STANLEY, M.A. (2002). Affective correlates of trichotillomania. *Behaviour Research and Therapy*, *40*, 1305-1315.

FIRST, M.B., GIBBON, M., SPITZER, R.L., WILLIAMS, J.B.W. & BENJAMIN, L.S. (1997). *Structured clinical interview for DSM-IV axis II personality disorders (SCID-II)*. Washington, DC: American Psychiatric Press.

FIRST, M.B., SPITZER, R.L., GIBBON, M. & WILLIAMS, J.B. (1995). *Structured Clinical Interview for DSM-IV Axis I disorders-Patient edition (SCID-I/P, version 2.0)*. New York, NY: Biometrics Research Department.

FLESSNER, C.A., WOODS, D.W., FRANKLIN, M.E., CASHIN, S.E., KEUTHEN, N.J., TRICHOTILLOMANIA LEARNING CENTER-SCIENTIFIC ADVISORY BOARD (TLC-SAB) (2007b). The Milwaukee Inventory for Subtypes of Trichotillomania – Adult version (MIST-A): Development of an instrument for the assessment of "focused" and 'automatic' hair pulling. *Journal of Psychopathology and Behavioral Assessment*, *30* (1), 20-30.

FRANKLIN, M.E., TOLIN, D.F. & DIEFENBACH, G.J. (2006). Trichotillomania. In E. Hollander & D.J. Stein (Eds.), *Clinical manual of impulse control disorders*. Washington, DC: American Psychiatric Press.

GHISI, M., BOTTESI, G., SICA, C., OUIMET, A.J. & SANAVIO, E. (2013). Prevalence, phenomenology and diagnostic criteria of hair-pulling in an Italian non-clinical sample: A preliminary study. *Journal of Obsessive-Compulsive and Related Disorders*, *2*, 22-29.

GRATZ, K.L. & ROEMER, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. *Journal of Psychopathology and Behavioral Assessment*, *26*, 41-54.

HAYES, S.C., STROSAHL, K., WILSON, K.G., BISSETT, R.T., PISTORELLO, J., TOARMINO, D. et al. (2004). Measuring experiential avoidance: A preliminary test of a working model. *The Psychological Record*, *54*, 553-578.

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

KEUTHEN, N.J., TUNG, E.S., REESE, H.E., RAIKES, J., LEE, L. & MANSUETO, C.S. (2015). Getting the word out: Cognitive-behavioral therapy for trichotillomania (hair-pulling disorder) and excoriation (skin-picking) disorder. *Annals of Clinical Psychiatry*, *27*, 10-15.

LINEHAN, M.M. (1993). *Cognitive-behavioral treatment of borderline personality disorder.* New York, NY: The Guilford Press.

NORBERG, M.M., WETTERNECK, C.T., WOODS, D.W. & CONELEA, C.A. (2007). Experiential avoidance as a mediator of relationships between cognitions and hair-pulling severity. *Behavior Modification*, *31*, 367-381.

NUNNALLY, J.C. & BERNSTEIN, I.H. (1994). *Psychological theory*. New York, NY: MacGraw-Hill.

PENNATO, T., BERROCAL, C., BERNINI, O. & RIVAS, T. (2013). Italian version of the Acceptance and Action Questionnaire-II (AAQ-II): Dimensionality, reliability, convergent and criterion validity. *Journal of Psychopathology and Behavioral Assessment*, *35*, 552-563.

POZZA, A. (2018). The role of alexithymia traits in compulsive skin picking behaviors. In R. Teixeira, P.P. Moorman & B. Bermond (Eds.), *Current developments in alexithymia: A cognitive and affective deficit*. New York, NY: Nova Publishers.

POZZA, A., GIAQUINTA, N. & DÈTTORE, D. (2016). Borderline, avoidant, sadistic personality traits and emotion dysregulation predict different pathological skin picking subtypes in a community sample. *Neuropsychiatric Disease and Treatment*, *12*, 1861.

POZZA, A., GIAQUINTA, N. & DÈTTORE, D. (2017). Impulsività e alessitimia nei sottotipi di Skin Picking: Uno studio su un ampio campione non-clinico. *Giornale Italiano di Psicologia*, *44*, 917-938.

POZZA, A., MAZZONI, G.P., PATASSINI, I., BERARDI, D., MOLINARO, F., PARENTI, G. & DÈTTORE, D. (2016). Studio sulle proprietà psicometriche della versione italiana del Milwaukee Inventory for the Dimensions of Adult Skin Picking (MIDAS) in campioni clinici e non clinici. *Psicoterapia Cognitiva e Comportamentale*, *22*, 35-63.

PURDON, C. & CLARK, D.A. (2000). White bears and other elusive intrusions. *Behavior Modification*, *24*, 425-453.

SHUSTERMAN, A., FELD, L., BAER, L. & KEUTHEN, N. (2009). Affective regulation in trichotillomania: Evidence from a large-scale internet survey. *Behaviour Research and Therapy*, *47*, 637-644.

SICA, C. & GHISI, M. (2007). The Italian versions of the Beck Anxiety Inventory and the Beck Depression Inventory-II: Psychometric properties and discriminant power. In M.A. Lange (Ed.), *New York leading-edge psychological tests and testing research*. New York, NY: NOVA Science Publishers.

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

SNORRASON, I., BELLEAU, E.L. & WOODS, D.W. (2012). How related are hair pulling disorder (trichotillomania) and skin picking disorder? A review of evidence for comorbidity, similarities and shared etiology. *Clinical Psychology Review*, *32*, 618-629.

TAYLOR, S., COX, B.J., HOLAWAY, R.M., STEWART, S.H., ENG, W., ARRINDELL, W.A., ZVOLENSKY, M.J., DEACON, B., ABRAMOWITZ, J.S., SANDIN, B., COLES, M., DALY, E.S. & BOUVARD, M. (2007). Robust dimensions of anxiety sensitivity: Development and initial validation of the Anxiety Sensitivity Index-3. *Psychological Assessment*, *19*, 176-188.

WALTHER, M.R., FLESSNER, C.A., CONELEA, C.A. & WOODS, D.W. (2009). The Milwaukee Inventory for the Dimensions of Adult Skin Picking (MIDAS): Initial development and psychometric properties. *Journal of Behavior Therapy and Experimental Psychiatry*, *40*, 127-135.

WETTERNECK, C.T., LEE, E.B., FLESSNER, C.A., LEONARD, R.C. & WOODS, D.W. (2016). Personality characteristics and experiential avoidance in trichotillomania: Results from an age and gender matched sample. *Journal of Obsessive-Compulsive and Related Disorders*, *8*, 64-69.

WOODS, D.W., WETTERNECK, C.T. & FLESSNER, C.A. (2006). A controlled evaluation of acceptance and commitment therapy plus habit reversal for trichotillomania. *Behaviour Research and Therapy*, *44*, 639-656.

DOI: 10.26387/bpa.283.4

Shared leadership: The Italian version of an overall cumulative scale

Salvatore Zappalà, Ferdinando Toscano, Simone Donati, Alessandro Malinconico, Ilaria Papola

Department of Psychology, University of Bologna

ABSTRACT. La leadership condivisa è un fenomeno in cui il ruolo e l'influenza propri della leadership sono distribuiti tra i membri di un gruppo. Questo studio presenta la versione italiana della scala di Leadership condivisa (LC) sviluppata da Muethel e Gehrlein (2009). La versione italiana della scala, composta da 7 item, è stata proposta a due campioni di studenti universitari italiani impegnati a svolgere attività di gruppo, per un totale di 444 rispondenti e 118 team. L'analisi fattoriale esplorativa e quella confermativa hanno confermato la struttura unidimensionale della scala. Inoltre, la scala è correlata significativamente con misure di processo (identificazione di gruppo e fiducia di gruppo) e di risultato (prestazione di gruppo e soddisfazione di gruppo), mostrando buona validità nomologica. Lo studio mostra che la versione italiana della scala ha buona validità interna e affidabilità, e costituisce un primo strumento per la misura della LC nel contesto italiano.

SUMMARY. Shared leadership is defined as an emergent team phenomenon where leadership roles and influence are distributed among team members. This study presents the Italian version of the Shared leadership (SL) scale developed by Muethel and Gehrlein (2009). The one-dimensional, seven-item scale was presented to two samples of Italian university students involved in team projects and team assignments, with a total of 444 respondents and 118 teams. Exploratory and confirmatory factor analyses confirmed the original one-factor model. The SL scale also shows good nomological validity because it is significantly related to team identification and team trust, as well as to team performance and team satisfaction. The study shows that the SL scale has good internal validity and reliability and can be considered a useful tool to measure SL in the Italian context.

Keywords: *Shared leadership, Team performance, Team processes, Scale validation*

INTRODUCTION

Leadership in organizations is no longer concentrated only in the hands of specific individuals who manage organizations and teams in a top-down way. Multidisciplinary teams, task forces, virtual teams, cross-functional, and inter-organizational teams require each member to make his/her own contribution by sharing knowledge and know-how to reach team objectives. In such teams, members tend to exercise, formally or informally, some forms of collaborative behaviors that yield horizontal and synergistic ways of performing leadership.

Shifting their focus from top-down, vertical influence processes to horizontal processes shared among team members, scholars introduced the concept of shared leadership. Shared leadership has been defined as "a dynamic, interactive influence process among individuals for which the objective is to lead one another to the achievement of group or organizational goals or both" (Pearce & Conger, 2003, p.1). What distinguishes shared leadership from traditional forms of leadership is that the process of influencing team members is no longer a skill or role attributed to a single person, the appointed or elected leader; instead, it is broadly distributed within the team and involves downward and upward influences as well as peer or lateral ones (Barnett & Weidenfeller, 2016; Pearce & Conger, 2003).

Other definitions of shared leadership have been provided to describe this phenomenon (Carson, Tesluk & Marrone, 2017; D'Innocenzo, Mathieu & Kukenberger, 2016). Most of them highlight that each team member, based on his or her skills and abilities, leads some team activity and follows the other team members when they are leading other activities. Three common aspects of the many existing definitions are that: 1) shared leadership involves lateral influence among peers; 2) it is an emergent team phenomenon; and 3) leadership roles and influences are distributed across team members (Zhu, Liao, Yam & Johnson, 2018).

Three recent meta-analyses concluded that shared leadership has a moderate, but significant, positive correlation with team performance (D'Innocenzo et al., 2016; Nicolaides et al., 2014) and team effectiveness (Wang, Waldman & Zhang, 2014), and these effects range between .21 and .35 in the three studies. In particular, shared leadership is more related to attitudinal (such as team satisfaction, team commitment, or team identification) and behavioral (such as team coordination) team outcomes than to subjective or objective team performance measures (Wang et al., 2014). A complex and reciprocal relationship with trust has also been observed. Small and Rentsch (2010) showed that trust is an antecedent of shared leadership, whereas Robert and You (2018) found that shared leadership promotes trust, which, in turn, has a direct effect on team satisfaction.

One important theoretical and methodological question has to do with the type of leadership shared among team members or, in other words, "what is shared in shared leadership". Wang et al. (2014) and Zhu et al. (2018) noted that some studies focus on the sharing of specific leadership styles (for instance, shared transformational, shared charismatic, or shared transactional leadership), whereas other studies focus on a "cumulative, overall" shared leadership, where team members assess how much their team relies on its members "for leadership". This latter case does not specify what type of leadership is enacted, but teammates have the "shared perception that, in general, members show leadership towards each other" (Wang et al., 2014, p. 184).

To assess and measure shared leadership, two major approaches have been observed: the aggregation approach and the social network approach (D'Innocenzo et al., 2016; Wang et al., 2014; Zhu et al., 2018). In the case of aggregation approach, when examining whether a specific leadership style (e.g. transformational leadership) was shared within a whole team, some scholars adapted well-established individual leadership questionnaires by changing the item referent from "my supervisor" to "my team members" and then aggregating members' ratings to the team level. For instance, Gockel and Werth (2010) used traditional questionnaires of aversive, directive, empowering, transactional, and transformational leadership that asked respondents to assess how much their teammates, or the team, shared one of these specific leadership styles. The same approach was used in other cases where new questionnaires were developed to assess specific functional leadership behaviors: respondents had to mentally aggregate the behavior of their different teammates for each item, and scholars derived an overall estimation of the team's shared leadership. For example, Muethel, Gehrlein & Hoegl (2012) developed a questionnaire to assess team members' proactive behaviors directed towards other teammates and towards their own area of responsibility. Grille & Kauffeld (2015) measured to what extent leadership behaviors such as assigning tasks, promoting team cohesion, or presenting inspiring ideas were shared among team members.

In the case of the social network approach, each team member assesses each of the other team members in terms of his/her respective leadership behavior. This is a more analytical approach because it sums up the influence of each member and provides a richer and more informative measure of shared leadership. Network density and network centralization indices have been used, although rarely in conjunction, to assess, respectively, how much leadership is being shared and the distribution pattern within the team or, in other words, if leadership is evenly distributed or concentrated in a few people (Carson et al., 2007; D'Innocenzo et al., 2016; Gockel & Werth, 2010).

Both the aggregate and network approaches have advantages and limitations (D'Innocenzo et al., 2016; Zhu et al., 2018). The limitations of the aggregate approach are based on the mental combinations that team members have to perform to provide a single representation of the team, as well as the adaptation of traditional vertical leadership constructs at the team level. The network approach more accurately reflects the complexities of shared leadership, but it is time consuming (because team members have to assess every other team member), and it is not efficient in assessing the many behaviors (such as planning, problem solving, suggesting ideas, or team support) that characterize leadership.

Considering the relevance of teams in modern organizations, the need to manage distributed, virtual, or even inter-organizational teams, and the relative lack of tools to measure shared leadership, this paper aims to provide the Italian community of scientists and practitioners with the Italian version of the shared leadership questionnaire proposed by Muethel and Gehrlein (2009). Following the aggregate approach, these two authors developed a scale to measure shared leadership behaviors. This is a seven-item, one-factor scale that the authors used to assess shared leadership in geographically disperse project teams working in software development companies. Five items were developed by the authors, based on their literature review, and two were adapted from two other different studies. The items assess an overall perception of shared leadership and address proactive initiatives undertaken by team members to anticipate other team members' information needs, facilitate task interdependencies, and encourage information flow, in order to revise and adapt team strategies to the environment.

Most previous studies have used traditional individual leadership scales aggregated at the team level, whereas the scale proposed here uses an overall cumulative approach. In addition, it focuses on proactive and goal-oriented behaviors of team members that facilitate task coordination and information flow. The scale is also suitable for research and practice because it is shorter than other scales that assess multiple leadership functions, such as the one used by Grille & Kauffeld (2015). Furthermore, it showed good internal consistency, with a Cronbach's alpha of .86.

This study aims to test the construct validity and reliability of the Italian version of the scale. To investigate construct validity, we tested factorial validity by running an exploratory analysis and a confirmatory factor analysis with two different samples of respondents. Then, we tested the nomological validity by examining whether shared leadership is positively correlated with specific team processes and team outcomes, as suggested in the literature (Nicolaides et al., 2014; Wang et al., 2014). Nomological validity is a component of construct validity, and Nunnally & Bernstein (1994) state that "any proof of the extent to which a measure defines a construct would have to come from determining how well the measure fits lawfully into a network of expected relationships" (p. 91). Accordingly, in the case of team processes, we expect shared leadership to be related to: a) affective team commitment; b) team identification; c) a propensity to trust team members, and d) the elaboration of team information. For the team outcomes, we expect shared leadership to be related to: a) team performance and b) team satisfaction.

METHODS

Participants

The present research was conducted in an Italian university with university students working in teams to carry out academic projects (such as research papers, group projects, or internship projects in community services). Specifically, two studies were conducted, in the 2015-16 and 2017-18 academic years, with two different samples of respondents. The first sample, attending master programs at the school of Psychology, was used for the exploratory factor analysis. It was composed of 224 participants, 31% males, with an average age of 23.9 years (*range* = 21-48; *SD* = 2.5), belonging to 62 different work teams (average team size = 5.09, *SD* = 1.1, *range* = 2-8).

The second sample, attending bachelor and master programs in different schools (Psychology, Sociology, Engineering, and Architecture), was used for the confirmatory factor analysis. It was composed of 220 participants, 30% males, with an average age of 21.9 years (*range* = 19-54; *SD* = 2.85), belonging to 56 different work teams (average team size = 5.01, *SD* = 2.16, *range* = 3-10).

Procedure

The Italian version of the shared leadership questionnaire was translated into Italian by two experts on the topic and back translated by three other people (a native English speaker and two non-Italians) into English. Professors who assigned team projects were contacted; after obtaining their approval, their students were contacted during lectures and invited to answer a paper and pencil questionnaire or its online version. Information about anonymity was given to all respondents; in order to maintain anonymity but aggregate data at the team level, participants were invited to agree on and share a fictitious name for their team to use when filling out the questionnaire.

Measures

The following three measures were used in both the first and second studies.

Shared leadership: Shared leadership (SL) was measured using the Italian version of the scale developed by Muethel and Gehrlein (2009). It consists of seven items rated on a 5-point Likert scale (from 1 "strongly disagree" to 5 "strongly agree"). Six items refer to anticipating team members' information needs and facilitating task interdependencies; the last item refers to how much the team relied on all the team members for leadership (the complete list of items is reported in Table 1).

Propensity to trust: the six-item subscale of the 21-item instrument developed by Costa & Anderson (2011) was used to measure trust within teams. The subscale refers to respondents' propensity to trust each other (item example: "Most people on this team do not hesitate to help a person in need").

Work group satisfaction: it was measured using Smith & Barclay's (1997) scale, composed of six items that assess the extent to which team members are satisfied with their teamwork. An example of an item is: "We are satisfied with each other's contribution to the team".

The following two measures were used only in the first study:

Team performance: the nine-item scale developed by Hoegl & Gemuenden (2001) was used to assess the perception of team effectiveness and efficiency. An example of an item is: "Considering the results, this team can be considered a success".

Team identification: it was measured using a version adapted to the team of the Organizational Identification scale by Mael and Ashforth (1992), validated in the Italian language by Bergami and Bagozzi (2000). It consists of six items, and an example of an item is: "The success of this team is my success".

The following two measures were used only in the second study:

Team affective commitment: we used the five items from the Italian version (Battistelli, Mariani & Bellò, 2006) of the affective commitment subscale of Meyer & Allen's (1991) Organizational Commitment questionnaire. Items were adapted to the team context (e.g., "This group has a great deal of personal meaning for me").

Team information elaboration: we used the four-item scale developed by Kearney, Gebert & Voelpel (2009) to assess the sharing of task-relevant information among team members. An example of an item is: "The members of this team complement each other by openly sharing their knowledge".

All the above-mentioned scales were assessed on a 5-point Likert scale ranging from 1 "strongly disagree" to 5 "strongly agree".

Data analysis

To assess the factorial validity of the Italian version of the SL scale, we performed an exploratory factor analysis (EFA) using Maximum Likelihood parameter estimates with SPSS 23, and then a confirmatory factor analysis (CFA) with Amos 23. Based on the literature (Bollen & Long, 1993), the model was assessed by using several goodness-of-fit criteria: the chi-square value (χ^2); the Root Mean Square Error of Approximation (RMSEA); the Standardized Root Mean Square Residual (SRMR); the Comparative Fit Index (CFI); the Tucker-Lewis Index (TLI); the Adjusted Goodness of Fit (AGFI) and the Normed Fit Index (NFI). Cronbach's

Table 1 – Descriptive statistics of the items of the Shared leadership scale in Sample 1 (N = 224) and Sample 2 (N = 220)

	Sample 1			Sample 2		
Items	**Mean (SD)**	**Skewness**	**Kurtosis**	**Mean (SD)**	**Skewness**	**Kurtosis**
1. Tutti i membri del gruppo si impegnano in comportamenti di guida del gruppo [*All team members engaged in leadership behavior*]	2.96 (1.13)	−.01	−.74	3.08 (1.04)	.01	−.58
2. Tutti i membri del gruppo offrono suggerimenti agli altri membri del gruppo per migliorare la prestazione del team [*All team members offered advice to other team members to improve team performance*]	3.51 (1.06)	−.46	−.49	3.68 (.99)	−.44	−.21
3. Tutti i membri del gruppo vanno incontro ai bisogni degli altri membri affinché quest'ultimi possano agire nel migliore dei modi [*All team members anticipated action needs of other team members*]	3.57 (1.08)	−.45	−.52	3.66 (.99)	−.41	−.39
4. Ogni membro del gruppo agisce tempestivamente affinché lo stesso gruppo si adatti ad influenze esterne [*All team members initiated action*s to adapt to external influences]	3.26 (1.00)	−.24	−.43	3.33 (.94)	−.10	−.21
5. Tutti i membri del gruppo anticipano le necessità operative del gruppo nel suo complesso [*All team members anticipated action needs of the team as a whole]*	3.07 (.96)	.02	−.47	3.17 (.98)	.05	−.24
6. Tutti i membri del gruppo avviano azioni che vanno oltre quanto richiesto dagli obiettivi di lavoro al fine di favorire una migliore prestazione dello stesso gruppo [*All team members initiated actions to foster team performance beyond their own works scope*]	2.87 (1.14)	.01	−.87	3.02 (1.06)	.12	−.61
7. Il gruppo fa affidamento su tutti i suoi membri per potersi guidare [*The team relied on all team members for leadership*]	3.25 (1.28)	−.25	−1.05	3.42 (1.12)	−.36	−.47
Mean of the scale (SD)	3.21 (.87)			3.37 (.79)		

alpha was used to test reliability. To test the possibility of aggregating the Shared leadership scale at the team level, we computed the inter-rater agreement rwg(j) (James, Demaree & Wolf, 1984) and the intraclass correlation coefficients ICC1 and ICC2 (Bliese, 2000; James, 1982). Finally, correlations at the team level were performed separately for the two studies to verify the association between the SL scale and the other variables used in this study.

RESULTS

Table 1 shows the descriptive statistics for the seven items on the SL scale for both Sample 1 and Sample 2. All the skewness and kurtosis indices in the two samples are within the range of −1 and + 1, indicating the absence of violations of normality assumptions. Accordingly, EFA was performed on Sample 1 using Maximum Likelihood parameter estimates.

The Kaiser-Meyer-Olkin (KMO) value (.897) and the significant Bartlett test results (χ^2 = 879.5 (21), p<.001) indicated that the sample was adequate for factor analysis. The factor solution yielded one factor with an eigenvalue greater than one, explaining 63.7% of the variance. Loadings, reported in Table 2, ranged between .68 and .82.

A confirmatory factor analysis was conducted on the second sample of respondents. Factor loadings are reported in Table 2 and ranged between .65 and .78. The cut-off value for CFI and TLI indices is .95; it is .90 for AGFI and NFI and below .08 for SRMR (Hu & Bentler, 1999); the rule of thumb for RMSEA is .08 or less (Browne & Cudeck, 1993). The single factor model showed an acceptable fit to the data: (χ^2 (14, N = 220) = 40.86, p<.001; (χ^2/df = 2.92; RMSEA = .09, RMR = .04, CFI = .96 and TLI = .95 (see Table 3, Model 1). The scale also showed good reliability, with Cronbach's alphas equal to or greater than .88 in the two samples.

In order to improve the fit of the model, and particularly the RMSEA and AGFI, we considered two other models, taking into account: a) modification indices suggesting the addition of an error covariance between items 5 and 6 (Model 2) and b) the removal of item 6 because, although with a factor loading of .65, it has the lowest loading compared to the other items (Model 3). Table 3 shows an improvement in the goodness-of-fit indices from Model 1 to Model 2 to Model 3, with Model 3 reporting satisfactory indices; however, we notice that Model 2 already presents satisfactory and acceptable fit indices (Raykov & Marcoulides, 2011).

In order to assess nomological validity, a component of construct validity, we checked whether the SL scale had the expected correlations with other constructs. Descriptive

Table 2 – Factor loadings of exploratory and confirmatory factor analyses of the Shared leadership scale

Items	Sample 1 *EFA*	Sample 2 *CFA*
Item 1	.68	.74
Item 2	.78	.77
Item 3	.80	.78
Item 4	.75	.72
Item 5	.79	.74
Item 6	.69	.65
Item 7	.82	.75
Alpha	.90	.89
Explained variance (%)	63.7	60.2

Table 3 – Fit indices of the confirmatory factor analysis of the Shared leadership scale (Sample 2, N = 220)

MODEL	χ^2	df	p	RMSEA (CI 90%)	SRMR	CFI	TLI	AGFI	NFI
Model 1 7 items	40.858	14	.000	.094 (.061 .128)	.039	.96	.95	.89	.95
Model 2 7 items Correlated errors: e5-e6	27.664	13	.01	.072 (.034 .109)	.032	.94	.97	.92	.96
Model 3 6 items	15.867	9	.07	.059 (.000 .106)	.025	.99	.98	.95	.97

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

statistics and alphas of the variables that we expect to be conceptually related to the SL scale are reported in Table 4: reliabilities are satisfactory for all the measures. In any case, before computing their correlations, using a direct consensus model (Chan 1998), we tested whether there was enough agreement across team members to justify the aggregation of the individual scores to the team level.

First, from the data set, we removed teams in which less than 40% of the team members answered the questionnaire. Thus, Sample 1 had a total of N = 211 respondents, for a total of 54 teams, and Sample 2 had a total of N = 176 respondents and 52 teams. Second, we computed the interrater agreement index, rwg (James et al., 1984) and the intraclass correlations coefficients, ICC1 and ICC2 (Bliese, 2000) for the SL scale in both samples. Results of the SL scores show the following: in the first sample, rwg(j) = .82, ICC1 = .30, ICC2 = .63; in the second sample, results are: rwg(j) = .81, ICC1 = .21, ICC2 = .47. These results, and those for the other scales, show a high degree of consensus across team members, and so we aggregated our measures at the team level[1].

Table 4 – Means, standard deviations and Cronbach's alpha of examined variables

	Study 1			Study 2		
	M	SD	Alpha	M	SD	Alfa
1. Team satisfaction	3.48	.86	.90	3.60	.66	.84
2. Propensity to trust	3.31	.72	.67	3.53	.62	.73
3. Team identification	3.50	.59	.78			
4. Team performance	3.74	.78	.89			
5. Affective commitment				3.64	.74	.80
6. Team elaboration				3.70	.72	.82

1 Results of rwg(j) and ICC1 and ICC2 for the other scales in the study are available from the first author.

Tables 5 and 6 show, as expected, that SL is significantly and positively correlated with the team processes and team outcomes variables we considered. Specifically, the higher the shared leadership, the greater the team satisfaction (r = .80 and r = .87, respectively, in Samples 1 and 2) and team propensity to trust (r = .51 and r = .77, respectively, in Samples 1 and 2). In addition, SL is positively related to team identification and team performance (Table 5) and to team affective commitment and team information elaboration (Table 6).

CONCLUSIONS

The main goal of this study was to investigate the factorial validity of the Italian version of the Muethel and Gehrlein (2009) scale of Shared leadership, one of the first instruments to assess how much team members believe that their team relies on the overall cumulative leadership of its members. Results of the present study support the good psychometric properties of the SL questionnaire in the Italian context, confirming the seven-item, one-factor model proposed by the authors. In order to have very good fit indices of the model, we considered to remove item 6 because, although the very good factor loading of .65 (higher than the suggested .40; Raykov & Marcoulides, 2011), error covariance of item 6 was correlated to other items (among them, the higher value was with item 5). We observed improved fit indices but we consider this an excess of zeal. In fact, even the model with the covariance between items 5 and 6 presents so good internal consistency, fit indices, and reliability that, taking into account also the useful suggestions from an anonymous reviewer, we decided to maintain the full scale. The covariance between the error terms of items 5 and 6 seems reasonable because these two items refer to proactive behaviors designed to improve operational team performance. In addition, removal of item 6 decrease minimally the explained variance and Cronbach's alphas; all this suggests to maintain item 6 and to use the complete seven-item scale. However, future studies using the scale with other samples should take into account the fit of the seven- vs six-item version of the scale.

Our two studies also suggest good nomological validity. The SL scale shows significant and consistent correlations with team processes and team outcome indicators. Our results support the literature and show that shared leadership is related to team identification (Muethel & Gehrlein, 2009) and trust towards the team (Robert & You, 2018; Small & Rentsch, 2010), and it is also related to team performance and team satisfaction, as team outcomes (Nicolaides et al., 2014; Wang et al., 2014).

This study has some limitations. First, this validation of the SL scale is restricted to students. Although the students were engaged in real teamwork where team performance was assessed (and marked by professors), it is possible that

Table 5 – Study 1: correlations between shared leadership and team processes and outcomes variables (N = 54)

	Team satisfaction	Propensity to trust	Team identification	Team performance
Shared leadership	.80**	.51**	.48**	.52**

Note. ***p*<.01.

Table 6 – Study 2: correlations between shared leadership and team processes and outcomes variables (N = 52)

	Team satisfaction	Propensity to trust	Affective commitment	Team elaboration
Shared leadership	.87**	.77**	.62**	.80**

Note. ***p*<.01.

the academic setting and the short-term nature of the project might undermine the generalizability of the results. For this reason, this study should be replicated with other teams and, particularly, teams of employees, in order to examine whether the SL scale is generalizable and can be used in professional contexts. Second, the validity assessment of the Italian version of the SL scale was limited, in this study, to internal consistency and nomological validity. It is necessary to investigate validity by testing correlations with another Shared leadership scale and with a measure of shared leadership obtained with a different method, in addition to testing concurrent validity by using some external criterion such as project teams or dispersed teams where leadership roles are shared within the team. In addition, we did not consider constructs negatively related to shared leadership, such as centralization. Third, interrater agreement and the intraclass coefficients showed that there was enough intra-team consensus to justify aggregating the answers at the team level; in addition, at the same time, ICCs suggested that there was also a group effect. Group comparison was not an aim of this paper, but future studies will have to examine the measurement invariance of this scale across different teams, by conducting, for instance, a multi-group confirmatory factor analysis.

Despite these limitations and considering the lack of similar scales in the Italian language, our results are promising. They suggest that the SL scale is a reliable and valid instrument to assess how much teams rely on the whole team for leadership.

References

BARNETT, R.C. & WEIDENFELLER, N.K. (2016). Shared leadership and team performance. *Advances in Developing Human Resources*, *18* (3), 334-351. DOI: 10.1177/1523422316645885

BATTISTELLI, A., MARIANI, M. & BELLÒ, B. (2006). Normative commitment to the organization, support and self-competence. In G. Minati, E. Pessa & M. Abram (Eds.), *Systemics of emergence: Research and development*. Boston: Springer. DOI: 10.1007/0-387-28898-8_36

BERGAMI, M. & BAGOZZI, R.P. (2000). Self-categorization, affective commitment and group self-esteem as distinct aspects of social identity in the organization. *British Journal of Social Psychology*, *39* (4), 555-577. DOI: 10.1348/014466600164633

BLIESE, P.D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K.J. Klein & S.W.J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions*. San Francisco, CA: Jossey-Bass.

BOLLEN, K.A. & LONG, J.S. (1993). *Testing structural equation models*. Newbury Park, CA: Sage.

BROWNE, M.W. & CUDECK, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models*. Newbury Park, CA: Sage.

CARSON, J.B., TESLUK, P.E. & MARRONE, J.A. (2007). Shared leadership in teams: An investigation of antecedent conditions and performance. *Academy of Management Journal*, *50* (5), 1217-1234. DOI: 10.2307/20159921

CHAN, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology*, *83* (2), 234-246. DOI: 10.1037/0021-9010.83.2.234

COSTA, A.C. & ANDERSON, N. (2011). Measuring trust in teams: Development and validation of a multifaceted measure of formative and reflective indicators of team trust. *European Journal of Work and Organizational Psychology*, *20* (1), 119-154.

DOI: 10.1080/13594320903272083

D'INNOCENZO, L., MATHIEU, J.E. & KUKENBERGER, M.R. (2016). A meta-analysis of different forms of shared leadership-team performance relations. *Journal of Management*, *42* (7), 1964-1991. DOI: 10.1177/0149206314525205

GOCKEL, C. & WERTH, L. (2010). Measuring and modeling shared leadership. *Journal of Personnel Psychology*, *9* (4), 172-180. DOI: 10.1027/1866-5888/a000023

GRILLE, A. & KAUFFELD, S. (2015). Development and preliminary validation of the Shared Professional Leadership Inventory for Teams (SPLIT). *Psychology*, *6* (1), 75-92. DOI: 10.4236/psych.2015.61008

HOEGL, M. & GEMUENDEN, H.G. (2001). Teamwork quality and the success of innovative projects: A theoretical concept and empirical evidence. *Organization Science*, *12* (4), 435- 449. DOI: 10.1287/orsc.12.4.435.10635

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

JAMES, L.R. (1982). Aggregation bias in estimates of perceptual agreement. *Journal of Applied Psychology*, *67* (2), 219-229. DOI: 10.1037/0021-9010.67.2.219

JAMES, L.R., DEMAREE, R.G. & WOLF, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, *69* (1), 85-98. DOI:10.1037/0021-9010.69.1.85

KEARNEY, E., GEBERT, D. & VOELPEL, S.C. (2009). When and how diversity benefits teams: The importance of team members' need for cognition. *Academy of Management Journal*, *52* (3), 581-598. DOI: 10.5465/amj.2009.41331431

MAEL, F. & ASHFORTH, B. (1992). Alumni and their alma mater: A partial test of the reformulated model of organizational identification. *Journal of Organizational Behavior*, *13* (2), 103-123. DOI: 10.1002/job.4030130202

MEYER, J.P. & ALLEN, N.J. (1991). A three-component conceptualization of organizational commitment. *Human Resource Management Review*, *1* (1), 61-89. DOI: 10.1016/1053-4822(91)90011-Z

MUETHEL, M. & GEHRLEIN, S. (2009). Motivational aspects of shared leadership: The role of team identification and team self-efficacy. *Proceedings of the Academy of Management Annual Conference*, Chicago, IL (7-11 August).

MUETHEL, M., GEHRLEIN, S. & HOEGL, M. (2012). Socio-demographic factors and shared leadership behaviors in dispersed teams: Implications for human resource management. *Human Resource Management*, *51* (4), 525-548. DOI: 10.1002/hrm.21488

NICOLAIDES, V.C., LAPORT, K.A., CHEN, T.R., TOMASSETTI, A.J., WEIS, E.J., ZACCARO, S.J. & CORTINA, J.M. (2014). The shared leadership of teams: A meta-analysis of proximal, distal, and moderating relationships. *The Leadership Quarterly*, *25* (5), 923-942. DOI: 10.1016/j.leaqua.2014.06.006

NUNNALLY, J.C. & BERNSTEIN, I.H. (1994). *Psychometric Theory* (*3rd ed.*). New York, NY: McGraw-Hill.

PEARCE, C.L. & CONGER, J.A. (2003). *Shared leadership: Reframing the hows and whys of leadership*. Thousand Oaks, CA: Sage. DOI: 10.4135/9781452229539

RAYKOV, T. & MARCOULIDES, G.A. (2011). *Introduction to psychometric theory*. New York, NY: Routledge.

ROBERT, L.P. & YOU, S. (2018). Are you satisfied yet? Shared leadership, individual trust, autonomy, and satisfaction in virtual teams. *Journal of the Association for Information Science and Technology*, *69* (4), 503-513. DOI: 10.1002/asi.23983

SMALL, E.E. & RENTSCH, J.R. (2010). Shared leadership in teams. *Journal of Personnel Psychology*, *9* (4), 203-211. DOI: 10.1027/1866-5888/a000017

SMITH, J.B. & BARCLAY, D.W. (1997). The effects of organizational differences and trust on the effectiveness of selling partner relationships. *Journal of Marketing*, *61* (1), 3-21. DOI: 10.2307/1252186

WANG, D., WALDMAN, D.A. & ZHANG, Z. (2014). A meta-analysis of shared leadership and team effectiveness. *Journal of Applied Psychology*, *99* (2), 181-198. DOI: 10.1037/a0034531

ZHU, J., LIAO, Z., YAM, K.C. & JOHNSON, R.E. (2018). Shared leadership: A state-of-the-art review and future research agenda. *Journal of Organizational Behavior*, *39* (7), 834-852. DOI: 10.1002/job.2296

DOI: 10.26387/bpa.283.5

Factorial validity of the Italian version of the Contextual Sensation Seeking Questionnaire for Skiing and Snowboarding (CSSQ-S)

Claudia Marino[1,2], Sergio Agnoli[3], Luca Scacchi[2], Maria Grazia Monaci[2]

[1] Dipartimento di Psicologia dello Sviluppo e della Socializzazione, Università degli Studi di Padova
[2] Dipartimento di Scienze Umane e Sociali, Università della Valle d'Aosta, Aosta
[3] Marconi Institute for Creativity, Bologna

ABSTRACT. La ricerca di sensazioni è stata spesso associata alla messa in atto di comportamenti rischiosi, come gli sport invernali fuoripista. Lo scopo del presente studio è di esaminare la validità fattoriale della versione italiana del *Contextual Sensation Seeking Questionnaire* per sciatori e snowboarder (CSSQ-S). La scala sembra avere buone caratteristiche psicometriche e la struttura fattoriale sembraessere parzialmente invariante tra sciatori e snowboarder di diverse età che praticano attività fuoripista abitualmente *vs* occasionalmente.

SUMMARY. Sensation seeking has been often associated with at risk snow behaviors. Recent research has highlighted the need to develop a specific measure to assess sensation seeking in specific activities like skiing and snowboarding. The aim of the current study was to examine the factorial validity of the Italian version of the Contextual Sensation Seeking Questionnaire for Skiing and Snowboarding (CSSQ-S). 434 skiers and snowboarders (aged 18-84 years) participated in the study. Confirmatory factor analyses were performed in order to assess the factorial validity of the scale. Results revealed that the factor structure of the CSSQ-S provided a good fit to the data. This study found the CSSQ-S to be reliable, partially invariant across occasional and habitual skiers and snowboarders, and have concurrent and convergent validity. This scale provides a useful tool to assess sensation seeking among skiers and snowboarders in Italian-speaking population.

Keywords: *Sensation seeking, Factorial validation, Skiing, Snowboarding, Multigroup analysis*

INTRODUCTION

Over the last decades, downhill winter sports have become very popular worldwide, with 6 to 15 millions of skiers and snowboarders registered in the U.S. and Europe in 2018 (Kopp, Wolf, Ruedl & Burtscher, 2016; Statista, 2018). Due to the huge amount of people engaging in such high-risk sports, concerns over the psychological factors involved in backcountry skiing and snowboarding have risen along with the increasing number of fatal or severe injuries, especially in the Alps (Höller, 2017). For example, rates of avalanche accidents in Italian mountains have increased over the last decade with a pick of deaths registered in the last few years (e.g., Pfeifer, Höller & Zeileis, 2018; Servizio Valanghe Italiano, 2018), leading for calls to understand backcountry skiers' behaviors, identify risk-seeking people, and actively prevent those accidents (Marengo, Monaci & Miceli, 2017).

Specific skiers' and snowboarders' profiles are more likely to take up the inherent risk of these sports as they have peculiar personality characteristics, like high levels of sensation seeking (Zuckerman, 2007). Beyond the several available methods to measure general sensation seeking (e.g., ZKPQ ImpSS, Sensation Seeking Scale-SSS; De Pascalis & Russo, 2003; Manna, Faraci & Como, 2013; Rossi & Cereatti, 1993), Thomson, Morton, Carlson and Rupert (2012) have recently argued that, in the context of winter sports, the adoption of a context-specific measure of sensation seeking is important to assess the specific psychological processes involved in skiing and snowboarding. The authors have validated the *Contextual Sensation Seeking Questionnaire for Skiing and Snowboarding* (*CSSQ*-S; Thomson et al., 2012) showing strong psychometric properties and sustaining that the actual behavioral tendencies are better predicted by a contextual propensity to engage in a given activity rather than by general personality traits. In this view, the 10 items of the CSSQ-S appear to "measure a person's tendency to seek out new, thrilling, or physically stimulating experiences while engaged in downhill sports, regardless of potential hazards" (Thomson et al., 2012, p. 515).

The scale has been successfully used in several recent studies (Garner, Haegeli & Haider, 2016; Maher, Thomson & Carlson, 2015; Thomson & Carlson, 2015; Thomson, Rajala, Carlson & Rupert, 2014). However, as the majority of these studies focused on North American and young populations, the current study deals with the question on the validity of the scale in other cultures and contexts, offering a contribution to its validation, by testing the factorial validity supported in previous studies in Italy (Marengo et al. 2017). Specifically, the aim of the present study is twofold: (1) to present the psychometric properties of the CSSQ-S for Italian skiers and snowboarders, and (2) to show the measurement invariance of the scale across habitual and occasional backcountry skiers/snowboarders. Whereas the first aim concerns the exploration of the internal validity of the scale in the Italian context, through the second aim the current study intended to analyze the scale measurement properties across two skiers/snowboarders profiles that are typically located in the Alps context and that are expected to be exposed to different rates of risky behaviors.

In the present study several winter activities involving skiing and snowboarding have been taken into account. As briefly outlined above, literature suggests that some forms of downhill winter sports are characterized by higher sensation seeking levels than others (e.g., ski mountaineering and snowboard vs alpine ski; Kopp et al., 2016). However, here we propose that in the analysis of sensation seeking in winter sports not only the type of sport should be taken into account, but also the risk taking propensity in performing such activities. A well defined trend has taken hold indeed in the Alps: winter sports, such as ski mountaineering and freeride, which were traditionally related to backcountry environments (i.e., to avalanche terrains), begun to be performed within ski resorts or on prearranged paths within controlled areas. Contrary to classic winter backcountry recreationists who took a risk in skiing in avalanche terrains, recreationists following to this new trend are not interested in the adventurous dimension, but are instead moved by the performance, the physical fitness, or the well-being generated by these sport activities (Perrin-Malterre & Chanteloup, 2018). Adventurous terrains lose all interests for members of this second winter recreationists category, who tend instead to perform their activities in controlled, low-risk, prearranged terrains which easily allow them to train or to increase their physical fitness. Therefore, two categories can be defined based on the frequency of reported backcountry activities: habitual or classic backcountry skiers/snowboarders and occasional backcountry skiers/snowboarders. Testing the invariance of the CSSQ-S scale in these two categories of skiers/snowboarders characterized by different attitudes toward risk allows to test the soundness of this instrument across two apparently different sensation seekers profiles.

In the present study the CSSQ-S factorial structure as well as its convergent and concurrent validity were tested. To this aim, measures of sensation seeking and risk taking along with questions on participants' habits and behaviors related to mountain activities were included in a comprehensive set of questionnaires containing also the CSSQ-S Italian version.

Moreover, in Thomson et al.'s study (2012) the sample mean age was only 27.1 years (*SD* = 4.8). However, Breivik (2010) has recently showed that "adventurous sports" (including skiing and snowboarding) have been developing in the last 30 years along with an increasing popularity among young as well as old people. The author suggested that seeking for excitement is relevant both for youths and old people (Breivik, 2010). Therefore, sensation seeking might be a valid construct to be investigated among (older) adults, especially in Italian alps where a broad range of active athletes and leisure sportsmen and women are over 50 years people (FISI, 2018-2019). For these reasons, the factorial structure of the scale was also explored across younger vs older adults.

METHODS

Procedure

The sample was recruited online by sending the link of a questionnaire to thematic e-mail lists (e.g., ski schools' attenders) and sharing the link in social network groups as well as in the Facebook page of a nonprofit foundation based in Valle d'Aosta, Italy (Marengo et al., 2017). The survey was accessible online from 8th May 2017 to 10th August 2017. Participants were asked to give their consent in the first page of the study website, which explained the purpose of the study and assured the anonymity of the responses. Participants were then directed to a second page containing demographic information and a series of self-report scales (see *Measures* section).

Participants

A total of 450 people accessed the questionnaire. Sixteen participants declared not to be involved in skiing nor snowboarding and were excluded from analyses, which were run on a final sample of 434 skiers and snowboarders (311 males, 123 females, M_{age} = 41,34, *SD* = 13.42, *range* = 18-84; the 93% of the sample is below 60 years old). The majority of the sample (97%) was Italian and 41.7% of the sample had high education degrees (that is, at list graduated). Half of the participants (50.7%) reported to have been involved in mountain professional activities during the last winter season (such as, alpine guides, ski instructors, pisteur-secouriste).

Measures

At the beginning of the questionnaire, participants were asked to complete a brief demographic section (e.g., age, gender, education, nationality) and a series of mountain-related questions about ability, and habits regarding the frequency of backcountry activities. Then, they were asked to complete the CSSQ-S and a few other questions in order to evaluate criterion-related validity of the CSSQ-S.

- *Contextual Sensation Seeking Questionnaire for Skiing and Snowboarding* (*CSSQ-S*). The CSSQ-S comprised ten items related to personal experience in skiing and snowboarding and developed by Thomson and colleagues (2012). Items were translated from English to Italian and back-translated in English by a bilingual psychologist expert in the field. Participants were asked to rate the extent to which they agreed with each item on a 5-point scale (from 1 = definitely disagree to 5 = definitely agree). Items were averaged to obtain a score of contextual sensation seeking. Higher scores indicate higher levels of sensation seeking. The full list of items (both in English and Italian) is reported in Table 1.
- *Impulsive sensation seeking* (*ImpSS*). Participants' general impulsive sensation seeking was assessed using the impulsive sensation seeking subscale (ImpSS) of the Italian version of the Zuckerman-Kuhlman Personality Questionnaire (ZKPQ; De Pascalis & Russo, 2003; Zuckerman et al., 1993). The scale comprised 19 items (e.g., "I often do things on impulse"; "I enjoy getting into new situations where you cannot predict how things will turn out"). Participants rated their agreement with each item related to sensation seeking and impulsive behavior in everyday situations on a 5-point scale (from 1 = definitely disagree to 5 = definitely agree). Responses were averaged and higher scores reflected more impulsivity and sensation seeking. The Cronbach's alpha for the scale was .90 (95% CI .88-.91).

Table 1 – Standardized factor loadings for the CSSQ-S (N = 434)

Items (Italian)	Items (English)	Standardized factor loadings
1. Mi piace andare veloce	1. I like to ski/ride fast	.527
2. Mi piace fare discese che non ho mai affrontato prima	2. I like to ski/ride down runs that I have never been down before	.428
3. Mi piace iniziare una discesa anche se non riesco a vedere come si presenta (ad es., una grossa cornice in ingresso)	3. I like to start a run even if I cannot see what lies ahead (i.e., big cornice)	.451
4. Mi piace andare all'esterno delle piste controllate e aperte	4. I like to ski/ride out of bounds	.573
5. Mi piace tentare salti anche se non sono certo della qualità della neve che troverò all'atterraggio	5. I like to attempt jumps even if I'm not sure of the quality of the landing area	.572
6. Mi piace spingermi oltre i miei limiti	6. I like to push my boundaries when I ski/ride	.675
7. Se perdo il controllo, non tento subito di rallentare, ma mi lascio andare	7. If I lose control, I don't try to immediately slow down, I just go with it	.414
8. Se una discesa prevede il passaggio in una lunga strettoia in rettilineo, l'affronto senza esitazione anche se so che dovrò andare molto veloce	8. If the only way down is a straight line through a narrow pass, I go for it without hesitation even if I know I will have to go fast	.616
9. Cerco sempre di trovare modi nuovi ed eccitanti di affrontare una discesa	9. I am always trying to find new and exciting ways down a run	.733
10. Un cliff di 4m non è un salto troppo alto per me	10. A 15-foot high drop off a cliff isn't too high a jump for me	.493

Note. Response format from 1 = definitely disagree to 5 = definitely agree.
Instruction for the Italian version: "Di seguito sono riportate alcune affermazioni che descrivono diversi modi di affrontare l'ambiente innevato su sci o snowboard. Le chiediamo di leggere con attenzione e, pensando alla Sua esperienza, di barrare la casella che meglio esprime il Suo grado di accordo".
Instruction for the English version: "The following statements describe different ways to deal with the snowy environment on skis or snowboards. Carefully read each question and tick the response that best expresses your level of agreement, thinking of your own experience".

– *Risk taking propensity.* Risk propensity was assessed with 8 ad hoc items (e.g., "I evaluate both the difficulty of the track and the snow conditions before going downhill"; "I reduce the speed if visibility is limited"). Participants were asked to rate the extent to which they agreed with each of the items on a 5-point scale (from 1 = definitely disagree to 5 = definitely agree). Items were averaged to obtain a total score of risk taking propensity. Higher scores indicate lower levels of risk taking propensity. The Cronbach's alpha was .67 (95% CI .62-.72).
– *Avalanche danger*. Participants were asked to indicate both "the most frequent" and the "highest" danger level

in which they engaged in their skiing/snowboarding activities during the last year. The response scale was based on the European scale of Avalanche danger (available at http://www.avalanches.org/eaws/en/main_layer.php?layer=basics&id=2) that was included in the questionnaire before the presentation of the questions pertaining the avalanche danger. This scale ranged from 1 = low danger to 5 = very high danger. Through these two questions, two self-reported indicators were obtained: (1) *the most frequent danger level*, and (2) the *highest danger level*.

Statistical analysis

First, a Confirmatory Factor Analysis (CFA) using the Lavaan package of software R was run, using Weighted least estimation with robust standard errors and mean and variance (WLSMV) estimator for ordinal items. The following indices were used to assess the fit of the model: (1) chi-square (χ^2); (2) Comparative Fit Index (CFI; acceptable fit ≥.90); (3) Goodness of Fit Index (GFI; acceptable fit ≥.90); (4) Tucker-Lewis Index (TLI; acceptable fit ≥.90); and (5) Root Mean Square Error of Approximation (RMSEA; acceptable fit ≤.08). Cronbach's alpha was employed to assess internal consistency of the scale.

Second, using rank analysis, two different groups were identified based on the frequency of reported backcountry activities (i.e., ski, ski mountaineering, snowboard, and freeride). Therefore, the model was tested separately on the two groups: occasional backcountry skiers and snowboarders (N = 248) and habitual backcountry skiers and snowboarders (N = 186) (labeled occasional *vs* habitual) to establish configural invariance (Van de Schoot, Lugtig & Hox, 2012). After this, a multi-group CFA was performed to examine measurement invariance of the CSSQ-S across the two groups. A hierarchical approach was adopted by successively constraining model parameters and comparing changes in model fit (Van de Schoot et al., 2012). Metric and scalar models were also estimated. Measurement invariance was established when: (a) the change in values for fit indices ($\Delta\chi^2$, ΔCFI, ΔTLI, ΔRMSEA) was negligible (that is, a significant $\Delta\chi^2$, ΔCFI and ΔTLI larger than .01, and a change larger than .015 in RMSEA are indicative of non-invariance; Cheung & Rensvold, 2002; Gilson et al., 2013; Van de Schoot et al., 2012); and (b) the multi-group model fit indexes indicated a good fit (Beaujean, Freeman, Youngstrom & Carlson, 2012).

Then, the above described procedure was followed to test the invariance of the model across two age groups: younger adults (N = 258; aged between 18 and 45 years) and older adults (N = 176; aged between 46 and 84 years).

Third, according to the procedure applied by the authors of the original version of the scale (Thomson et al., 2012), we tested the association of CSSQ-S with education. We also performed an independent-samples t-test in order to test the mean difference of CSSQ-S scores between professional and recreational skiers and snowboarders.

Finally, Pearson's correlation was used to test the association between CSSQ-S and ImpSS to establish evidence of concurrent validity. Finally, the correlations between CSSQ-S and risk taking propensity and the two indicators of avalanche danger level were computed to test for convergent validity.

RESULTS

Confirmatory Factor Analysis: measurement invariances

Results of CFA for the global model showed an adequate fit to the data: $\chi^2_{(35)}$ = 71.49, p<.001, CFI = .979, GFI = .998, TLI = .973, RMSEA = .049 [.033-.065]. Standardized loadings ranged between .41 and .73 (see Table 1). The internal consistency of the scale's scores was α = .81 (95% CI .79-.84). Moreover, results (see Table 2) demonstrated that the model fit was adequate to excellent for both groups of backcountry skiers/snowboarders (occasional: $\chi^2_{(35)}$ = 43.637, p = .15, CFI = .991, GFI = .998, TLI = .989, RMSEA = .032 [.000-.059]; habitual: $\chi^2_{(35)}$ = 56.871, p = .01, CFI = .963, GFI = .997, TLI = .952, RMSEA = .058 [.028-.085]).

Regarding model invariance, the fit indices of the unconstrained multi-group model ($\chi^2_{(70)}$ = 100.51, p = .01, CFI = .981, TLI = .975, RMSEA = .045 [.023-.064]) demonstrated the configural invariance of the model across groups, suggesting that the factor structure is similar across the two groups. In the subsequent metric model, all item loadings were constrained to equality and differences in fit indexes did not reveal globally a significant reduction in model fit ($\Delta\chi^2_{(5.15)}$ = 5.64, p = .36, ΔCFI = .014, ΔTLI = .012, ΔRMSEA = .009), suggesting that the meaning of the construct assessed by CSSQ-S is similar across both

Table 2 – Fit indices for measurement invariance tests on the CSSQ-S (occasional and habitual backcountry skiers and snowboarders)

Model	N	χ^2(df)	$\Delta\chi^2$(df)	CFI	ΔCFI	TLI	ΔTLI	RMSEA	ΔRMSEA
Occasional	248	43.64(35)	-	.991	-	.989	-	.032	-
Habitual	186	56.87(35)	-	.963	-	.952	-	.058	-
Model 1	434	100.51(70)*	-	.981	-	.975	-	.045	
Model 2	434	131.24(80)*	5.64(5.15)	.967	.014	.963	.012	.054	.009
Model 3	434	160.07(87)*	19.16(7.81)*	.932	.035	.932	.031	.074	.020
Model 4	434	160.07(87)*	12.04(7.29)	.953	.014	.952	.011	.062	.008

Legenda. df = degree of freedom; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index.

Note: *p<.01; Model 1 = Configural invariance; Model 2 = Metric invariance; Model 3 = Scalar invariance; Model 4 = Partial invariance with unconstrained thresholds of item 4 and 10.

occasional and habitual backcountry skiers. Finally, all the item thresholds were constrained across groups to test for scalar invariance. Results showed that the fit of the scalar model is significantly worse than the metric model ($\Delta\chi^2_{(7.81)}$ = 19.16, p<.01, ΔCFI = .035, ΔTLI = .031, ΔRMSEA = .020). Therefore, a further model was tested constraining all item thresholds except for the two thresholds with the largest unstandardized difference. That is, thresholds of item 4 and item 10 were released to try to establish partial scalar invariance (Steenkamp & Baumgartner, 1998). The fit of this new model was not significantly worse than the previous one ($\Delta\chi^2_{(7.29)}$ = 12.04, p = .12, ΔCFI = .014, ΔTLI = .011, ΔRMSEA = .008), thus supporting partial invariance (Van de Schoot et al., 2012).

With regard to invariance across age groups (see Table 3), results of the tested model in both groups separately showed that the fit indices in both cases are excellent, suggesting that the scale could constitute an overall good measure for both age groups: younger adults: $\chi^2_{(35)}$ = 56.04, p = .013, CFI = .981, GFI = .998, TLI = .975, RMSEA = .048 [.022-.071]; older adults: $\chi^2_{(35)}$ = 42.06, p = .192, CFI = .982, GFI = .997, TLI = .977, RMSEA = .038 [.000-.067].

The multi-group analysis comparing the two age groups showed that that fit indices were very good, indicating that the construct holds across the two groups: $\chi^2_{(70)}$ = 98.10, p = .015, CFI = .981, GFI = .998, TLI = .976, RMSEA = .043 [.020-.062].

However, the metric invariance was not totally supported. Therefore, in line with the above presented findings on habitual/occasional skiers, we constrained all the items except for item 4 and 10. Also item 5 (which showed the greater difference in loadings) was released. In this way, the metric invariance was partially supported ($\Delta\chi^2_{(4.81)}$ = 4.70, p = .43, ΔCFI = .011, ΔTLI = .011, ΔRMSEA = .009) as well as the partial scalar invariance ($\Delta\chi^2_{(6.45)}$ = 9.98, p = .15, ΔCFI = .007, ΔTLI = .004, ΔRMSEA = .002).

CSSQ-S validity

With regard to demographic variables, results (see Table 4) showed a significant negative association between CSSQ-S scores and age (r = −.40, p<.001) and a non-significant association with education (r = −.04, p>.05).

Table 3 – Fit indices for measurement invariance tests on the CSSQ-S (young and old backcountry skiers and snowboarders)

Model	N	χ^2(df)	$\Delta\chi^2$(df)	CFI	ΔCFI	TLI	ΔTLI	RMSEA	ΔRMSEA
Younger adults	258	56.04(35)	-	.981	-	.975	-	.048	-
Oder Adults	176	42.06(35)	-	.982	-	.977	-	.038	-
Model 1	434	98.10(70)*	-	.981	-	.976	-	.043	
Model 2	434	162.60(80)*	10.75(5.97)	.944	.037	.937	.039	.069	.026
Model 3	434	121.45(77)*	4.70(4.81)	.970	.011	.965	.011	.052	.009
Model 4	434	140.95(87)*	9.98 (6.45)	.963	.007	.961	.004	.054	.002

Legenda. df = degree of freedom; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index.

Note. *p<.01; Model 1 = Configural invariance; Model 2 = Metric invariance; Model 3 = Partial metric invariance; Model 4 = Partial scalar invariance with unconstrained thresholds of item 4, 5, and 10.

Table 4 – Means and bivariate correlations (N = 434)

	M (SD)	1	2	3	4	5	6
1. CSSQ-S score	2.71(.70)	-					
2. Age	41.34(13.42)	−.40*	-				
3. Education	4.27(1.02)	−.04	−.12*	-			
4. ImpSS	2.27(.66)	.56*	−.35*	−.02	-		
5. Risk taking propensity	4.37(.49)	−.41*	.18*	−.01	−.30*		
6. Avalanche danger 1	2.12(.82)	.21*	−.18*	.02	.19*	−.13	-
7. Avalanche danger 2	2.84(1.06)	.27*	−.29*	.01	.28*	−.17*	.68*

Legenda. ImpSS = Impulsive Sensation Seeking.

Note. *p<.001; Avalanche danger 1 = the most frequent danger level; Avalanche danger 2 = the highest danger level.

Results of the *t*-test ($t_{(432)}$ = 2.79, p<.05) indicated that professionals (M = 2.81, SD = .73; N = 220) have slightly higher scores in CSSQ-S than recreationists (M = 2.62, SD = .65; N = 214).

With respect to the association between CSSQ-S and ImpSS, the correlation was .56 (p<.001), thus providing evidence for concurrent validity. In line with Thomson et al. (2012), a moderate correlation sustains the association between the two constructs without being overlapped. Moreover, the association between CSSQ-S and risk taking propensity (r = −.41, p<.001) indicates the concurrent validity of the scale. Also, Table 4 shows the low to moderate associations between CSSQ-S and the two indicators of avalanche danger levels suggesting convergent validity for CSSQ-S scores.

DISCUSSION

This study was designed to answer two questions: (a) has the Italian version of the CSSQ-S good psychometric properties? and (b) is the factor structure of the scale invariant across skiers/snowboarders characterized by different sensation seeking profiles (i.e., habitual *vs* occasional backcountry skiers/snowboarders)? Overall, results demonstrated good factorial structure of the CSSQ-S (Italian version) for the total sample and for both habitual and occasional backcountry skiers and snowboarders. This result suggests that the scale is suitable for research in the Italian context. Moreover, with regard to the validity of the CSSQ-S, our results partially replicated the findings from the original validation paper (that is, a non-significant association between CSSQ-S and education; significant links with professional activity, general impulsive sensation seeking, risk taking propensity, and avalanche danger levels).

Regarding invariance across habitual and occasional backcountry skiers/snowboarders, although the fit indices of the multiple group analyses emerged to be only sufficient to probe an adequate fit, taken together such indices did not support a statistically significant reduction of the metric model fit, showing that the items were similarly interpreted across the two groups of skiers and snowboarders. However, for scalar invariance, results revealed that item 4 ("I like to ski/ride out of bounds") and item 10 ("A 15-foot high drop off a cliff isn't too high a jump for me") seem to negatively influence the fit of the model. A partial scalar invariance for CSSQ-S was established, showing that the scale operates in a similar fashion across the two groups with the exception of item 4 and 10. The specific content of these two items appears more appropriate for the habitual group of backcountry sportspersons than for the occasional group, who is likely to be less prone to engage as well as less experienced in skiing out of bounds and in doing jumping cliffs. Therefore, this result suggests that the scale may be improved through close analyses of the item content, which could allow a specific adaptation of the scale for peculiar skiers/snowboarders' profiles, for example by considering the exclusion of certain items for athletes not interested or experienced in backcountry activity. Thus, these results highlighted the need for researchers who want to measure athletes' sensation seeking, to carefully take into account not only the different types of winter sports, but also the relevant difference between the activities in controlled terrain and off-piste downhill sports (e.g., Martha, Sanchez & Gomà-i-Freixanet, 2009). Moreover, it has been found that item 4, 5 ("I like to attempt jumps even if I'm not sure of the quality of the landing area") and 10 might have slightly different meanings for younger vs older skiers and snowboarders and, subsequently, they have different mean levels across groups. Following the same line of reasoning, it could be argued that jump-related items could have different meanings for older people, in that old skiers and snowboarders may tend to jump a cliff less often than younger ones. In this view, although the scale shows good properties across both age groups, researchers and practitioners who are willing to use the scale among Italian-speaking older adults, should take answers to item 4, 5, and 10 cautiously, as they might have slightly different meanings and levels for younger vs. older adults.

This study has some limitations. For example, it does not provide information about test-retest reliability nor predictive validity, which were provided by Thomson et al. (2012). Future studies are therefore needed to analyze these forms of validity in the Italian version of CSSQ-S. These limitations notwithstanding, the present study offers new insight on the statistical properties of this scale which could be used by researchers and practitioners to gain an in-depth understanding of sensation seeking in winter activities in the Italian context, with a specific focus on activities performed both in controlled terrains and in backcountry areas.

Funding sources: No financial support was received for this study.
Conflict of interest: The authors declare no conflict of interest.

References

BREIVIK, G. (2010). Trends in adventure sports in a post-modern society. *Sport in society*, *13* (2), 260-273.

DE PASCALIS, V. & RUSSO, P.M. (2003). Zuckerman-Kuhlman Personality Questionnaire: Preliminary results of the Italian version. *Psychological Reports*, *92*, 965-974.

FISI Italian Federation of Winter Sports (2018-2019). https://www.fisi.org/images/federazione/documenti/agenda/AGENDA-COMPLETA-2018-2019.pdf last access: December 13rd 2018.

HÖLLER, P. (2017). Avalanche accidents and fatalities in Austria since 1946/47 with special regard to tourist avalanches in the period 1981/82 to 2015/16. *Cold regions science and technology*, *144*, 89-95.

KOPP, M., WOLF, M., RUEDL, G. & BURTSCHER, M. (2016). Differences in sensation seeking between alpine skiers, snowboarders and ski tourers. *Journal of Sports Science & Medicine*, *15* (1), 11-16.

MAHER, A.M., THOMSON, C.J. & CARLSON, S.R. (2015). Risk-taking and impulsive personality traits in proficient downhill sports enthusiasts. *Personality and Individual Differences*, *79*, 20-24.

MANNA, G., FARACI, P. & COMO, M. (2013). Factorial structure and psychometric properties of the Sensation Seeking Scale – Form V (SSS-V) in a sample of Italian adolescents. *Europe's Journal of Psychology*, *9*, 276-288.

MARENGO, D., MONACI, M.G. & MICELI, R. (2017). Winter recreationists' self-reported likelihood of skiing backcountry slopes: Investigating the role of situational factors, personal experiences with avalanches and sensation-seeking. *Journal of Environmental Psychology*, *49*, 78-85.

MARTHA, C., SANCHEZ, X. & GOMÀ-I-FREIXANET, M. (2009). Risk perception as a function of risk exposure amongst rock climbers. *Psychology of Sport and Exercise*, *10* (1), 193-200.

PERRIN-MALTERRE, C. & CHANTELOUP, L. (2018). Ski touring and snowshoeing in the Hautes-Bauges (Savoie, France): A study of various sports practices and ways of experiencing nature. *Journal of Alpine Research – Revue de géographie alpine*, *106* (4).

PFEIFER, C., HÖLLER, P. & ZEILEIS, A. (2018). Spatial and temporal analysis of fatal off-piste and backcountry avalanche accidents in Austria with a comparison of results in Switzerland, France, Italy and the US. *Natural Hazards and Earth System Sciences*, *18* (2), 571-582.

ROSSI, B. & CEREATTI, L. (1993). The sensation seeking in mountain athletes as assessed by Zuckerman's Sensation Seeking Scale. *International Journal of Sport Psychology*, *24* (4), 417-431.

SERVIZIO VALANGHE ITALIANO (2018). http://www.cai-svi.it/ last access: July 12nd 2018.

STATISTA (2018). https://www.statista.com/statistics/801008/europe-number-of-people-skiing-by-country/ https://www.statista.com/topics/1770/winter-sports/ last access June 30th 2018.

THOMSON, C.J. & CARLSON, S.R. (2015). Increased patterns of risky behaviours among helmet wearers in skiing and snowboarding. *Accident Analysis & Prevention*, *75*, 179-183.

THOMSON, C.J., MORTON, K.L., CARLSON, S.R. & RUPERT, J.L. (2012). The Contextual Sensation Seeking Questionnaire for skiing and snowboarding (CSSQ-S): Development of a sport specific scale. *International Journal of Sport Psychology*, *43*, 503-521.

THOMSON, C.J., RAJALA, A.K., CARLSON, S.R. & RUPERT, J.L. (2014). Variants in the dopamine-4-receptor gene promoter are not associated with sensation seeking in skiers. *PloS ONE*, *9* (4), e93521.

Van DE SCHOOT, R., LUGTIG, P. & HOX, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*, 486-492.