TPM - TESTING, PSYCHOMETRICS, METHODOLOGY IN APPLIED PSYCHOLOGY

CONTENTS

Vol. 24 No. 4 December 2017

Journal Staff: Via Valerio Flacco n. 10, 35128 Padova (Italy); tel. (0039) 049 80.74.522; telefax (0039) 049 80.74.492; website: www.tpmap.org, e-mail: tpmap@tpmap.org.

Meeting point: 4-5 Gough Square (second floor), London EC4A 3DE.

Quarterly journal registered at the Court of Padova on the 29th of November 1990 n. 1253.

Publisher: Cises Srl, Via Valerio Flacco n. 10, 35128 Padova (Italy).

Journal articles are abstracted in PsycINFO, Scopus, and Ebsco.

Subscription rates: 4 issues € 180,00 (print and online, for libraries and institutions), € 140 (print, for libraries and institutions), € 90 (online, for libraries and institutions), € 90 (print, for individuals), € 50 (online, for individuals).
The cost per single article is € 5,00.
Payment by bank transfer c/o Banca Popolare di Vicenza, Padova branch – ABI: 05728 – CAB: 12100 – CIN: A – c/c 022570525370 – IBAN: IT 85 A 05728 12100 022570525370 – SWIFT: BPVIIT2202.

TPM Vol. 24, No. 4, December 2017.

Printed by the Publisher at Via Valerio Flacco n. 10, Padova.

TPM®

IMPACT OF QUESTIONNAIRE FORMAT ON RELIABILITY, VALIDITY, AND HYPOTHESIS TESTING

MADHUBALAN VISWANATHAN
UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

UJWAL KAYANDE
MELBOURNE BUSINESS SCHOOL

RICHARD P. BAGOZZI
UNIVERSITY OF MICHIGAN

SAM RIETHMULLER
COLES FLYBUYS, MELBOURNE

SHIRLEY Y. Y. CHEUNG
HONG KONG BAPTIST UNIVERSITY

Researchers use a variety of questionnaire formats to collect data on measures of constructs for theory testing. For example, researchers may label measures of constructs, present measures of different constructs on different pages, or intersperse items from different constructs. Such questionnaire format choices are often guided by commonly held beliefs and conventions. However, there is little if any empirical research evaluating such different formats, precluding an informed view of the appropriateness of a format for a study. To our knowledge, our research represents the first systematic empirical investigation of how different formats affect research outcomes. We conduct a series of studies to systematically examine the effects of questionnaire format on the psychometric properties of measures of constructs and the relationships between constructs. Using multiple group confirmatory factor analysis (MGCFA), we find that the measures are largely invariant to questionnaire format when using student samples, and recommend that researchers should reconsider the rationale provided for choosing specific formats.

Key words: Measurement; Reliability; Theory-testing; Multiple group confirmatory factor analysis; Questionnaire formats.

Correspondence concerning this article should be addressed to Madhubalan Viswanathan, University of Illinois Urbana-Champaign, 183 Wohlers Hall, 1206 South Sixth Street, Champaign, IL 61820, USA. Email: mviswana@illinois.edu

Researchers commonly use questionnaires with individual or organizational foci and collect data on a variety of measures of constructs. Such questionnaires can be structured in many different ways. Measures of dimensions of a construct could be labeled, or presented separately on different pages. Items from measures of multiple dimensions or even different constructs could be interspersed. The sequencing of items within a measure of a dimension could be changed. Although

the actual layout of measures in a questionnaire is not commonly discussed in a published paper, researchers tend to make a conscious decision about the structure — that is, format — based on commonly held beliefs and conventions.

Underlying the different practices followed by researchers is the rationale that each of these practices ameliorates a measurement-related problem that eventually affects research outcomes (e.g., relationships between variables in tests of hypotheses). Yet, there has not been a systematic empirical investigation of whether and how different format practices affect research outcomes. As a consequence, systematic guidance on the choice of questionnaire format is very limited. In essence, the lack of systematic research in this area calls into question the reliability and validity of measures and inferences about relationships between those constructs in past research, based on the varying use of practices such as labeling or interspersing. Given the lack of empirical work on this topic, current research design practice rests on conceptual arguments and implicit knowledge about possible outcomes of different questionnaire format alternatives.

The task of identifying methodological factors that affect reliability and validity is itself challenging, and relevant factors may often be considered too idiosyncratic to study systematically. Moreover, much of current knowledge or working hypotheses in this realm may well be tacit or implicit in nature, based on individual researchers' experiences. However, such tacit working hypotheses need to be explicated and studied systematically. Our main objective here is to systematically examine the effects of methodological factors stemming from questionnaire format on the psychometric properties of measures of constructs and on the relationships between constructs. Specifically, we investigate the impact of (i) sequencing of items within and across measures, and (ii) grouping or separation of items and measures through labeling, pagination, or contiguous placement. Additionally, we investigate whether questionnaire length (i.e., the burden placed on respondents), different types of response scales (respondent- or stimulus-centered), or different types of samples (student versus non-student adults) moderate the impact of format on psychometric properties of measures of constructs.

In placing this research in perspective, research on the effects of a variety of factors on responses in survey research with its focus on estimating accurate means (e.g., Sudman, Bradburn, & Schwarz, 1996) provides a noteworthy parallel. This literature has systematically identified and studied a number of factors that influence responses to individual questions. In this paradigm, the emphasis has been on eliciting unbiased responses that do not deviate from the "true" value. However, such "additive error" is relatively less problematic in research, such as in many types of academic research, which emphasizes accurate estimates of *relationships* between variables rather than accurate estimates of absolute values per se (Groves, 1991). Relevant for such research is an understanding of methodological factors relating to format, such as labeling of constructs, on the *psychometric properties of measures, as well as on the relationships between constructs*. Relationships between constructs are typically not affected by some additive error, rather, a nuanced understanding of correlational systematic error is required (Viswanathan, 2005) — error that affects associations between variables.

Our research is in the tradition of methodological articles on such topics as common method variance (Cote & Buckley, 1988; Malhotra, Kim, & Patil, 2006; Podsakoff, Podsakoff, MacKenzie, & Lee, 2003) and the effects of measure design (e.g., measure development process), sample characteristics, and scale design on the reliability of rating scales (Churchill & Peter, 1984;

Weng, 2004). We differ, however, in our focus on how format factors, such as labeling of constructs or interspersion of items from different constructs, affect reliability and validity. Thus, our work lies at the middle of the continuum from examining common method variance at one end (i.e., at the unit of analysis of the overall method) and examining micro-level measure design or scale properties, such as response category anchors or number of response categories at the other end (i.e., at the level of response scales or scale development procedures as the unit of analysis). There is, of course, some overlap, between the format factors we study and research on either end of this continuum — for example, intermixing of items of constructs in the context of common method variance (Kline, 2000; Podsakoff et al., 2003) referred to here as interspersion, and time and location of measurement of constructs (Podsakoff et al., 2003) also addressed here. We also note that Bradlow and Fitzsimons (2001) represent one exception in having examined what we refer to as format factors of labeling and grouping, discussed subsequently.

This paper is organized as follows. Following a general articulation of measurement error, we discuss different methodological practices in terms of their assumed effect on types of measurement error. We report on a number of empirical studies. We introduce an application of multiple group confirmatory factor analysis to test whether these assumptions are borne out empirically, or whether measures are invariant across the different methodological practices. We then discuss five studies to test these assumptions, followed by a description of the analysis and results. Finally, we conclude with specific prescriptions for empirical research.

Measurement Error and Methodological Factors in Research

The implicit rationale used by researchers in choosing one format over another is that it ameliorates a potential measurement-related problem arising from measurement error. Here, we discuss different types of measurement error that could occur and how each might affect the observed relationship between variables. We then discuss methodological practices relating to format and discuss how such practices are assumed to impact measurement. We conclude this section by introducing multiple group confirmatory factor analysis as an approach to empirically test whether commonly used practices actually impact the psychometric properties of measures.

Measurement Error

We first present a review of the different types of measurement error by way of background for the rest of the paper. *Additive systematic error* occurs with deviations from the true score by a constant magnitude (e.g., extreme means) and may influence observed relationships when decreased item variance reduces covariation with other items (Viswanathan, 2005). For research with a focus on relationships between constructs, this type of systematic error may be less problematic when compared to contexts, such as opinion research, where the onus is on estimating absolute values (Groves, 1991).

With *correlational systematic error*, responses vary consistently and by different degrees beyond true differences in the measured construct (Viswanathan, 2005). Response categories, such as *very good* to *very bad*, may be used in consistent but different ways by different individuals,

wherein *very bad* is more or less negative for different respondents. Correlational systematic error may strengthen or weaken observed relationships (Nunnally, 1978).

Within-measure correlational systematic error arises between different items of the same construct, with examples including stronger observed relationships between items of a construct when using the same response format (Viswanathan, 2005). Halo error is an example wherein a global impression is employed to complete ratings on measures of distinct dimensions (Lance, LaPointe, & Stewart, 1994). A "halo" can be created by responses to one or two items, and lead to consistent responses to other items of the construct. Responses to later items in a scale may, therefore, be more polarized and consistent (Feldman & Lynch, 1988; Knowles, 1988; Simmons, Bickart, & Lynch, 1993) and more reliable, by following responses to earlier items (Knowles & Byers, 1996).

Across-measure correlational systematic error leads to inaccurate but consistent observed relationships, increasing or decreasing correlations (Viswanathan, 2005). Common method factors, such as placement of items of different measures on one page (Lennox & Dennis, 1994), represent examples of this type of error.

Methodological Practices and Measurement Error

Multiple practices are typically followed by researchers in the social sciences, in structuring their questionnaires. Items measuring different constructs can be interspersed — for example, Parameswaran, Barnett, Greenberg, Bellenger, and Robertson (1979) interspersed items from different domains of lifestyles; Szybillo, Binstok, and Buchanan (1979) interspersed items measuring attitude importance with other items; please also see Smith, Haugtvedt, and Petty (1994). Items measuring a construct can be placed contiguously with or without labels (e.g., Bradlow & Fitzsimons, 2001), or items measuring different constructs can be placed on different pages (e.g., Mittal 1995). Each of these practices may have an impact on correlational systematic error, leading to a potential effect on the estimates of relationships between constructs. For example, items that are used to test theories of relationships between constructs can be interspersed to mitigate inflation of correlations that could result from halo error. However, as these practices have received little systematic inquiry a clear theoretical rationale is not available for why researchers use one format over another. Thus, rather than express formal hypotheses about the effects of each condition, we present typical arguments or assumptions currently held in practice and in the literature.

We examined a total of eight conditions that varied on two methodological factors: (i) item sequencing within and across constructs, and (ii) separation or grouping of items and constructs through labeling, pagination, or contiguous placement. Treating the contiguous placement of items and measures (for multidimensional constructs) as the baseline condition, we then compared the effect of adding labels to measures, paginating measures, and interspersing (or resequencing) items and measures. We present a summary of the conditions in the Appendix A.

In Condition 1, referred to as "contiguous," items from each construct/dimension (the latter if the construct is multidimensional) are presented in proximity contiguously, a common practice and a useful baseline. Condition 2, referred to as "contiguous and labeled," is similar to Condition 1 with the addition of labels for measures/subscales of unidimensional constructs/dimensions of multidimensional constructs. In Condition 3, referred to as "contiguous and paginated," measures/subscales for each construct/dimension are presented on a different page, without labels. Thus,

Conditions 2 and 3 test the effects of labeling and pagination, respectively. Each of these approaches serves to provide a logical division between measures/subscales of dimensions of a construct and the constructs themselves, labeling more explicitly and pagination more subtly. Such division may lead to greater consistency among items within measures of specific constructs, or within subscales of specific dimensions of a construct, exploiting within-measure correlational systematic error. Stability reliability may also be enhanced, due to consistency over time among items. Because subscales of individual dimensions, rather than the measure of the overall multidimensional construct, are labeled, items are expected to have higher loadings on respective factors representing dimensions. Labeling may have stronger effects than pagination due to the explicit naming of constructs and dimensions. In this regard, Bradlow and Fitzsimons (2001) found that labeling and grouping (of items on a screen), each led to reduced variance within a subscale. In terms of relationships across measures of different constructs, labeling and pagination serve to separate measures of constructs (or dimensions of a construct); therefore, they may reduce observed correlations when compared to Condition 1.

In Condition 4, referred to as "interspersed," items from measures of different constructs were completely interspersed. Podsakoff et al. (2003) discusses the possibility of interspersion, referred to it as intermixing of items, which increases inter-construct correlation and reduces intra-construct correlation. Kline, Sulsky, and Rever-Moriyama (2000) include interspersion as a possible solution for reducing common method variance. Interspersion may detract from the halo effect within a measure of a construct when responses to later items are based on a general impression created by earlier items, likely reducing consistency across items within a measure representing a construct or a dimension. Interspersion can potentially cut both ways: it can serve to separate items within a construct, but can also create confusion and have the opposite effect of labeling or pagination. In this regard, the within-measure-of-a-construct halo effect serves to increase consistency of responses to items within a measure of a construct. Subsequent items in a construct are interpreted in light of earlier items. Researchers have shown increased reliability for later items in a construct (Knowles, 1988). However, interspersion also has the potential benefit of minimizing blurring across items from measures (subscales) representing different dimensions of a construct when compared to the contiguous condition, suggesting higher fit for multidimensional models with confirmatory factor analysis (CFA). Interspersion may also decrease observed relationships between measures of different constructs. This decrease when compared to Condition 1 may be greater than the decrease due to labeling or pagination.

In Condition 5, items were presented contiguously within measures (subscales) representing dimensions of the same construct as in Condition 1, but the sequence of measures/dimensions across the questionnaire was different. For example, two related measures of different constructs may be contiguous here but non-contiguous in Condition 1. Moreover, Condition 5 was sequenced so that no two measures (subscales) of dimensions of the same construct were contiguously placed, thereby testing the degree to which correlations between measures (subscales) representing dimensions of the same construct are influenced by contiguous placement by comparing to other conditions. Condition 6 was similar to Condition 5 with the addition of labeling which could lead to stronger relationships across related constructs when compared to Condition 5.

In Condition 7, referred to as "resequenced contiguous," items within subscales/measures of dimensions or constructs were sequenced differently when compared to the sequencing during their validation, but items of measures/subscales representing each construct or dimension were

still presented contiguously. The aim here was to assess the extent to which validated measures should be used with item sequencing identical to those at validation. The sequencing used at validation capitalizes on within-measure correlational systematic error due to sources such as a halo effect in responses. A different, unvalidated sequencing of items may detract from this effect. Condition 8, referred to as "resequenced, contiguous, and labeled," was similar to Condition 7 with the addition of labeling of measures (subscales) representing dimensions. This was designed to assess whether labeling would help overcome any detrimental effect due to resequencing of items.

Assessing the Impact of Questionnaire Format

A natural way to test whether various formatting conditions impact the psychometric properties of measures is multiple group confirmatory factor analysis (MGCFA; Steenkamp & Baumgartner, 1998). This procedure allows us to test whether measures are invariant across the conditions previously identified. Consider a vector of k items for the c^{th} condition. The data model, x_c, is given by the following formula:

$$x_c = \tau_c + \Lambda_c\xi_c + \delta_c \quad (1)$$

where Λ_c is the matrix of factor loadings relating the vector of latent variables ξ_c to the observed items, τ_c is a vector of intercepts, and δ_c is a vector of measurement errors. Given a fixed condition, the covariance structure of the data Σ_c is computed as follows:

$$\Sigma_c = \Lambda_c\Phi_c\Lambda_c^T + \Theta_c \quad (2)$$

where Φ_c is the covariance matrix of the latent variables, Λ_c^T is the transpose of the matrix Λ_c and Θ_c is the covariance matrix of the error terms. Testing for measurement invariance essentially involves constraining specific parameters in equations (1) and (2) to be the same across conditions. Adopting the process suggested by Steenkamp and Baumgartner (1998), we impose increasingly more strict parameter constraints, in order to assess whether the psychometric properties of measures are invariant to the format used for collecting the data (Appendix B). Byrne, Shavelson, and Muthén (1989) suggest that there are two primary approaches to assessing invariance — measurement invariance and structural invariance, both of which encompass different components. The level of invariance required for a given model depends on the aims of the research. Measurement invariance encompasses configural, metric, scalar, and error variance invariance. *Configural* invariance exists when the patterns of loadings in the factor loading matrices (Λ_1, Λ_2, …, Λ_c) indicate that the items in each condition load on to the same factor representing an underlying dimension. At the most basic level, the equivalence of factor loading structures implies that a construct can be conceptualized in the same way in terms of underlying dimensions under different conditions. It is tested by constraining factor loading structure to be identical across conditions, and comparing this constrained model to a benchmark model with the factor loading structures not constrained to be the same. More strict tests are for *metric* invariance, which occurs when the factor loadings are the same across conditions (i.e., $\Lambda_1 = \Lambda_2 = \ldots = \Lambda_c$), and for *scalar* invariance, which occurs when the intercepts are the same across conditions (i.e., $\tau_1 = \tau_2 = \ldots = \tau_c$). Once this level of invariance has been established, mean scores can be compared across conditions. *Error variance invariance* occurs when there is an approximately equal amount of measurement error across conditions (i.e., $\Theta_1 = \Theta_2 = \ldots = \Theta_c$), and along with metric invariance, suggests that a measure is equally reliable across conditions.

Structural invariance consists of factor variance and factor covariance invariance, and implies that the covariation or correlation between constructs is equivalent across conditions. *Factor variance invariance* is when the variance of the latent constructs is equal across conditions (i.e., $\phi_{i1} = \phi_{i2} = \ldots = \phi_{ic}$), whereas *factor covariance invariance* is when the latent constructs share the same covariance structure (i.e., $\phi_{ij1} = \phi_{ij2} = \ldots = \phi_{ijc}$ $i \neq j$).

The strength of MGCFA for our study is that it allows us to formally test whether questionnaire formatting impacts measures. MGCFA also allows us to formulate tests of format equivalence that are specifically appropriate for the different objectives of the research (Steenkamp & Baumgartner, 1998). When research aims to explore the structure of a construct and how various items relate to that construct, then lower levels of invariance such as configural and metric invariance may be sufficient. When the aim of research is to examine differences in mean scores between groups of individuals, then it may also be necessary to show scalar invariance. If, on the other hand, the aim of research is to test the relationships between constructs, as is often the case, higher levels of invariance are required. Researchers may wish to uncover the true scores of respondents without systematic bias. If measures depend on the format of the questionnaire rather than the underlying phenomena, different conclusions would be drawn, depending on the format used for collecting the data. The formal testing procedure is also particularly suited to our study since there is little theory to guide an understanding of the impact of different formats. The procedure used to test for invariance begins by allowing each condition to have a different model, and progressively applying each of the six constraints. In other words, in each of the eight conditions, six different invariance models were applied to the data and compared to establish which best explained the data. This allows us to identify where differences between conditions arise.

Following recommendations from the literature (e.g., Cheung & Rensvold, 2002; Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000), we used standard statistics of goodness-of-fit for testing the six different invariance models. Whereas a standard chi-squared test of difference can be used to compare different models, such a test is extremely sensitive as the sample size increases, and is criticized as an impractical test of model fit (Cheung & Rensvold, 2002). Thus, we use root mean squared error of approximation (RMSEA), comparative fit index (CFI), Tucker-Lewis index (TLI), and Bayesian information criterion (BIC) to determine whether a scale is invariant, with a far greater reliance on BIC as the most appropriate way to assess relative model fit.

Empirical Studies

We conducted five different studies to examine the impact of questionnaire formats on the psychometric properties of measures. Study 1 comprehensively tested all eight conditions across thirteen measures, using student samples. In Study 2, we examined whether an increase in cognitive demands on the respondent changes the pattern of results. In Study 3, we used stimulus-centered measures instead of respondent-centered measures that were used in Studies 1 and 2. In Study 4, we used stimulus-centered measures, but the correlation matrix for analysis was computed across stimuli, rather than across respondents (as in Study 3). Finally, in Study 5, we examined whether a non-student sample changes the pattern of results. Together, these five studies, eight conditions, and multitude of measures represent a large number of ways in which we sought to examine the implicit assumptions made by researchers about questionnaire formats. The five studies are linked

in testing different conditions in Study 1 with a student sample for respondent-centered measures, and extending to conditions of higher cognitive demand (Study 2), stimulus-centered measures (Studies 3 and 4), and a non-student sample (Study 5). Thus, we test our predictions for different study conditions, different types of measures, and different samples.

Study 1

Method

We tested questionnaire formats corresponding to the eight conditions (Appendix A), described previously, using a number of previously validated scales. The scales presented in Table 1 include coupon proneness, value consciousness, sale proneness, and price consciousness (Lichtenstein, Ridgway, & Netemeyer, 1993), consumer independent judgment-making and consumer novelty (Manning, Bearden, & Madden, 1995), consumer susceptibility to interpersonal influence (Bearden, Netemeyer, & Teel, 1989), material values — defining success, acquisition centrality, and pursuit of happiness (Richins & Dawson, 1992), consumer ethnocentrism (Shimp & Sharma, 1987), and need for cognition (Cacioppo, Petty, & Kao, 1984). We selected scales for which the response formats were similar to facilitate the presentation and analysis of various conditions described below. We used a test-retest approach with a 6-week interval. Data were collected from undergraduate students enrolled in introductory business classes at a large university with sample sizes in conditions ranging from 160-180. Students were given extra credit for participation. Each session took about 10-15 minutes to complete. Because of the sample sizes involved and the test-retest element, typically, data collection on one condition was completed during a semester.

TABLE 1
Scales and examples of items in studies

	Number of items	Sample items	Comment
	Studies 1 and 2		
Consumer independent judgment making (Manning et al., 1995)			
Consumer independent judgment making	6	Prior to purchasing a new brand, I prefer to consult a friend that has experience with the new brand	
Consumer novelty	8	I often seek out information about new products and brands	
Material values (Richins & Dawson, 1992)			
Defining success	6	I admire people who own expensive homes, cars, and clothes	
Acquisition centrality	7	I like a lot of luxury in my life	
Pursuit of happiness	5	My life would be better if I owned certain things I don't have	

(Table 1 continues)

Table 1 (continued)

	Number of items	Sample items	Comment
Value consciousness (Lichtenstein et al., 1993)			Also in Study 5
Value consciousness	7	I am very concerned about low prices, but I am equally concerned about product quality	
Price consciousness	5	I will grocery shop at more than one store to take advantage of low prices	
Coupon proneness	8	Redeeming coupons makes me feel good	
Sale proneness	5	If a product is on sale, that can be a reason for me to buy it	
Consumer ethnocentrism (Shimp & Sharma, 1987)	17	American people should always buy American-made products instead of imports	
Need for cognition (Cacioppo et al., 1984)	18	I find satisfaction in deliberating hard and for long hours	
Consumer susceptibility to interpersonal influence (Bearden et al., 1989)	12	I often consult other people to help choose the best alternative available from a product class	
		Study 3	
Service quality (Parasuraman, Zeithaml, & Berry, 1988)			
Tangibility	4	McDonald's has up-to-date equipment	
Reliability	5	What McDonald's promises to do something by a certain time, it does so	Also in Study 4
Responsiveness	4	You do not receive prompt service from McDonald's employees (R)	Also in Study 4
Assurance	4	You can trust the employees of McDonald's	
Empathy	5	McDonald's does not give you individual attention (R)	
Affective response to advertising (Holbrook & Batra, 1987)			
Pleasure	9	I felt grateful	
Arousal	9	I felt excited	
Domination	9	I felt afraid	
Endorser evaluation (Ohanian, 1990)		Please rate Michael Jordan as a celebrity endorser for Wheaties Cereal on the scales below	
Attractiveness	5	unattractive --- attractive	
Trustworthiness	5	undependable --- dependable	
Expertise	5	not an expert --- expert	

(Table 1 continues)

Table 1 (continued)

	Number of items	Sample items	Comment
Retail service quality (Dabholkar, Thorpe, & Rentz, 1996)			
Physical aspect	6	This store has modern-looking equipment	
Reliability	5	When this store promises to do something by a certain time, it will do so	
Personal interaction	9	Employees in this store give prompt service to customers	
Problem solving	3	This store willingly handles returns and exchanges	
Policy	5	This store offers high quality merchandise	
Perceived value			Also in Study 5
Perceived quality (Grewal, Monroe, & Krishnan, 1998)	3	The laptop appears to be of good quality	
Perceived transaction value (Grewal et al., 1998)	3	I would get a lot of pleasure knowing that I would save money at this reduced sale price	
Perceived acquisition value (Grewal et al., 1998)	9	This laptop would be a worthwhile acquisition because it would help me use it at a reasonable price	
Perceived sacrifice (Teas & Agarwal, 2000)	2	If I purchased the laptop for the indicated price, I would not be able to purchase some other products I would like to purchase now	
Perceived price (Zeithaml, 1988)	2	The price of this laptop is high	
Involvement (McQuarrie & Munson, 1986)			
Importance	5	The product is… important --- unimportant	
Interest	5	The product is… unexciting --- exciting	

Note. R = reverse coded.

Analysis and Results

A number of different types of analyses were conducted on the data to examine means and variances for subscales representing dimensions for each measure, item-to-total correlations and coefficient alphas for subscales representing dimensions for each measure, item level and overall test-retest correlations, exploratory and confirmatory factor analyses, and correlations across sub-scales representing dimensions and across measures of different constructs.

Means and Reliability

Means at test and at retest, as well as differences between means at test versus retest, did not suggest any consistent pattern across conditions. An examination of test-retest correlations across the conditions for each scale suggests no striking pattern across conditions that held for all measures (means of test-retest correlations across scales for each condition ranged from .68 to .75). Similarly, an examination of item level test-retest correlations across conditions suggests no striking pattern at this broad level (means of item level test-retest correlations across scales for each condition ranged from .49 to .57).

Similar results were found for item-to-total correlations and coefficient alpha. Means of item-to-total correlations across scales for each condition ranged from .54 to .76, with six conditions in the narrow band of .71-.76 and the continuous paginated condition and the measures resequenced labeled conditions in the 0.5 range. Means of coefficient alpha across scales for each condition ranged from .78 to .88 (see Table 2). Even interspersed versus labeling conditions did not lead to consistent differences across conditions (.84 versus .86, respectively), contrary to the assumption that the interspersed condition would lead to low internal consistency, at least for this particular sample and duration of administration. Lower internal consistency was found for some measures of scales in some conditions, such as material values — acquisition centrality scale in the paginated condition (coefficient alpha of .61 at test, for example). Similarly, item-to-total correlations were lower for some items of measures in some conditions. However, there was no dominant effect that held across conditions.

In order to establish whether these differences are statistically significant, we sought to conduct statistical tests to compare the variety of estimates of reliability across conditions. We note, however, that many of the estimates do not have well-defined distributions. Therefore, we use bootstrapping to construct 95% confidence intervals for all estimated statistics, described in Appendix C. The 95% confidence intervals were overlapping across all conditions and scales.

Factor Analyses

The next step in the analysis was to conduct factor analyses to examine the factor structure of various measures under different conditions. It is quite possible that differences across format conditions do not emerge in terms of means, internal consistency reliability, and stability reliability; yet, emerge in terms of blurring of items across subscales representing dimensions within and across measures. Indeed, internal consistency reliability assumes unidimensionality, which can be tested explicitly by factor analysis.

We began with exploratory factor analysis (EFA) and followed with confirmatory factor analysis (CFA), using appropriate multidimensional versus unidimensional models for measures of multidimensional and unidimensional constructs, respectively. Using CFAs, such an approach led to unsatisfactory levels of fit in all conditions. In an effort to boost to satisfactory levels at least in some conditions, we employed an approach involving parceling of items. We followed the rationale and recommendations for parceling suggested by Bagozzi and Edwards (1998, pp. 79-82) and Little, Cunningham, Golan, and Widaman (2002). We first used the results of EFA to select scales for further analyses where item loadings were high on appropriate subscales representing

TPM Vol. 24, No. 4, December 2017
465-498

TABLE 2
Study 1 – Test-retest correlations and coefficient alphas

Test-retest correlation across scales	Contiguous	Contiguous and labeled	Contiguous and paginated	Interspersed items	Resequenced measures	Resequenced and labeled measures	Resequenced items, contiguous measures	Resequenced items, contiguous and labeled measures
Consumer independent judgment making	.58	.50	.64	.54	.46	.44	.59	.60
Consumer novelty	.63	.66	.70	.70	.71	.57	.66	.68
Material values – Defining success	.79	.83	.75	.65	.82	.72	.83	.72
Material values – Acquisition centrality	.77	.73	.77	.77	.84	.70	.80	.73
Material values – Pursuit of happiness	.63	.74	.57	.76	.83	.72	.81	.78
Value consciousness	.69	.82	.59	.80	.77	.69	.78	.66
Price consciousness	.67	.72	.64	.61	.70	.69	.69	.68
Coupon proneness	.81	.73	.76	.80	.53	.67	.80	.62
Sale proneness	.70	.70	.68	.54	.68	.65	.76	.69
Consumer ethnocentrism	.78	.78	.78	.78	.81	.71	.78	.79
Need for cognition	.86	.85	.86	.79	.84	.78	.81	.82
Consumer susceptibility to interpersonal influence	.71	.75	.69	.73	.73	.75	.75	.70
Mean test-retest correlation	.72	.73	.70	.70	.73	.68	.75	.71

Coefficient alpha	test	retest	test	retest	test	retest	test	retest	Test	retest	test	retest	test	retest	test	retest
Consumer independent judgment making	.85	.91	.86	.89	.78	.80	.81	.86	.90	.89	.79	.63	.85	.89	.86	.89
Consumer novelty	.89	.91	.89	.91	.83	.85	.88	.90	.92	.91	.86	.69	.88	.92	.89	.88
Material values – Defining success	.85	.89	.84	.84	.73	.77	.80	.54	.84	.88	.75	.76	.86	.87	.83	.84
Material values – Acquisition centrality	.81	.82	.77	.76	.61	.65	.75	.82	.81	.81	.64	.67	.76	.82	.74	.73
Material values – Pursuit of happiness	.76	.75	.75	.84	.65	.43	.83	.83	.83	.84	.71	.64	.77	.84	.83	.84
Value consciousness	.84	.89	.87	.87	.84	.86	.82	.85	.88	.85	.88	.88	.85	.86	.85	.87
Price consciousness	.78	.85	.80	.80	.71	.66	.76	.83	.82	.83	.99	.71	.82	.83	.80	.84

(Table 2 continues)

Table 2 (continued)

Coefficient alpha	test	retest	test	retest	test	retest	test	retest	test	retest	test	retest	test	retest	test	retest
Coupon proneness	.90	.93	.91	.91	.88	.89	.89	.90	.84	.56	.84	.84	.90	.90	.86	.87
Sale proneness	.79	.82	.84	.78	.84	.83	.81	.83	.85	.84	.82	.82	.83	.86	.83	.83
Consumer ethnocentrism	.96	.97	.96	.97	.95	.95	.96	.96	.93	.96	.97	.95	.97	.97	.97	.97
Need for cognition	.90	.92	.90	.92	.84	.81	.89	.92	.88	.93	.79	.84	.90	.84	.89	.92
Consumer susceptibility to interpersonal influence	.78	.91	.87	.90	.89	.89	.86	.88	.89	.91	.90	.91	.87	.89	.89	.90
Mean coefficient alpha	.84	.88	.86	.87	.79	.78	.84	.84	.87	.85	.81	.78	.86	.87	.86	.87

dimensions (≥ .40) and not high on inappropriate subscales representing dimensions (< .25). By this criterion, we excluded the need for cognition scale and the susceptibility to interpersonal influence scale from further analysis, which did not yield interpretable factor structures according to specifications for the original scales. We then used a parceling approach that was not based on idiosyncratic item content and that did not vary across scales, essentially using generic rules for combining contiguous items into different parcels. For subscales representing dimensions with relatively few items, two parcels for each subscale representing a dimension were used comprised of odd versus even numbered items, respectively. For the longer ethnocentrism scale, we used four parcels, each consisting of every fourth item with a different starting point (e.g., Item 1, 5, 9 … in the first parcel; 2, 6 … in parcel 2, etc.).

Having established measurement models, we conducted MGCFA to test for invariance. For each scale, we estimated three sets of MGCFAs — a model combining test and retest, a model for test alone, and a model for retest alone. Results are presented in Table 3. We discuss the first case for the "test" set of the Materialism scale. We began by testing for configural invariance, which is when the factor loading structure is the same across conditions. This is considered appropriate since the model has acceptable fit (RMSEA = .071; CFI = .991; TLI = .970). Testing for metric invariance, the results support the conclusion that the factor loadings are invariant across conditions (RMSEA = .051; CFI = .993; TLI = .985); and we find similar support for scalar invariance (RMSEA = .074; CFI = .979; TLI = .968). This finding implies that if the aim of a study was to compare means, it would not matter which of the questionnaire formats were implemented. We find support for factor variance invariance (RMSEA = .073; CFI = .963; TLI = .968) and for factor covariance (RMSEA = .078; CFI = .967; TLI = .964). These results suggest that choice of questionnaire format will not lead to different conclusions when examining relationships between constructs. We also find support for error variance invariance (RMSEA = .074; CFI = .951; TLI = .968). Along with metric invariance, this result suggests that the eight conditions are similarly reliable, reinforcing the previous findings. Further support for measurement invariance is provided by the BIC, which is minimized at the final level of invariance (35247.34).

We obtained similar results for the remainder of the measures. 12 out of 12 MGCFAs (three sets for each of the four scales), achieved all six levels of invariance using goodness-of-fit criteria. Furthermore, in 11 of the cases the BIC was minimized in the most restrictive invariance model (all six levels; the one exception was the retest set of the consumer independent judgment and decision making scale, which did achieve the first five levels of invariance), suggesting consistence in the overall pattern of results.

Cross-Dimensional Relationships

To further examine the impact of questionnaire format on the relationships between dimensions within a construct, we computed the correlations among the set of dimensions or constructs within each of three multidimensional construct measures or subscales (Table 4A). Again, no discernible differences were found across conditions, with bootstrapping yielding an overwhelming pattern of overlapping confidence intervals. Our earlier discussion of the motivations for different formats argued for the possibility of labeling separating each distinct measure (subscale) of individual dimensions in respondents' minds, thus lowering the observed relationship but

TABLE 3
Study 1 – Multiple group confirmatory factor analysis

	Invariance	Test				Retest				Test-Retest			
Measure	Type	RMSEA	CFI	TLI	BIC	RMSEA	CFI	TLI	BIC	RMSEA	CFI	TLI	BIC
Material values	Configural	.071	.991	.970	35764.12	.066	.993	.978	35584.09	.128	.941	.879	68652.81
	Metric	.051	.993	.985	35673.71	.078	.985	.969	35518.58	.121	.939	.892	68496.23
	Scalar	.074	.979	.968	35617.79	.093	.972	.956	35467.68	.121	.930	.892	68391.41
	Factor covariance	.078	.967	.964	35514.32	.094	.959	.955	35371.82	.105	.924	.918	67795.99
	Factor variance	.073	.963	.968	35393.30	.094	.947	.955	35272.83	.104	.916	.920	67601.92
	Error variance	.074	.951	.968	35247.34	.092	.934	.957	35133.07	.101	.909	.925	67308.49
Consumer ethnocentrism	Configural	.233	.978	.935	25865.55	.222	.981	.942	25880.06	.174	.949	.925	49958.80
	Metric	.163	.975	.968	25755.90	.155	.978	.971	25766.78	.156	.948	.940	49720.51
	Scalar	.162	.962	.968	25715.01	.133	.975	.979	25658.69	.151	.941	.944	49567.87
	Factor covariance	.162	.962	.968	25715.01	.133	.975	.979	25658.69	.149	.940	.945	49528.65
	Factor variance	.157	.960	.970	25685.28	.127	.974	.981	25620.03	.148	.938	.946	49470.30
	Error variance	.156	.943	.971	25622.15	.117	.968	.984	25485.14	.144	.929	.949	49267.45
Consumer independent judgment making	Configural	.101	.995	.968	26989.69	.196	.982	.890	27011.37	.096	.975	.949	52517.70
	Metric	.167	.960	.913	26991.33	.216	.939	.867	27043.86	.086	.974	.959	52346.44
	Scalar	.057	.992	.990	26822.22	.174	.935	.914	26967.61	.090	.967	.956	52224.91
	Factor covariance	.057	.991	.990	26782.22	.161	.934	.926	26928.33	.082	.965	.963	51974.67
	Factor variance	.070	.982	.985	26718.63	.081	.978	.981	26718.04	.088	.955	.958	51871.22
	Error variance	.104	.940	.966	26652.03	.140	.902	.944	26758.98	.101	.926	.944	51715.86
Value consciousness	Configural	.068	.986	.967	49136.09	.067	.988	.971	48046.10	.111	.930	.875	94741.07
	Metric	.073	.979	.962	49032.59	.064	.985	.974	47925.79	.108	.928	.883	94503.70
	Scalar	.108	.943	.916	49045.04	.094	.962	.944	47900.31	.116	.908	.864	94483.88
	Factor covariance	.094	.942	.937	48789.34	.086	.956	.953	47664.39	.102	.901	.895	93351.34
	Factor variance	.092	.933	.939	48649.97	.092	.942	.947	47555.58	.102	.894	.896	93094.90
	Error variance	.092	.918	.939	48449.59	.089	.932	.949	47338.54	.101	.884	.898	92693.79

Note. RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker-Lewis index; BIC = Bayesian information criterion.

TPM Vol. 24, No. 4, December 2017
465–498

TABLE 4A
Average cross-dimensional correlations

		Contiguous	Contiguous and labeled	Contiguous and paginated	Interspersed items	Resequenced measures	Resequenced and labeled measures	Resequenced items, contiguous measures	Resequenced items, contiguous and labeled measures
		$N = 183$	$N = 165$	$N = 160$	$N = 180$	$N = 161$	$N = 170$	$N = 170$	$N = 159$
Study 1 – test	Consumer novelty	−.36	−.13	−.35	−.15	−.25	−.33	−.20	−.27
	Materialism	.46	.38	.39	.44	.58	.37	.55	.47
	Value consciousness	.34	.31	.31	.43	.48	.39	.45	.43
Study 1 – retest	Consumer novelty	−.27	−.32	−.48	−.30	−.33	.00	−.29	−.25
	Materialism	.51	.42	.38	.40	.60	.37	.57	.56
	Value consciousness	.39	.40	.39	.44	.45	.39	.48	.51
Study 2	Consumer novelty	−.10	.00		−.29				
	Materialism	.46	.43		.56				
	Value consciousness	.44	.33		.50				
Study 3	Service quality	.48	.38		.52				
	Affect	.31	.27		.22				
	Endorsement	.48	.56		.74				
	Retail service quality	.46	.49		.46				
	Computer evaluation	.25	.31		.35				
	Involvement	.62	.48		.65				
Study 5	Value consciousness		−.02		.10				
	Computer evaluation		.32		.29				

not with contiguous or interspersed conditions. Whereas interspersion serves to separate items, labeling may serve to group items within a subscale representing a dimension and distinguish them from items from other subscales representing other dimensions, and attenuate relationships. However, the results did not support these assumptions, with no consistent pattern indicating that a particular format impacts the strength or direction of relationships between dimensions. The measures resequenced conditions where subscales representing dimensions were not presented contiguously did not differ from the other conditions.

Cross-Construct Relationships

Next, correlations across constructs were examined. In particular, the relationship between consumer independent judgment making and susceptibility to interpersonal influence was examined. We expected a negative relationship between these two constructs at a conceptual level, independence being likely to be negatively related to susceptibility to influence. All the items in the former measure related to seeking advice or consulting friends before purchase. Several items in the susceptibility to interpersonal influence relate to seeking advice from friends. In the measures resequenced conditions, consumer independent judgment making and susceptibility to interpersonal influence were presented contiguously. Whereas correlations were statistically significant across most conditions, there were no consistent differences across conditions (Table 4B), as assessed with a *z*-test. For example, the difference between correlations in the contiguous and contiguous labeled for the "test" set is significant (–.37 vs. –.03, $p < 0.01$), whereas the same difference becomes insignificant for the "retest" set (–.36 vs. –.37). A similar inconsistent pattern was found with correlations of susceptibility to interpersonal influence and consumer novelty.

Discussion

The overwhelmingly striking and perhaps surprising pattern from Study 1 was the invariance of measures across conditions for a student sample. This implies that one could conduct theory testing by comparing subsets of individuals, or examining relationships between variables, and regardless of the format of the questionnaire, the substantive conclusions would be similar. Additionally, it appears that, irrespective of condition, measures performed comparably across conditions in terms of test-retest reliability and internal consistency reliability. These results indicate that our study does not provide evidence of the impact of questionnaire format on the psychometric properties of the scale. Therefore, the rationale leading to predictions of the impact of format differences did not find support in our data.

In interpreting these findings, even the interspersed condition was not different from the labeled condition. In other words, in one condition, items are disguised by interspersing with other items, and in the other, items are labeled under a subscale representing a dimension or a scale representing a unidimensional construct and presented together. On the one hand, this pattern points to the quality of the measures used. However, this pattern may also be a consequence of the administration in Study 1, which involved completion of these measures by students for a length

TABLE 4B
Average cross-construct correlations

		Contiguous	Contiguous and labeled	Contiguous and paginated	Interspersed items	Resequenced measures	Resequenced and labeled measures	Resequenced items, contiguous measures	Resequenced items, contiguous and labeled measures
		N = 183	*N* = 165	*N* = 160	*N* = 180	*N* = 161	*N* = 170	*N* = 170	*N* = 159
Study 1 – test	SUS-CON	–.38	−.03	−.44	−.37	−.20	−.20	−.15	−.27
	SUS-NOV	.28	.35	.28	.30	.29	.31	.14	.09
Study 1 – retest	SUS-CON	−.36	−.37	−.37	−.41	−.20	−.08	−.32	−.33
	SUS-NOV	.38	.38	.25	.26	.42	.19	.27	.21
Study 2	SUS-CON	−.50	−.20		−.36				
	SUS-NOV	.38	.12		.22				
Study 3	PTRAN-PPRI	.27	.32		.53				
	PACV-PPRI	.35	.54		.68				
	PTRAN-PQUAL	.54	.39		.53				
	PACV-PQUAL	.66	.52		.73				
	PTRAN-PSAC	−.17	.06		−.11				
	PACV-PSAC	−.16	.09		.01				
	PPRI-PQUAL	.16	.19		.35				
	PPRI-PSAC	.26	.27		.16				
Study 5	PTRAN-PPRI		.09		.32				
	PACV-PPRI		−.06		.07				
	PTRAN-PQUAL		.62		.55				
	PACV-PQUAL		.67		.72				
	PTRAN-PSAC		.40		.20				
	PACV-PSAC		.25		−.05				
	PPRI-PQUAL		.18		.15				
	PPRI-PSAC		.11		.39				

Note. SUS-CON = Consumer susceptibility to interpersonal influence-Consumer independent judgment making; SUS-NOV = Consumer susceptibility to interpersonal influence-Consumer novelty; PTRAN-PPRI = Perceived transaction value-Perceived price; PACV-PPRI = Perceived acquisition value-Perceived price; PTRAN-PQUAL = Perceived transaction value-Perceived quality; PACV-PQUAL = Perceived acquisition value-Perceived quality; PTRAN-PSAC = Perceived transaction value-Perceived sacrifice; PACV-PSAC = Perceived acquisition value-Perceived sacrifice; PPRI-PQUAL = Perceived price-Perceived quality; PPRI-PSAC = Perceived price-Perceived sacrifice.

of time that was typically between 10 and 15 minutes. In other words, the demands on the respondents were not burdensome in terms of administration procedures. We use Study 2 to examine whether a greater burden on respondents might generate a different pattern of results.

Study 2

Method

To disentangle the effect of the overall cognitive demands of the administration and examine its role as a potential moderating factor, a second study was conducted. A student sample completed the same questionnaire used in Study 1 as a part of data collection of approximately three times the duration (about 45 minutes). The relationship between questionnaire length and response quality has been investigated in the methodological literature (Burchell & Marsh, 1992; Herzog & Bachman, 1981; Krosnick 1999), and recommendations on this issue are common in survey research. Students were assigned to one of three conditions considered the most distinctly different: the contiguous, the labeled, and the interspersed conditions (Appendix A). Whereas the labeled provides a sharp logical division, the interspersed is diametrically opposite; together these two conditions represent the most distinctly different conditions. The contiguous condition provides a moderate baseline for comparison. There was a test phase but no retest phase in this study. Students completed the same measures as in Study 1, but as part of a larger study that typically took about 45 minutes to complete. The questionnaire of interest for this study took about 15 minutes, as in Study 1, and was typically administered at the middle or the end of the administration of the larger study.[1] The sample size ranged from 129 to 131 for each condition.

Analysis and Results

There were no sizable differences in means or standard deviations between the three conditions. In terms of internal consistency, coefficient alphas were similar in magnitude across conditions (.81 to .84). In some instances, the interspersed condition had lower item-to-total correlations (e.g., value consciousness: .80 for interspersed vs. .86 for labeled and .89 for contiguous). However, this was an exception rather than evidence for an overwhelming or a consistent pattern. When compared to Study 1, where the cognitive demands were lower, lower coefficient alphas were found only for the need for cognition scale. Once again, most bootstrapped 95% confidence intervals were overlapping, and where there was no overlap, there was no pattern.

Using procedures identical to those in Study 1, items were parceled for each of the subscales representing dimensions. There were also no consistent differences across the conditions in terms of MGCF (Table 5). As in Study 1, the CFI and the TLI supported invariance across conditions, while the BIC preferred the model with all six levels of invariance imposed. The only exception was for the consumer judgment scale, which had poor fit across all the conditions.

Another set of findings related to relationships across dimensions within a construct. Correlations were computed among the set of dimensions or constructs within each of three multidimensional measures. No consistent difference emerged across conditions (Table 4A). Correlations across

constructs were also examined and did not lead to any striking differences across conditions (Table 4B). These results indicate that our data do not provide evidence of differences across conditions.

TABLE 5
Study 2 – Multiple group confirmatory factor analysis

Measure	Invariance type	RMSEA	CFI	TLI	BIC
Material values	Configural	.000	1.000	1.006	10057.99
	Metric	.011	1.000	.999	10039.58
	Scalar	.040	.995	.991	10023.06
	Factor covariance	.018	.999	.998	9990.703
	Factor variance	.020	.998	.998	9961.434
	Error variance	.000	1.000	1.001	9919.028
Consumer ethnocentrism	Configural	.137	.992	.975	7884.139
	Metric	.099	.991	.987	7855.016
	Scalar	.112	.984	.984	7839.252
	Factor covariance	.112	.984	.984	7839.252
	Factor variance	.106	.984	.985	7829.319
	Error variance	.084	.986	.991	7785.621
Consumer independent judgment making	Configural	.344	.945	.668	8894.712
	Metric	.237	.939	.842	8879.913
	Scalar	.187	.940	.902	8858.877
	Factor covariance	.180	.934	.909	8853.658
	Factor variance	.171	.922	.918	8843.898
	Error variance	.150	.912	.936	8812.934
Value consciousness	Configural	.058	.991	.979	15593.56
	Metric	.069	.985	.971	15572.28
	Scalar	.084	.974	.958	15557.78
	Factor covariance	.067	.978	.973	15491.76
	Factor variance	.085	.960	.957	15477.55
	Error variance	.094	.940	.947	15444.44

Note. RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker-Lewis index; BIC = Bayesian information criterion.

Discussion

The results of Study 2 suggest that cognitive load arising from the duration of data collection procedures did not contribute to the lack of differences observed in Study 1. The same pattern of results in terms of measurement invariance, cross-dimensional relationships, and cross-construct relationships were observed as in Study 1. It appears that, irrespective of a diverse set of conditions, student respondents are able to respond appropriately based on the content of items, even when the duration of data collection is extended. Compared to others, students may be highly motivated to fill-out questionnaires, more familiar with, and practiced in filling out questionnaires, and more cognitively complex or skilled in processing abstract concepts often comprising questionnaires.

Study 3

Study 3 was conducted to examine an important factor that may moderate the results, the nature of the measures on which data are collected. The measures used in earlier studies are respondent-centered, relating to individual differences, rather than stimulus-centered or relating to characteristics of stimuli (Cox, 1980). The nature of the measures possibly overwhelms the effects of format factors. In particular, respondents may be more certain and knowledgeable about traits and characteristics pertaining to themselves. Literature from a number of areas of research including self-concepts, self-referencing, and autobiographical memory (e.g., Krishnamurthy & Sujan, 1999) supports this line of reasoning, arising from the highly-organized memory structure of the self (Greenwald & Banaji, 1989) and leading to advantages in elaboration of incoming information and memory. When using validated scales with items relating to the self, student respondents are perhaps able to complete questionnaires based on item content, minimizing the effect of format factors. Therefore, a study was designed using stimulus-centered scales under the three conditions employed in Study 2 (Appendix A). The rationale for choosing these three conditions was discussed earlier; that is, comparing the most distinctly different labeling and interspersed conditions along with the baseline contiguous condition. Because each set of measures relating to a construct pertained to a different stimulus, interspersion was carried out between items from subscales of different dimensions of a construct rather than across constructs.

Method

A range of stimulus-centered scales was used in the study including measures of multidimensional constructs: service quality (Parasuraman et al., 1988) for McDonalds fast food restaurants with five dimensions and associated subscales — tangibility (four items), reliability (five items), responsiveness (four items), assurance (four items), and empathy (five items); affective response to an ad (Holbrook & Batra, 1987) for a health club with three dimensions and associated subscales — pleasure (six items), arousal (six items), and dominance (six items); perceived expertise (five items), attractiveness (five items), and trustworthiness (five items) of a celebrity endorser (Michael Jordan on a Wheaties cereal box; Ohanian, 1990); retail service quality (Dabholkar et al., 1996) of Walmart stores with five dimensions and associated subscales — physical aspects (six items), reliability (five items), personal interaction (nine items), problem solving (three items), and policy (five items); evaluation of a computer based on a picture and description on perceived quality (three items; Grewal et al., 1998), perceived transaction value (three items; Grewal et al., 1998), perceived acquisition value (nine items; Grewal et al., 1998), perceived sacrifice (two items; Teas & Agarwal, 2000), and perceived price (two items adapted from multiple sources; construct discussed in Zeithaml, 1988, and other literature); involvement (McQuarrie & Munson, 1986) in a smart phone based on a picture and description with two dimensions and associated subscales — importance (five items) and interest (five items). Stimuli were generally selected to be moderately positive to allow for variation on the scales. The sample was again made up of students, with the sample size ranging from 127 to 131 for each condition.

Analysis and Results

Means and standard deviations were comparable across conditions. The interspersed condition had slightly lower mean coefficient alphas across scales (.75) when compared to the contiguous and contiguous labeled conditions (.82 and .82), with this pattern being accentuated for some dimensions (see Table 6; e.g., dominance dimension of affect: .70 for contiguous labeled vs. .48 for interspersed). For stimulus-centered scales, internal consistency may be somewhat lower for interspersed items, with interspersion leading to separation between individual items within a subscale capturing a dimension. On the other hand, this effect does not occur for contiguous and labeled conditions. The bootstrapped 95% confidence intervals were overlapping for the most part, with no consistent pattern.

TABLE 6
Study 3 – Alpha coefficients

		Contiguous	Contiguous and labeled	Interspersed items
Service quality (McDonald's)	Tangibility	.69	.68	.59
	Reliability	.82	.80	.74
	Responsiveness	.77	.68	.69
	Assurance	.75	.77	.70
	Empathy	.77	.71	.64
	Mean	.76	.73	.67
Affective response to ad (Print ad/health club)	Pleasure	.87	.89	.79
	Arousal	.75	.76	.63
	Dominance	.60	.70	.48
	Mean	.74	.78	.64
Endorser evaluation (Michael Jordan for Wheaties cereal)	Attractiveness	.83	.86	.83
	Trustworthiness	.87	.92	.87
	Expertise	.90	.92	.91
	Mean	.87	.90	.87
Retail service quality (Walmart)	Physical aspect	.78	.78	.73
	Reliability	.84	.84	.74
	Personal interaction	.89	.89	.84
	Problem solving	.77	.85	.74
	Policy	.53	.47	.21
	Mean	.76	.77	.65
Perceived value (Computer)	Perceived quality	.90	.91	.79
	Perceived transaction value	.87	.89	.92
	Perceived acquisition value	.97	.97	.96
	Perceived sacrifice	.86	.86	.83
	Perceived price	.86	.92	.78
	Mean	.89	.91	.86
Involvement (PalmOne)	Importance	.94	.93	.88
	Interest	.94	.91	.91
	Mean	.94	.92	.90
Mean		.82	.82	.75

In terms of MGCFA, odd-even parceling was used as in previous studies. Again, almost all of the measures met the criteria of invariance, with the CFI greater than .92 for all models, the TLI greater than .93 for all models, and the BIC minimized when all six forms of invariance imposed (Table 7). The RMSEA was acceptable for all models with the only exception being for the involvement construct, which had a RMSEA < .10 when testing for factor variance invariance. Despite this, there were again no overwhelming differences in patterns of results that held across conditions.

TABLE 7
Study 3 – Multiple group confirmatory factor analysis

Measure	Invariance type	RMSEA	CFI	TLI	BIC
Service quality (McDonald's)	Configural	.059	.989	.967	11820.58
	Metric	.048	.991	.978	11798.86
	Scalar	.065	.981	.960	11787.44
	Factor covariance	.057	.976	.969	11692.84
	Factor variance	.067	.961	.958	11656.52
	Error variance	.083	.931	.936	11642.01
Affective response to ad (Print ad/health club)	Configural	.095	.975	.938	11256.08
	Metric	.082	.975	.953	11226.48
	Scalar	.098	.956	.934	11212.91
	Factor covariance	.088	.957	.946	11182.23
	Factor variance	.092	.945	.941	11162.46
	Error variance	.094	.926	.939	11118.68
Endorser evaluation (Michael Jordan for Wheaties cereal)	Configural	.046	.997	.993	10426.44
	Metric	.042	.997	.994	10397.26
	Scalar	.053	.993	.990	10372.98
	Factor covariance	.057	.991	.989	10347.22
	Factor variance	.093	.972	.970	10349.47
	Error variance	.095	.963	.969	10305.73
Retail service quality (Walmart)	Configural	.065	.974	.959	15507.18
	Metric	.064	.972	.960	15452.82
	Scalar	.068	.964	.953	15409.58
	Factor covariance	.064	.964	.960	15310.54
	Factor variance	.073	.951	.948	15290.44
	Error variance	.074	.942	.946	15199.55
Perceived value (Computer)	Configural	.073	.975	.963	16135.9
	Metric	.070	.975	.966	16066.42
	Scalar	.069	.973	.966	16004.98
	Factor covariance	.067	.972	.969	15909.01
	Factor variance	.078	.960	.958	15905.09
	Error variance	.091	.938	.942	15866.13
Involvement (PalmOne)	Configural	.062	.999	.994	6840.614
	Metric	.052	.998	.996	6821.706
	Scalar	.068	.996	.993	6805.935
	Factor covariance	.080	.993	.990	6800.328
	Factor variance	.116	.980	.978	6799.402
	Error variance	.127	.964	.974	6781.989

Note. RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker-Lewis index; BIC = Bayesian information criterion.

Another set of findings related to relationships across subscales of dimensions within a construct and across measures of different constructs. Although not universally the case, there were instances of lower correlations for the labeled condition when compared to the contiguous or interspersed conditions (e.g., Tables 4A and 4B). This may be due to the logical division of a dimension or a construct that labeling may achieve, as discussed earlier.

Discussion

Overall, Study 3 suggests that the nature of scales (stimulus-centered vs. respondent-centered) is not a significant factor in moderating the effects of format factors on responses for a student sample. Therefore, the lack of differences across conditions is not due to the nature of items pertaining to the self.

Study 4

Method

To further examine stimulus-centered scales, Study 4 used a design where correlations were computed on ratings *across stimuli* rather than respondents, again comparing the labeled, interspersed, and contiguous conditions (Appendix A), the rationale for choosing these three conditions having been discussed earlier. For each condition, five versions were created. Each version involved approximately 20 respondents rating a total of 20 restaurants. Each restaurant was rated on two dimensions (each subscale capturing a dimension with three items): reliability and responsiveness. The aim here was to keep the length of the questionnaire comparable to the lengths in the first three studies, yet collect data on multiple dimensions of service quality. Three items from measures (subscales) of each dimension were used leading to a total of 120 items (six items each for 20 restaurants). Across five versions, the total number of restaurants rated was 100, which becomes the effective sample size. Means were computed for each restaurant on each item. Correlations were then computed across restaurants, with a sample size of 100 providing a basis for MGCFA.

Analysis and Results

The results did not suggest any striking differences across the three conditions (mean coefficient alphas across conditions ranged from .86 to .89; Table 8). The bootstrapped 95% confidence intervals do not display a consistent pattern, with most intervals overlapping.

As there were only three items for each subscale per dimension, parceling was not used in specifying the MGCFA models. Although the RMSEA was high in this study, the CFI, TLI, and BIC all provided evidence for invariance across formatting conditions (Table 9), suggesting that differences across conditions do not emerge for analyses based on correlations across stimuli rather than individuals. For student samples used in our studies, the nature of the analysis, that is, across stimuli versus individuals, does not lead to differences. The high RMSEA is to be expected with the sample size of 20 respondents (as RMSEA does not account for the multiple stimuli being

responded to by these 20 respondents). The BIC is more appropriate here, which provides clear evidence of invariance.

TABLE 8
Alpha coefficients for Study 4

Alpha coefficients	Contiguous and labeled	Contiguous	Interspersed items
Reliability	.91	.92	.91
Responsiveness	.87	.80	.80
Mean	.89	.86	.86

TABLE 9
Study 4 – Multiple group confirmatory factor analysis

Measure	Invariance type	RMSEA	CFI	TLI	BIC
Service quality	Configural	.209	.930	.869	17121.39
	Metric	.184	.929	.900	17086.47
	Scalar	.169	.924	.915	17055.47
	Factor covariance	.164	.925	.920	17044.90
	Factor variance	.160	.922	.924	17029.95
	Error variance	.145	.919	.937	16978.53

Note. RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker-Lewis index; BIC = Bayesian information criterion.

Discussion

Overall, Study 4 suggests that the nature of scales (stimulus-centered vs. respondent-centered) using stimuli as the unit of analysis is not a significant factor in moderating the effects of format factors on responses for a student sample.

Study 5

Method

Study 5 was conducted to examine the effect of another factor, the sample composition, by using a non-student adult sample. The first four studies employed student samples, producing few, if any, striking differences across conditions and thus demonstrating the robustness of published scales to variations in format. Stimulus- versus respondent-centered scales also did not lead to significant differences across conditions. The aim in this study was to examine whether the lack of

differences across conditions, particularly for the respondent-centered measures, is because of students' skills in taking questionnaires and frequent participation in studies, as well as their relatively high cognitive skills and tolerance for abstractions. In this regard, Churchill and Peter (1984) hypothesized but did not find results supportive of the notion that student samples lead to higher reliability. Peterson (2001) conducted a meta-analysis that identified differences in homogeneity and effect sizes between student and non-student samples. Research across several disciplines has examined this issue (e.g., Gordon, Slade, & Schmitt, 1986).

In Study 5, two conditions — labeled and interspersed, the two most distinctly different as discussed earlier — were compared (Appendix A), using a subset of respondent- and stimulus-centered scales (respondent-centered scales: coupon proneness, value consciousness, sale proneness, and price consciousness; stimulus-centered scales, identical to Study 3: evaluation of a computer based on a picture and description on perceived quality, perceived transaction value, perceived acquisition value, perceived sacrifice, and perceived price). Non-student adults were recruited in a university town through several means; by having volunteers at a local non-profit organization complete the questionnaire and by approaching employees of a large university. Questionnaires were distributed through supervisors who asked participants to complete them. One hundred and seventy-four individuals participated in this study, about equally distributed across the two conditions.

Analysis and Results

Analyses similar to previous studies were conducted. Whereas means and standard deviations were largely similar, differences in coefficient alpha emerged for some of the stimulus-centered scales, specifically, perceived transaction value and perceived price scales (Table 10).

TABLE 10
Alpha coefficients for Study 5

	Contiguous and labeled	Interspersed items
Value consciousness	.84	.78
Price consciousness	.56	.71
Coupon proneness	.90	.83
Sale proneness	.71	.70
Mean	.75	.76
Perceived quality	.89	.72
Perceived transaction value	.84	.58
Perceived acquisition value	.93	.86
Perceived sacrifice	.82	.79
Perceived price	.57	.18
Mean	.81	.63
Overall mean	.78	.68

Although the overall mean coefficient alphas (of .78 vs. .68) were somewhat different across conditions, these differences were driven by one construct (the perceived price scale) and should therefore be viewed with caution. As in all other studies, the bootstrapped 95% confidence intervals indicate no pattern, and are overlapping for most conditions.

In terms of MGCFA, the respondent-centered value consciousness scales failed to achieve reasonable fit even when testing for configural invariance again. This is largely due to the problematic measure of the perceived price dimension (Table 11). In particular, the interspersed condition for value consciousness led to an unsatisfactory level of fit in contrast to the contiguous labeled condition, the only divergence between conditions in terms of CFA results that we observed across our five studies. The stimulus-centered computer evaluation scales achieved factor covariance invariance, but not variance invariance or error variance invariance.

TABLE 11
Study 5 – Multiple group confirmatory factor analysis

Measure	Invariance type	RMSEA	CFI	TLI	BIC
Service quality (McDonald's)	Configural	.119	.943	.867	6549.633
	Metric	.153	.890	.780	6557.791
	Scalar	.170	.846	.730	6562.632
	Factor covariance	.151	.847	.786	6537.251
	Factor variance	.141	.849	.813	6519.832
	Error variance	.128	.854	.846	6493.026
Affective response to ad (Print ad/health club)	Configural	.088	.943	.914	6969.076
	Metric	.087	.940	.917	6942.218
	Scalar	.093	.927	.905	6927.056
	Factor covariance	.091	.922	.908	6890.448
	Factor variance	.103	.897	.884	6895.734
	Error variance	.117	.852	.849	6892.397

Note. RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker-Lewis index; BIC = Bayesian information criterion.

Discussion

Overall, Study 5 suggests that, for the most part, the nature of the sample composition is not a significant factor in moderating the effects of format factors on responses. However, some differences emerged such as the finding that the interspersed condition for the value consciousness scale led to a poor fit relative to the labeled condition.

GENERAL DISCUSSION

We set out to investigate whether and how different format practices affect research outcomes, and, in particular, the psychometric properties of measures and the observed relationships between constructs. Researchers use a variety of formats, reflecting implicit beliefs of how each of these practices ameliorates a measurement-related problem. The surprising and consistent findings

from our research suggest that implicit theories about labeling, interspersion and the like may not necessarily hold for student samples. By and large, we do not find differential effects of such presentation issues on measure reliability and validity. In fact, magnitudes of means and standard deviations, internal consistency reliability, stability reliability, dimensionality, cross-dimensional relationships, and cross-construct relationships are, for the most part, unaffected by such variations in format. In particular, formal statistical tests for measurement invariance support the view that formatting for the most part does not impact the psychometric properties of measures for the samples we employed. Arguments such as a possible grouping effect created by labeling versus the opposite created by interspersion do not hold up under empirical testing. The 95% confidence intervals overlap for the most part, an overwhelming pattern further indicating that the differences are statistically insignificant.

Where labeling and interspersion may have an effect, though, is perhaps with non-student samples and respondent-centered scales. Whereas differences across conditions are minimal for student samples, differences emerge for non-student samples. Specifically, we find that the interspersed condition for the value consciousness scale led to a poor fit relative to the labeled condition when we used a non-student sample. However, this difference must be put in the context of the large number of ways in which we explored possible effects across a variety of conditions, scales, and samples. In Study 1, we used 12 scales, compared eight formats for those scales, and found no differences of consequence. In Study 2 (higher cognitive burden relative to Study 1), we used four scales, compared three formats, and found no differences. In Study 3 (stimulus-centered scales vs. respondent-centered scales), we used six scales, compared three formats, and found no differences. In Study 4 (stimulus as unit of analysis), we used one stimulus-centered scale, compared three conditions, and again found no differences. In Study 5 (non-student sample), we used a mix of nine stimulus- and respondent-centered scales, compared two formats, and found a difference in fit across conditions for only one scale. In summary, across a large number of tests, we found only one difference, although only comparing two conditions in this study. This difference though, emphasizes the need for further research on non-student samples across a variety of conditions. Futhermore, there is a need for further research to understand effects at the level of specific measures.

Our research questions the implicit and explicit rationales about presentation effects. By and large, we did not find empirical support for presumed differences between such presentation factors as labeling and interspersion. Thus, when such presentation factors are suspected a priori, either empirical support — often impractical in substantive studies as it would involve experimental manipulation of methodological factors — or evidence through pilot-testing (such as through think-alouds) are recommended to verify the presence of presentational confounds. Such testing would assess whether validated measures display reliability and validity with interspersion, labeling, and other formats.

Our findings point to a recommendation relating to the interactive effects of sample composition and respondent-centered scales. Particularly in terms of dimensionality, we found that interspersion leads to unsatisfactory levels of fit for non-student samples for respondent-centered scales, pointing to the downside of such a format. Clearly, special effort should be taken when surveying non-student adults to motivate respondents, simplifying the length and complexity of questionnaires, and reducing the use of abstractions. However, as noted above, this isolated result should be viewed with caution and subjected to further empirical testing.

So far, our discussion has focused on broad patterns that relate to cross-construct relationships and reflect differences in format conditions. However, narrower patterns may apply for individual scales and reflect differences in format conditions, and even narrower patterns may hold for items within individual scales and reflect differences in format conditions. As an example of the former, the ethnocentrism scale in the resequenced items condition led to a drop in fit in CFA models (Table 3). Similarly, the consumer independent judgment making subscale had a low test-retest correlation, and several subscales of material values displayed relatively low internal consistency under the contiguous paginated condition (Table 2), and perceived transaction value had low internal consistency in the interspersed condition (Table 9). These findings suggest caution in assuming that measures can be presented in ways that differ from their presentation at validation. As format factors involve using items and constructs in ways different from their presentation at validation, implications of our research extend to researchers involved in measure development and validation in efforts to address these issues, as well as to researchers who use these measures in conditions different from those employed for validation.

Finally, we note that our finding of the lack of impact of format has a limitation. In a strict scientific sense, it does not indicate that the impact is absent — rather, it indicates that the evidence does not indicate any impact. We have strived to include as many conditions and moderators as possible, along with relatively large sample sizes, across five separate studies. Further, we have examined a large number of statistics to detect impact. MGCFA has overwhelmingly suggested that measures are consistent. We have also constructed 95% confidence intervals to examine statistical differences. This examination clearly indicates a lack of impact of format on psychometric properties. Although this conclusion is based only on our studies, the evidence does suggest that implicit beliefs about impact of format require reexamination. Study 5 has limitations as well, as we studied only two conditions with a non-student sample. Moreover, the non-student sample was not representative of any larger population and was relatively small in size. Nevertheless, the study expands to non-student samples and emphasizes the importance of further research on this topic with such samples. In general, we note that all the studies would benefit from increased sample sizes for the statistical tests we employed.

In summary, researchers use a variety of different formats to structure questionnaires in theory testing. Such format choices reveal implicit beliefs that labeling, interspersion, pagination, sequencing, and/or pagination affects covariation between measures of constructs of interest. Our empirical investigation, spread across five studies, is, to our knowledge, the first comprehensive systematic examination whether such format choices matter, and if so, how. Our overwhelmingly consistent finding, based on our studies, is that format choices do not affect model fit, particularly for student samples, suggesting that researchers' implicit beliefs may need to be reexamined.

NOTE

1. Across the studies, we designed methods such that variations in administration did not create differences across conditions. However, we caution conservatively that a potential confound exists in this study between the questionnaire being completed at the middle versus end of administration and the conditions. As reported subsequently, the similar pattern of results here when compared to Study 1 suggests that being part of a study of longer duration did not affect the basic pattern.

REFERENCES

Bagozzi, R. P., &. Edwards, J. R. (1998). A general approach to representing constructs in organizational research. *Organizational Research Methods*, *1*(1), 45-87.

Bearden, W. O., Netemeyer, R. G., & Teel, J. E. (1989). Measurement of consumer susceptibility to interpersonal influence. *Journal of Consumer Research*, *15*, 473-481. doi:10.1086/209186

Bradlow, E. T., & Fitzsimons, G. J. (2001). Subscale distance and item clustering effect in self-administered survey: A new metric. *Journal of Marketing Research*, *38*(2), 254-261.

Burchell, B., & Marsh, C. (1992). The effect of questionnaire length on survey response. *Quality & Quantity*, *26*, 233-244. doi:10.1007/BF00172427

Byrne, B. M., Shavelson R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456-466. doi:10.1037/0033-2909.105.3.456

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*(3), 306-307.

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

Churchill, G. A., Jr., & Peter, J. P. (1984). Research design effects on the reliability of rating scales: A meta-analysis. *Journal of Marketing Research*, *21*(4), 360-375.

Cote, J., & Buckley, M. R. (1988). Measurement error and theory testing in consumer research: An illustration of the importance of construct validation. *Journal of Consumer Research*, *14*(4), 579-582.

Cox, E. P. (1980). The optimal number of response alternatives for a scale: A review. *Journal of Marketing Research*, *17*(4), 407-422.

Dabholkar, P. A., Thorpe, D. I., & O Rentz, J. (1996). A measure of service quality for retail stores: Scale development and validation. *Journal of the Academy of Marketing Science*, *24*(1), 3-16.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York, NY: Chapman&Hall.

Feldman, J. M., & Lynch, J. G., Jr. (1988). Self-generated validity: Effects of measurement on belief, attitude, intention, and behavior. *Journal of Applied Psychology*, *73*, 421-435. doi:10.1037/0021-9010.73.3.421

Gordon, M. E., L. Slade, A., & Schmitt, N. (1986). The "science of the sophomore" revisited: From conjecture to empiricism. *The Academy of Management Review*, *11*(1), 191-207.

Greenwald, A. G., & Banaji, M. R. (1989). The self as a memory system: Powerful, but ordinary. *Journal of Personality and Social Psychology*, *57*(1), 41-54.

Grewal, D., Monroe, K. B., & Krishnan, R. (1998). The effects of price-comparison advertising on buyers' perceptions of acquisition value, transaction value, and behavioral intention. *Journal of Marketing*, *62*(2), 46-59.

Groves, R. M. (1991). Measurement error across disciplines. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement error in surveys* (pp. 1-25). New York, NY: John Wiley.

Herzog, A. R., & Bachman, J. G. (1981). Effects of questionnaire length on response quality. *Public Opinion Quarterly*, *45*, 549-559. doi:10.1086/268687

Holbrook, M. B., & Batra, R. (1987). Assessing the role of consumer responses to advertising. *Journal of Consumer Research*, *14*, 404-420. doi:10.1086/209123

Kline, P. (2000). *The handbook of psychological testing*. New York, NY: Routledge.

Kline, T. J., Sulsky, L. M., &. Rever-Moriyama, S. D. (2000). Common method variance and specification errors: A practical approach to detection. *Journal of Psychology*, *134*(4), 401-421.

Knowles, E. S. (1988). Item context effects on personality scales: Measuring changes the measure. *Journal of Personality and Social Psychology*, *55*, 312-320. doi:10.1037/0022-3514.55.2.312

Knowles, E. S., & Byers, B. (1996). Reliability shifts in measurement reactivity: Driven by content engagement or self-engagement? *Journal of Personality and Social Psychology*, *70*(5), 1080-1090.

Krishnamurthy, P., & Sujan, M. (1999). Retrospection versus anticipation: The role of the ad under retrospective and anticipatory self-referencing. *Journal of Consumer Research*, *26*(1), 55-69.

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, *50*, 537-567. doi:10.1146/annurev.psych.50.1.537

Lance, C. E., LaPointe, J. A., & Stewart, A. M. (1994). A test of the context dependency of three causal models of halo rater error. *Journal of Applied Psychology*, *79*, 332-340. doi:10.1037/0021-9010.79.3.332

Lennox, R. D., & Dennis, M. L. (1994). Measurement error issues in substance abuse services research: Lessons from structural equation modeling and psychometric theory. *Evaluation and Program Planning*, *17*(4), 399-407.

Lichtenstein, D. R., Ridgway, N. M., & Netemeyer, R. G. (1993). Price perceptions and consumer shopping behavior: A field study. *Journal of Marketing Research*, *30*, 234-235. doi:10.2307/3172830

Little, T. D., Cunningham, W. A., Golan S., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, *9*, 151-173. doi:10.1207/S15328007 SEM0902_1

Malhotra, N. K., Kim, S. S., & Patil, A. (2006). Common method variance in IS research: A comparison of alternative approaches and a reanalysis of past research. *Management Science*, *52*, 1865-1883. doi:10. 1287/mnsc.1060.0597

Manning, K. C., Bearden, W. O., & Madden, T. J. (1995). Consumer innovativeness and the adoption process. *Journal of Consumer Psychology*, *4*(4), 329-345.

McQuarrie, E. F., & Munson, J. M. (1986). The Zaichkowsky Personal Involvement Inventory: Modification and extension. In P. Anderson & M. Wallendorf (Eds.), *Advances in consumer research* (pp. 36-40). Provo, UT: Association of Consumer Research.

Mittal, B. (1995). A comparative analysis of four scales of consumer involvement. *Psychology and Marketing*, *12*, 663-682. doi:10.1002/mar.4220120708

Nunnally, J. C. (1978). *Psychometric theory*. New York, NY: McGraw-Hill.

Ohanian, R. (1990). Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. *Journal of Advertising*, *19*(3), 39-52.

Parameswaran, R., Barnett A., Greenberg, D., Bellenger, N., & Robertson, D. H. (1979). Measuring reliability: A comparison of alternative techniques. *Journal of Marketing Research*, *16*(1), 18-25.

Parasuraman, A., Zeithaml, V., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, *64*(1), 12-40.

Peterson, R. A. (2001). On the use of college students in social science research: Insights from a second-order meta-analysis. *Journal of Consumer Research*, *28*(3), 450-461.

Podsakoff, N. P., Podsakoff, P. M., MacKenzie, S. B., & Lee, J.-Y. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879-903.

Richins, M. L., & Dawson, S. (1992). A consumer values orientation for materialism and its measurement: Scale development and validation. *Journal of Consumer Research*, *19*(3), 303-316.

Shimp, T. A., & Sharma, S. (1987). Consumer ethnocentrism: Construction and validation of the CETSCALE. *Journal of Marketing Research*, *24*(3), 280-289.

Simmons, C. J., Bickart, B. A., & Lynch, J. G. (1993). Capturing and creating public opinion in survey research. *Journal of Consumer Research*, *20*(2), 316-329.

Smith, S. M., Haugtvedt, C. P., & Petty, R. E. (1994). Attitudes and recycling: Does the measurement of affect enhance behavioral prediction? *Psychology&Marketing*, *11*(4), 359-374.

Steenkamp, J. B. E., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*(1), 78-90.

Sudman, S., Bradburn, N., & Schwarz, N. (1996). *Thinking about answers*. San Francisco, CA: Jossey-Bass.

Szybillo, G. J., Binstok, S., & Buchanan, L. (1979). Measure validation of leisure time activities: Time budgets and psychographics. *Journal of Marketing Research*, *16*(1), 74-79.

Teas, R. K., & Agarwal, S. (2000). The effects of extrinsic product cues on consumers' perceptions of quality, sacrifice and value. *Journal of the Academy of Marketing Science*, *28*, 278-290. doi:10.1177/0092070300282008

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4-70.

Viswanathan, M. (2005). *Measurement error and research design*. Thousand Oaks, CA: Sage.

Weng, L.-J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement*, *64*, 956-972. doi:10.1177/00131 64404268674

Zeithaml, V. A. (1988). Consumer perceptions of price, quality, and value: A means-end model and synthesis of evidence. *The Journal of Marketing*, *52*(3), 2-22.

APPENDIX A

Description of Experimental Conditions

Condition	Description
1. Contiguous	Items appear in the order in which they appeared in validity tests and related measures are contiguous
2. Contiguous and labeled	Same as condition 1, but measures are labeled
3. Contiguous and paginated	Same as condition 1, but measures are on separate pages
4. Interspersed	Items are completely interspersed in the questionnaire
5. Resequenced measures	Items are contiguous, but measures are resequenced
6. Resequenced and labeled measures	Items are contiguous, but measures are resequenced and labeled
7. Resequenced items, contiguous measures	Measures are contiguous, but items are resequenced
8. Resequenced items, contiguous, and labeled measures	Measures are contiguous and labeled, but items are resequenced

APPENDIX B

Models of Invariance
Six Models of Invariance in Order of Stringency of Requirements[a]

Model	Level of invariance	Description of invariance
1	Configural invariance	Similar pattern of factor loadings across conditions
2	Metric invariance	Same factor loadings across conditions
3	Scalar invariance	Intercepts are the same across conditions
4	Error variance invariance	Measurement error is the same across conditions
5	Factor variance invariance	Variance of latent constructs is equal across conditions
6	Factor covariance invariance	Covariance structure of latent constructs is the same across conditions

Note. [a] Testing for invariance requires the researcher to first compare the null model (i.e., a model without any constraints) against Model 1 (configural invariance) on the Bayesian information criterion, and then progressively compare Model 1 through to Model 6 on BIC to assess the type of invariance across conditions. Model 6 represents the most stringent level of invariance.

APPENDIX C

Bootstrapping Procedure

In order to test whether the reliability of measures is dependent on questionnaire formats, we bootstrapped the empirical data to form 95% confidence intervals around the estimated reliability coefficients. Bootstrapping is a commonly used statistical procedure to construct confidence intervals when the distribution of the statistic of interest is not known. Efron and Tibshirani (1993) provide a detailed review of bootstrapping procedures. We implemented a standard case re-sampling procedure, and constructed confidence intervals using percentile bootstrapping, as follows:

1. We first sampled, with replacement, *N* observations from the dataset with *N* observations. That is, if the dataset has 100 observations, we randomly drew 100 observations to construct a bootstrap sample. The "with replacement" procedure implies that we draw the first observation randomly, then put it back in the pool, pick the second observation, put it back in the pool, and so on until we sample 100 observations. This procedure may result in some observations being duplicated in a particular bootstrap sample while others may be omitted.
2. We then estimated the appropriate statistic using the bootstrap sample.
3. We repeated the above steps 500 to 1000 times to estimate each confidence interval. That is, we drew 500 to 1000 bootstrap samples from the original data. We repeated the procedure for 5000 resamples and achieved identical substantive results.
4. Finally, we calculated the percentiles for each of the statistics. In our case, we calculated the 2.5% and 97.5% percentiles to give us 95% confidence intervals around our estimates.

We then examined the confidence intervals of the estimated statistics under the conditions of each study. All bootstrapping results are available from the authors.

TPM®

USING THE GAGE R&R METHOD TO EVALUATE THE RELIABILITY AND ASSESSMENT PROCESS OF THE CREATIVE ENGINEERING DESIGN ASSESSMENT

SOPHIE MORIN
LOUIS-MARC BOURDEAU
JEAN-MARC ROBERT
POLYTECHNIQUE MONTRÉAL

Creativity, being a key competency of engineering, must be taught and assessed. Here, we verify the reliability of the Creative Engineering Design Assessment (CEDA), a psychometric test for engineering students, as well as clarify the assessment process. To test the former, the gage repeatability and reproducibility (R&R) method was applied innovatively. Our findings suggest that the use of the gage R&R method is relevant in a psychometric environment and that the measured total variation is satisfactory. In addition, control charts were used to further analyze the assessment strategy's reliability. We demonstrated that the assessment process for the two qualitative criteria (Originality and Usefulness) of the CEDA was in control, that is, the values' variation was caused by unpredictable but normal and inevitable events. The results demonstrate the test's reliability according to two concepts (repeatability and reproducibility) and allow for the refinement of the assessment process by defining the Likert scales with more precision.

Key words: Creativity; Assessment methodology; Engineering education; Gage R&R method; Reliability.

Correspondence concerning this article should be addressed to Sophie Morin, Polytechnique Montréal, 2900 Edouard-Montpetit Blvd, H3C 3A7 Montréal, QC, Canada. Email: sophie.morin@polymtl.ca

Creative and innovative people are recognized for their contribution to society's well-being. This is particularly the case for engineers who are often called upon to produce innovative ideas and thus participate in the improvement of their organization's products, services, and processes, to keep them competitive. It thus stands to reason that engineering schools must foster creativity among their students as an integral part of their curriculum.

In the context of research and training, the assessment of creativity is a major challenge (Clary, Brzuszek, & Fulford, 2011; Cropley, Kaufman, & Cropley, 2011; Plucker & Runco, 1998; Treffinger, Young, Selby, & Shepardson, 2002). In fact, most of the numerous tools that have been developed in the last 60 years (Clary et al., 2011; Kim, 2014; Treffinger et al., 2002), including the widely used Alternate Uses Test (Guilford, 1968), assess divergent thinking (DT)[1] rather than creativity.

Even though, according to some authors, DT is often confused with creativity, "this can be misleading because convergent thinking[2] is as important for creativity as divergent thinking." (Piffer, 2012, p. 260). As Piffer explains: "The name of the most popular creativity test, the Torrance Test of Creative Thinking, is exemplar. Its name suggests that other cognitive tests (e.g.,

Working memory tests, general knowledge, IQ tests) are not tests of creative thinking." (p. 260). In the same vein, authors like Gabora and Kaufman (2010) have argued that creative production relies just as heavily on knowledge and analytical thinking (which are associated with convergent thinking) as on imagination and divergent thinking.

To address the need for a more holistic evaluation of creativity in engineering, Charyton (2014) recently developed a tool specific to this context: the Creative Engineering Design Assessment (CEDA). The present study seeks to improve the CEDA's evaluation process and determine its reliability using the gage reproducibility and repeatability method (R&R), a statistical method widely used in engineering. Gage R&R is generally used on quantitative data such as length or voltage, but was applied here to the analysis of qualitative data yielded by the CEDA.

The paper is structured as follows. Firstly, we present the CEDA test; secondly, we describe the gage R&R method; thirdly, we explain the methodology and present the results. Finally, we present the discussion and the conclusion with suggestions for future research.

CREATIVE ENGINEERING DESIGN ASSESSMENT (CEDA) TEST

The CEDA is a psychometric test that relies on qualitative criteria requiring assessments made by observers (or judges). Recently developed in the USA by Charyton (2014) to assess engineering students' creative performance, it is based on the Purdue Creativity Test (PCT; Harris, 1960), a well-known, validated test. The CEDA requires participants to conceive concepts using the various geometric shapes presented. As with the PCT, results are assessed according to three criteria (Flexibility, Fluidity, and Originality). The CEDA, however, also includes a "Usefulness" criterion and adopts a more elaborate quantitative scale than the test it derived from. A review of available literature shows the CEDA assesses five important aspects of creativity (Figure 1) as well as confirming both content and face validity of the test (Charyton, 2014; Charyton, Jagacinski, & Merrill, 2008; Charyton, Jagacinski, Merrill, Clifton, & Dedios, 2011; Charyton & Merrill, 2009).

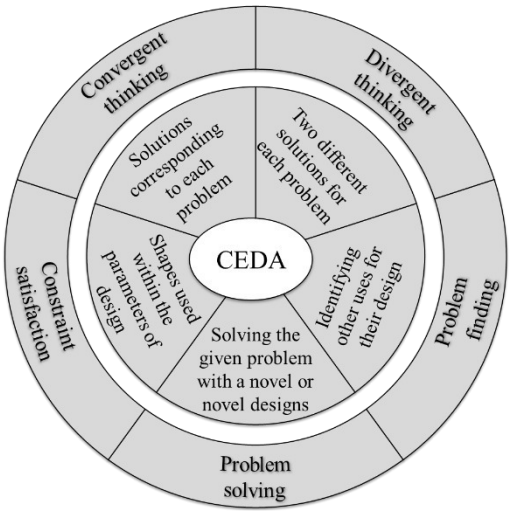

FIGURE 1
Five creativity aspects covered by the CEDA test (figure inspired by Charyton, 2014).

To evaluate creative performance, divided into the five aspects presented in Figure 1, the CEDA uses four criteria: Fluidity, Flexibility, Originality, and Usefulness. Fluidity is the number of different items produced by a person during a creative session; Flexibility is the number of categories covered by these items; Originality is the frequency with which the items are found (or repeated) across the sample; and Usefulness corresponds to how well an item responds to the general goal suggested. The first three (Fluidity, Flexibility, and Originality) are well known and often used to assess creativity (Clary et al., 2011; Kim, 2006; Treffinger et al., 2002). Charyton et al. (2011) added the Usefulness criterion due to its importance, indeed essentiality, in an engineering context. The test provides a numerical score, from zero to 284, with a higher score indicating higher creativity.

Scoring and interpretation of results obtained with the CEDA present certain drawbacks. Indeed, very limited quantitative data have been published, so a range of results is difficult to establish. Moreover, the Fluidity and Flexibility criteria are measured quantitatively, whereas Originality and Usefulness are measured qualitatively using Likert scales. Even with a somewhat detailed description of each level of these scales, subjectivity can still be present. To this day, to our knowledge, no results have been published to show, describe, and analyze links between assessors' work and participants' scores.

A final limitation of the CEDA lies in the actual instructions for assessment of the data collected. From our point of view, this test has several merits but lacks details in the assessing process. "As in the previous study, two judges scored each CEDA: one judge from engineering and one judge from psychology. There was a total of four engineering judges, who scored subsets of the CEDA's, and one psychology judge, who was a CEDA test developer. Two of the CEDA test developers trained the judges. Judges practiced scoring in a team environment; however, each judge evaluated the CEDA's separately" (Charyton et al., 2011, p. 785). From a practical viewpoint, guidelines are vague and difficult to apply. Very limited work has been published by other researchers to further understand the test's operationalization (Carpenter, 2016).

THE GAGE R&R METHOD

The gage R&R is a statistical method used in engineering to measure (or gage) the reliability of a measurement system (Ostle, Turner, Hicks, & McElrath, 1996; Wheeler, 2006). Typically, in engineering these would include the operator (human or not), the work piece, and the tool. A gage R&R study helps determine whether the measurement system's variability is small compared with the process' variability; how much variability is caused by the operators; and whether the measurement system is capable of discriminating between different parts.

Although the number of operators, parts, and trials varies for a particular application of the gage R&R method, the general procedure of measurement with this method consists of forming a part sample; randomly choosing three operators (if the operator is human, they should be trained and familiar with the process but be neither novice nor expert); proceeding with the first assessment (T1); repeating the assessment for the second trial (T2) (randomizing the order of the measurements).

If the total variance associated with the measurement system (repeatability and reproducibility) is less than 10%, the measurement system is judged acceptable. If the total variance is between 10% and 30%, the measurement system is judged acceptable depending on the

application, the cost of the measuring device, the cost of repair, or other factors. Finally, if it is more than 30%, the measurement system is judged unacceptable and should be improved (Automotive Industry Action Group, AIAG, 2010; Ostle et al., 1996; Wheeler, 2006).

Principal Elements

The gage R&R method is used for multiple purposes: to compare the measurement system variability to the process variability, calculate how much variability in the measurement system is caused by differences between operators, and determine if the measurement system is capable of discriminating between different parts.

Accuracy and Precision

The gage R&R method measures accuracy, defined as "hitting the right spot" (i.e., measuring the right thing), and precision, defined as "hitting the same spot every time" (i.e., obtaining the same result every time the measurement is done). In other words, the method can be used to determine if a measuring system consistently measures the right concept. As pertains to psychometric testing, the gage R&R method can thus be used to analyze variability in repeatability and reproducibility and to identify to what extent variance in the results is due to the measurement system. Figure 2 illustrates how repeatability and reproducibility are combined in this statistical method.

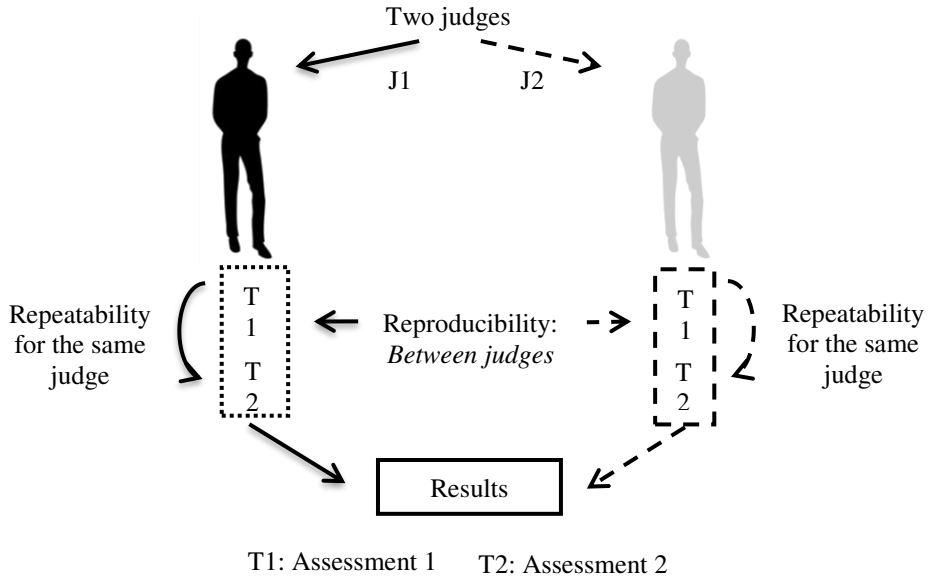

FIGURE 2
Repeatability and reproducibility constructs.

The accuracy aspects of the test, bias, linearity, stability (AIAG, 2010), are not addressed in this study. As presented above, we rely on previous studies by Charyton (2014) to establish that the CEDA has a good validity and indeed assesses creativity.

A measurement system's precision is evaluated by two constructs: repeatability and reproducibility. "The repeatability of a measuring device is the variability observed when repeated measurements are obtained by the same operator on the same unit or part" (Ostle et al., 1996, p. 337). "The reproducibility of the measurement process is estimated by considering the variability among the sample averages for the operators used in the study" (p. 340). The gage R&R method analyzes the two constructs' variability and identifies to what extent variance in the results is due to the measurement system.

Control Charts

One of the most useful features of the methodology of evaluation is the control charts (Figure 3). They allow the professionals in charge of the test (e.g., engineer, researcher, manager) to visually determine if the measured values are between the upper and lower control limits, that is, if a process is "in control." According to Wheeler and Chambers (1992), a process will inevitably include variation. However, two types of variation exist: controlled and uncontrolled. Controlled variation is due to "random" causes and uncontrolled variation is due to "assignable" causes. A process "not in control" is being affected by assignable causes that can be identified and eliminated. Control charts are the tool used to overcome these assignable causes and move beyond the barrier of process improvement. The data of several operators (judges) can be displayed on the same chart to get a global view of the assessments.

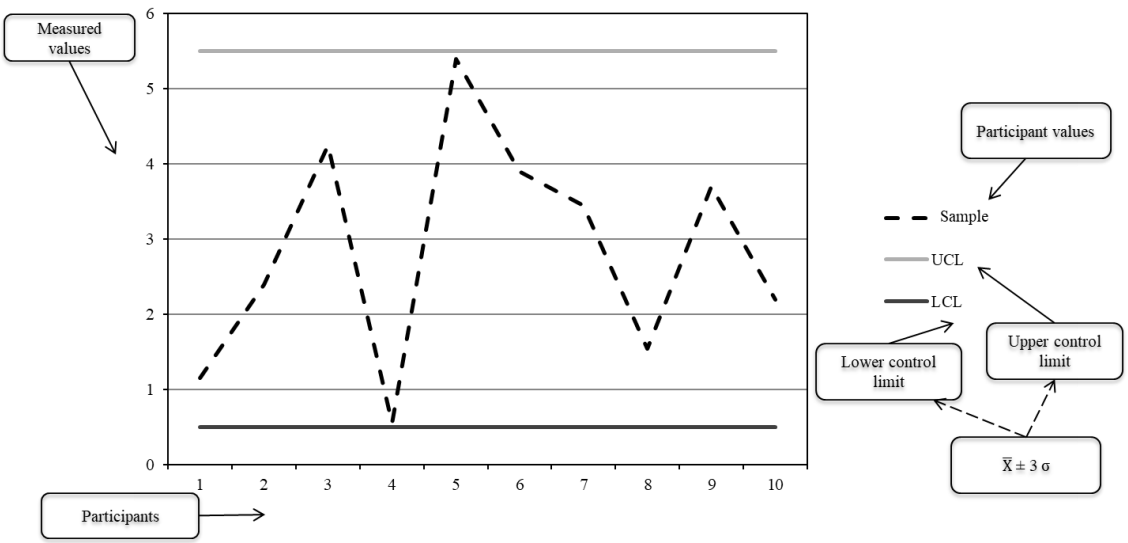

FIGURE 3
Principal elements of the control charts.

METHOD

This study had two objectives: to clarify the CEDA's assessment process and to test the reliability of the CEDA measurement system using the gage R&R method. The method presents

the procedure as a methodological design in two phases, the samples, the judges, and the test used for testing the reliability and the assessment process of the CEDA.

Procedure

Phase 1

In order to optimize participation, we translated the test into French. Hence, a pretest step was conducted to verify language and understanding of the guidelines.

Phase 1 aimed to establish a more detailed assessment scale for the Originality and Usefulness criteria than what was suggested by Charyton as well as to clarify the CEDA's assessment process. Participants were chosen from guests attending a workshop on creativity. They were invited through personal networking and included people interested in learning more about creativity and how it may be developed (e.g., project manager, sound technician, pedagogical consultant, marketing director, industrial engineer, receptionist, professor). The test was presented as an introduction to this conference and workshop. No compensation was given to the participants. Workshops consisted of three sessions of three hours each.

Twenty-two tests were evaluated by three judges, all female engineers aged 25-40 years. All three were educators with an interest in creativity as a competency and a background in art (circus, dance). They assessed the CEDA tests in two phases, two months apart to minimize the memory bias.

Confronted with assessment difficulties and questions, the judges met to clarify certain aspects of their evaluations. To diminish the confusion regarding the levels for Originality and Usefulness, copies of the designs created by the participants were made (sketch only) and classified according to the scale suggested. One of the judges suggested it would be helpful to define with more precision each level in regard to the test itself. To accomplish the task, they looked at each design and used the "think aloud" approach to describe their understanding. This helped them build common knowledge and become more confident with the assessment task.

Judges defined in more detail what each level meant with regard to the CEDA. All assessments were done individually, but a group discussion was arranged between the two assessments of each phase to resolve any remaining confusion or disagreements.

For the first analysis, the three judges used the CEDA's original scoring system. They followed the general instructions given by Charyton (2014). They used objective/quantitative measures for Fluidity and Flexibility, and subjective/qualitative measures (on two Likert scales) for Originality and Usefulness. They made one judgment for each design and a third one (global) for each problem (Figure 4-B; D1, D2, Global). They proceeded with the summation of the four criteria, with the equation suggested by Charyton: Fluidity + Flexibility + 2*Originality + 2*Usefulness (overall creativity score). We found confusing the use of the term "overall" to describe the "global" judgment for design 1 and 2 (D1 and D2) and not the overall creativity score obtained from the previous formula so we changed overall to global.

Moreover, this phase allowed us to conduct a first evaluation of the judges' performance and concordance with the gage R&R method. Two months after a first assessment, a second one was conducted (reproducibility aspect). The tests were randomly distributed to eliminate a possible interaction bias across tests. Control charts were built to visually compare reproducibility and repeatability for each judge.

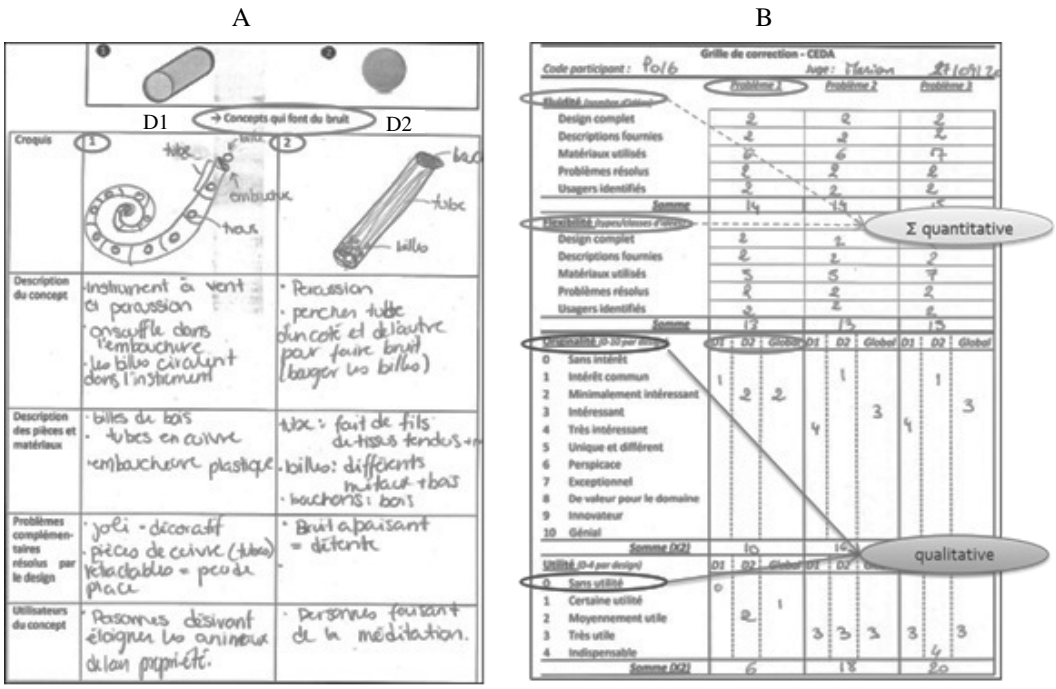

FIGURE 4
CEDA examples (A: answer sheet; B: assessment sheet).
D1 = design 1; D2 = design 2.

Phase 2

Phase 2 aimed to test the reliability of the CEDA measurement system using the gage R&R method. In Phase 2, a somewhat homogenous group (all were industrial engineering students in their second year at Polytechnique Montreal) completed the test. It was administered as a creativity exercise to prepare them for the semester's project. No financial or academic compensation was given, and all participants signed a consent form. The first author of this paper was invited by the professor in charge of the course "Integration project" as a creativity consultant to help students integrate creativity into their project.

Only two judges assessed the tests in Phase 2. The judges counted the items together instead of separately. This method allowed us to obtain faster assessments with greater calculation certainty. At that time, the judges had not discussed or made comments about the Originality or Usefulness of the designs; they kept those opinions for the individual assessments. After the first assessments were completed, judges met to discuss scores that were more than two levels apart (e.g., 2 = *somewhat interesting* and 4 = *very interesting*). They wanted to understand the differences in judgement and adapt the descriptions of the assessment criteria if necessary. Two months later, they proceeded to reassess Originality and Usefulness for the 98 tests, randomly redistributed.

Following the assessment, the gage R&R method was used a second time to evaluate more specifically the variability of the Originality and Usefulness scores. We also built control charts to verify if the scores obtained corresponded to an "in control" assessment process.

Participants

Table 1 describes the two samples of participants to our study. We had a total of 120 participants, 22 in the first phase and 98 in the second. In total, there were 54 females and 66 males, all with engineering backgrounds (94% are undergrads engineering students in different specialties). All but four participants were between 20 and 39 years old with a large majority (N = 102) between 20 and 29.

TABLE 1
Participants description

Sample description	Phase 1	Phase 2
Number of participants	22	98
Gender (female; male)	6; 16	48; 50
Engineering education (undergrads; grads)	16; 6	98; 0
Age (20-29; 30-39; 40-49; 50-59)	8; 11; 2; 1	94; 3; 1; 0

Test

Participants were given the test in one paper document. They could use any of their own crayons, pencils, erasers, and so forth. They had 30 minutes to complete the test. The CEDA is presented on four pages, three with problems and one with guidelines. There is a different general goal for each problem (i.e., design that can produce sound, design that can communicate, design that can travel). Participants had to describe two original designs for each problem (1-2) built around suggested objects (sphere, cube, cylinder, pyramid). In Figure 4-A, two designs can be seen (two columns); the general problem suggested is defined as “a concept that can produce a sound” and the two proposed objects are a sphere and a cylinder. This example represents the first of the three pages comprising the test.

RESULTS

In Phase 1, the assessment process is studied and clarified. Also, control charts allow us to compare judges’ concordance. In Phase 2, a statistical analysis shows how the test is reliable from two aspects, repeatability and reproducibility.

Assessment Clarification

The scale provided by Charyton (2014) to evaluate the two qualitative criteria, Originality and Usefulness, was a starting point but remained difficult to use because of the lack of specificity. With the sorting exercise, each level of the scale was defined more precisely according to the

answers at hand. Even with 22 tests, categories and patterns emerged (e.g., musical instruments, houses, cars, etc.). With these "subgroupings" of designs, it was possible to visualize what could be expected in the specific context of the CEDA and for each level. The same process was applied for Originality and Usefulness.

These discussions and findings were used to clarify and standardize the assessment strategy. Table 2 and Table 3 show what Charyton (2014) provided and what this study added. Another observation concerns the scale itself. No results over 6 were given or obtained.

TABLE 2
Description of the Originality criterion

From Charyton (2014)	Added in our study
0 – Dull	Does not correspond to the general goal suggested; common object (daily use)
1 – Common place	Designs that often reoccur in the tests
2 – Somewhat interesting	Minimal transformation or use of multiple suggested forms, multiple materials, added objects or materials
3 – Interesting	A more perceptive concept, but not developed enough
4 – Very interesting	Combination of two or more simple concepts
5 – Unique and different	Combination of more developed and complex concepts, concepts not existing in this suggested form
6 – Insightful	Well-developed idea, well-described (details), combining multiple concepts (different fields) in a novel way
7 – Exceptional 8 – Valuable to the field 9 – Innovative 10 – Genius	In all of our assessments, no designs obtained higher marks than 6, so we didn't have examples to discuss and compare scores 7 to 10.

TABLE 3
Description of Usefulness criterion

From Charyton (2014)	Added in our study
0 – Useless	Does not respond to the general goal suggested, does not present any possible uses
1 – Somewhat useful	Responds to the general goal suggested but has limited possible uses
2 – Useful	Relevant uses but for very specific cases
3 – Very useful	Existing concepts but not optimal/one solution among others, existing concept that needs elaboration
4 – Indispensable	Existing concepts, indispensable or integrated to modern life

The Gage R&R Method

Phase 1

Figure 5 illustrates the repeatability and reproducibility of creativity measures with the CEDA. Each chart shows the two measures made for each participant by one of the three judges. The first chart (at the top) shows that Judge 1 gave the most consistent judgments since the two lines are very close to each other. Judge 3 (bottom chart) is less consistent, and Judge 2 (middle chart) is even less consistent (see circles).

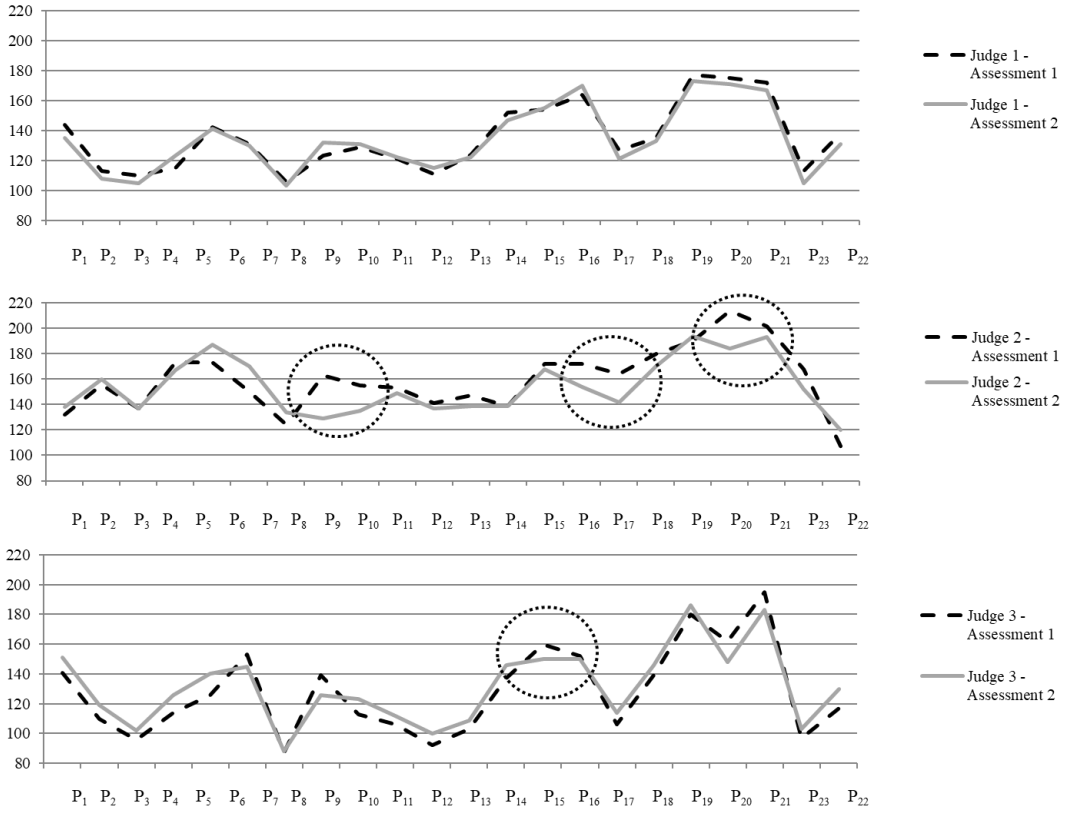

FIGURE 5
Creativity scores on the CEDA attributed by the three judges in Phase 1.
P = participant in ordinate; CEDA score in abscissa = min: 84, max: 214; P23 and P22 were reversed for analysis reasons (Judge 2 did not assess P22 accordingly).

The Pearson concordance coefficient between judges was calculated to show how similarly the three judges assessed the results for each participant (Kline, 2005). As Table 4 shows, the coefficient is higher for Judges 1 and 3. In line with this result and for organizational reasons (availability, time, cost), we felt comfortable proceeding with two judges (1 and 3), as was the case in previous studies by Charyton (Charyton et al., 2008, 2011; Charyton & Merrill, 2009). Furthermore, Charyton proceeded with two judges in all of her studies, so we believed it was a suitable and appropriate decision. Even though we have confidence that with more training Judge 2 could tighten her results, for the reasons mentioned above, we chose to continue the project with only two judges.

TABLE 4
Pearson concordance coefficient

	R^2		
Variable	$\bar{X}$ J1	$\bar{X}$ J2	$\bar{X}$ J3
$\bar{X}$ J1	1	.398	.828
$\bar{X}$ J2		1	.452
$\bar{X}$ J3			1

Note. R^2 = Pearson coefficient; J = judge.

Phase 2

The 98 test results from Phase 2 were divided randomly into four samples (25, 25, 25, 23) because the different constants used to calculate the variations in reproducibility and repeatability are established for small samples (N = 25). Control charts were built to verify if the measures were in control (i.e., predictable from a statistical point of view). Figure 6 shows visually that the values obtained are between the limits so that the assessment process is in control for the two qualitative criteria, Originality and Usefulness, confirming the test's reliability. This means that the values' variation is caused by unpredictable but normal and inevitable events and nothing specific can be done to further control the assessment process.

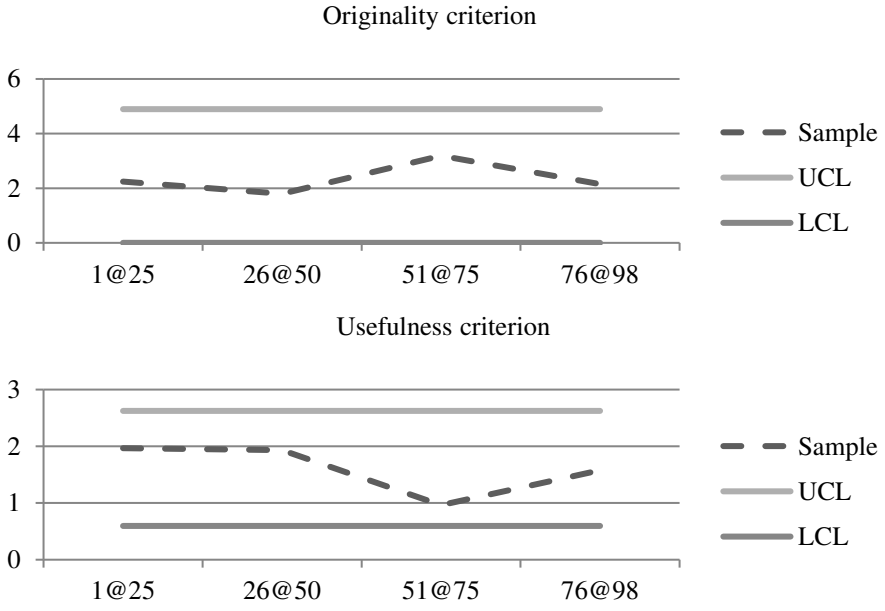

FIGURE 6
Control chart – Originality and Usefulness criteria (four samples).
UCL = upper control limit; LCL = lower control limit.

Using the same four groups of participants, Figure 7 and Figure 8 illustrate the percentages of evaluation variability related to three categories: repeatability, reproducibility, and participants. When using a measuring system, it is essential that the variability observed be due to the participants and not to the instrument itself or its use. This is what is shown in these graphs. For the two criteria Originality and Usefulness, between 85% and 95% of the variability is due to the participants and not the judges or the assessment process (repeatability and reproducibility).

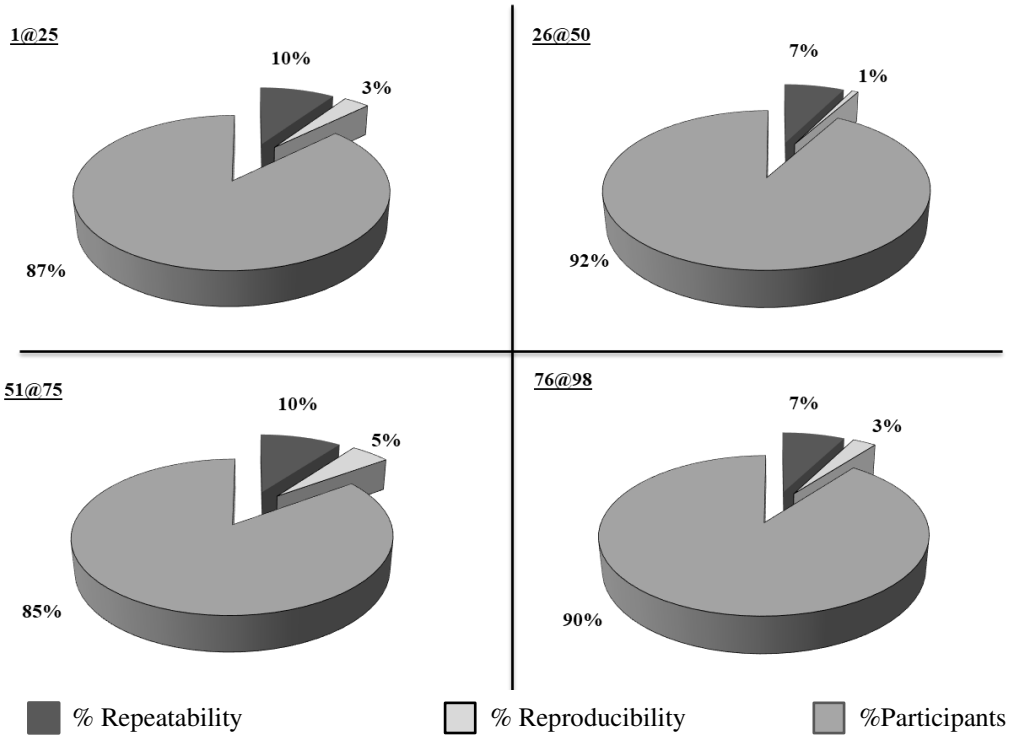

FIGURE 7
Percentages of evaluation variability for Originality (four samples) (Average global score: 88.50%).

In short, we calculated the standard deviation (*SD*), the variance, and their percentages. The measurement system is responsible for 8.50% of the total variance (see Table 5), 6.70% of the variance is caused by repeatability, and 1.80% by reproducibility. For an engineering process, a 10% limit is usually the maximum acceptable (AIAG, 2010; Ostle et al., 1996; Wheeler, 2006). This value is therefore considered acceptable. In other words, 91.50% of the variance is due to the participants' differences.

DISCUSSION AND CONCLUSION

Phase 1

With the assessment guidelines provided by Charyton (2014), the three judges performed a first round of assessments. The guidelines concerning the two criteria Fluidity and Flexibility, and how to count the items and categories, were reasonably easy to follow, but those regarding

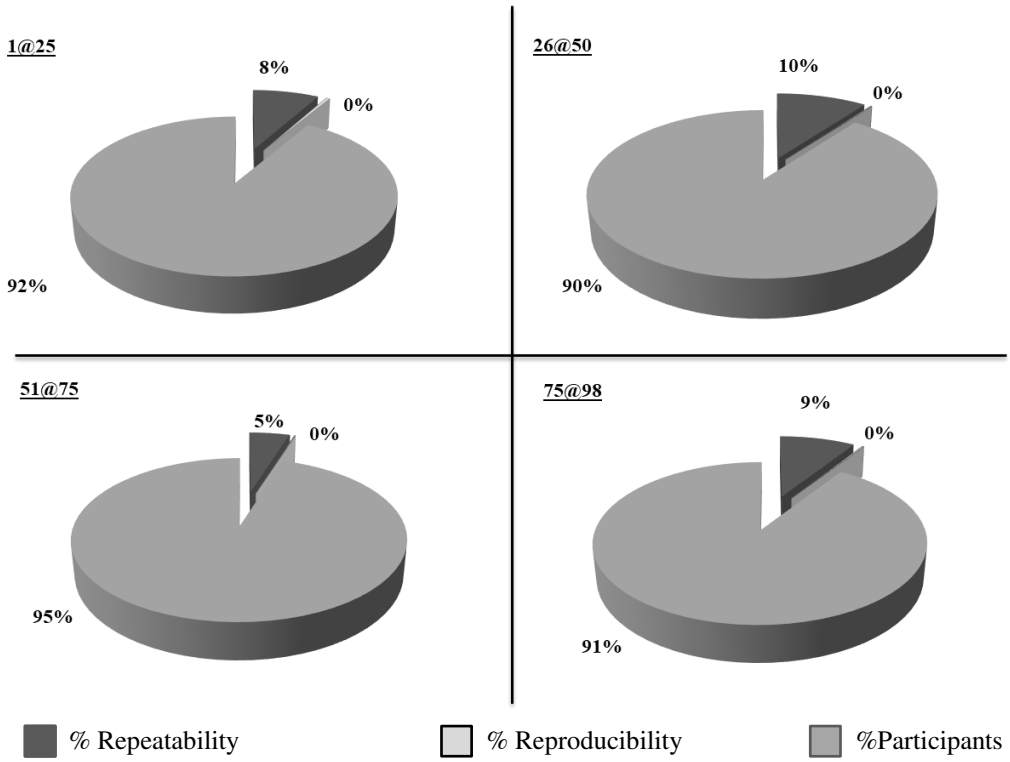

FIGURE 8
Percentages of evaluation variability for Usefulness (four samples) (Average global score: 92%).

TABLE 5
Types of variances

Variations	*SD*	Variance	% (gage R&R)	% (gage R&R total)
Repeatability: measure to measure	5.79	33.55	78.82	6.70
Reproducibility: judge to judge	3.00	9.02	21.18	1.80
Gage R&R total	8.80	42.57	100.00	8.50
Subject to subject	21.40	458.00		91.50
Total	30.20	500.57		100.00

Originality and Usefulness were much less practical. They caused assessment difficulties, requiring the judges to meet and discuss several answers. Even though a few words were provided to describe the Likert scale (Table 2 and Table 3), the judges felt they did not describe in enough detail the possibilities revealed in the tests. What is "somewhat interesting"? What is "moderately useful?" "To score Originality (Uniqueness), rate each design on the scale from 0 to 10. Scorers or judges should think of a word on your own that describes each design and then look on the rubric list to find the word and assign that number to the design" (Charyton, 2014, p. 21).

No completed assessment sheets (examples with designs and scores) have ever been published or made available to understand how the judges on Charyton's team worked. We are aware that the descriptions we came up with still leave room for personal interpretation. However, with training and practice supported by a manual containing examples (designs and scores), judges should be better equipped to make more precise assessments, specifically toward designs produced with the CEDA.

Another difficulty was the assessment of the participant's "global" performance suggested by Charyton. "Each design should be assessed separately (D1, D2). Then, an overall evaluation of the entire problem should be rated. The Originality score for the entire problem (global) will be the score that is analyzed and becomes the overall Originality score for the problem. Although each design score can be inputted and analyzed, we recommend using the overall problem score. It is also important to note that this process of scoring each design is pertinent towards making an assessment of the overall Originality score per problem" (Charyton, 2014, pp. 21-22). This explanation was not convincing and did not sufficiently describe how this overall score could or should be used. Also, according to Charyton, all scores should be added up (D1, D2, global). Given these obstacles, we made changes to this strategy in Phase 2, which we will discuss below.

A gage R&R statistical analysis was performed to see if the three judges were able to adequately repeat their assessments over time (period of two months) as well as come up with similar scores (overall precision of the assessment system). In this phase, the judges followed the aggregation method suggested by Charyton (2014). She proposes a formula adding the four scores (Fluidity, Flexibility, Originality, Usefulness), but the two scores of Originality and Usefulness are multiplied by 2. The explanation provided by the author is the following: "The correlations for the revised formula with Usefulness (2*Usefulness added to the original CEDA formula) illustrates similar findings with the new scoring of the revised CEDA compared with the previous scoring method without Usefulness" (p. 18). When asked in a personal exchange to provide more details, Charyton added "this formula was based on theory in relation to the conceptualization of Originality and Usefulness as integral components of creativity specific to engineering design" (personal communication, May 11, 2013). We still had serious reservations, so we made adjustments in Phase 2.

Phase 2

Charyton (2014) proposed an evaluation of every design but also added a third, more global, one to give an average score for each pair of designs (Figure 4-B). In collaboration with a statistician, we determined this score was unnecessary as it represented an average of two scores we already sum up. Moreover, it accentuated the gap between judges. It was an additional judgment that did not even assess a specific element of design.

An important problem arose when a specific situation occurred: if participants came up with only one of the two designs per problem, they would get a score for the first one (e.g., 3 = somewhat interesting) but would get 0 for the second one. What should the global score be? How does it adequately represent the participant's overall performance? To overcome this hurdle, we decided to eliminate the global score. Statistically it was pulling apart the judges' assessments, and theoretically it was not adding any information to the result.

As mentioned above, we felt uncomfortable using the CEDA's original scoring formula proposed by Charyton et al. (2011) because it lacked specifications. We found limited statistical and theoretical explanations as to why Originality and Usefulness numbers should be doubled as well as for the reference value of 100 for the Fluidity and Flexibility criteria. Therefore, in Phase 2, all four scores were calculated independently and no total scores were tabulated. This would allow us to conduct a more specific analysis of each criterion and keep differences between judges to a minimum. Even with these modifications, the CEDA remains a relevant tool since it provides an overall assessment of creativity (five creativity elements and four evaluation criteria), which is particularly rare in the literature on the subject.

Final Comments

This study had two objectives: to test the reliability of the CEDA measurement system and to clarify the assessment process of the CEDA. For the first objective, we used the gage R&R method to verify reliability regarding two aspects, repeatability and reproducibility. With 91.50% of the evaluation variation caused by the participants (not the judges or the test itself), the test was proven to be highly reliable. The second objective was achieved by organizing discussions between judges about the classification and scoring of the participants' design works. These yielded longer and better descriptions of the qualitative Likert scales for the Originality and Usefulness criteria, as well as a critique and a revision of the scoring process.

It was innovative to use the gage R&R method to analyze data from a psychometric test. To our knowledge, this is not a common application. It has an advantage over other types of analysis commonly used in social sciences (e.g., variance component analysis), as it can be used with a small sample.

For future research, we believe that a global score, a composite indicator of creativity, could be established with more precision and specifications. A single score would be easier to manage and work with (compare, rank, etc.) than four. Charyton (2014; Charyton et al., 2011) did use a global score, but it remains questionable for the reasons evoked above (multiplication by 2, overall problem score). Also, from a statistical point of view, the scores of Fluidity and Flexibility are always very close to each other and seem highly correlated. Should they be merged or should one be eliminated? Interesting research could be done in this direction. Finally, we started to build a visual guide with the different designs collected to facilitate and simplify the assessment process. Sketches with scores could be used to guide future judges in their assessments and help them provide comparable results from one study to another.

NOTES

1. According to Smith & Ward (2012, p. 465), divergent thinking is "The search for many varied and imaginative possible problem solutions."
2. According to Smith & Ward (2012, p. 465), convergent thinking is "Type of problem solving or reasoning in which cognitive operations are intended to converge upon the single correct answer."

FUNDING

We are grateful for funding to the first author from the Fonds de recherche du Québec – Nature et technologies (FRQNT 183790).

REFERENCES

Automotive Industry Action Group. (AIAG). (2010). *Measurement System Analysis (MSA)*. Retrieved from http://www.rubymetrology.com/add_help_doc/MSA_Reference_Manual_4th_Edition.pdf

Carpenter, W. A. (2016). *Engineering creativity: Toward an understanding of the relationship between perceptions and performance in engineering design.* (Unpublished doctoral dissertation). University of Akron, Ohio, OH.

Charyton, C. (2014). *Creative engineering design assessment: Background, directions, manual, scoring guide and uses*: London, UK: Springer.

Charyton, C., Jagacinski, R. J., & Merrill, J. A. (2008). CEDA: A research instrument for creative engineering design assessment. *Psychology of Aesthetics, Creativity, and the Arts*, *2*, 147-154. doi:10.1037/1931-3896.2.3.147

Charyton, C., Jagacinski, R. J., Merrill, J. A., Clifton, W., & Dedios, S. (2011). Assessing creativity specific to engineering with the revised creative engineering design assessment. *Journal of Engineering Education*, *100*(4), 778-799.

Charyton, C., & Merrill, J. A. (2009). Assessing general creativity and creative engineering design in first year engineering students. *Journal of Engineering Education*, *98*(2),145-156.

Clary, R. M., Brzuszek, R. F., & Fulford, C. T. (2011). Measuring creativity: A case study probing rubric effectiveness for evaluation of project-based learning solutions. *Creative Education*, *2*, 333-340. doi:10.4236/ce.2011.24047

Cropley, D. H., Kaufman, J. C., & Cropley, A. J. (2011). Measuring creativity for innovation management. *Journal of Technology Management & Innovation*, *6*(3), 13-30.

Gabora, L., & Kaufman, S. B. (2010). Evolutionary approaches to creativity. In J. C. Kaufman & R. J. Sternberg (Eds.), *Cambridge handbook of creativity* (pp. 279-300). New York, NY: Cambridge University Press.

Guilford, J. P. (1968). *Creativity, intelligence and their educational implications.* San Diego, CA: EDITS-Knapp.

Harris, D. (1960). The development and validation of a creativity test in engineering. *Journal of Applied Psychology*, *44*, 254-257. doi:10.1037/h0047444

Kim, J. (2014, August). *An assessment method to evaluate team project based engineering design*. Paper presented at the 9th International Conference on Computer Science & Education, Vancouver, Canada.

Kim, K. H. (2006). Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking (TTCT). *Creativity Research Journal*, *18*, 3-14. doi:10.1207/s15326934crj1801_2

Kline, T. J. B. (2005). *Psychological testing: A practical approach to design and evaluation*. Thousand Oaks, CA: SAGE Publications Ltd.

Ostle, B., Turner, K. V., Hicks, C. R., & McElrath, G. W. (1996). *Engineering statistics, the industrial experience*. Belmont, CA: Duxbury Presspages.

Piffer, D. (2012). Can creativity be measured? An attempt to clarify the notion of creativity and general directions for future research. *Thinking Skills and Creativity*, *7*, 258-264. doi:10.1016/j.tsc.2012. 04.009

Plucker, J. A., & Runco, M. A. (1998). The death of creativity measurement has been greatly exaggerated: Current issues, recent advances, and future directions in creativity assessment. *Roeper Review*, *21*, 36-39. doi:10.1080/02783199809553924

Smith, S. M., & Ward, T. B. (2012). Cognition and the creation of ideas. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 456-474). New York, NY: Oxford University Press.

Treffinger, D. J., Young, G. C., Selby, E. C., & Shepardson, C. (2002). *Assessing creativity: A guide for educators*. Retrieved from http://nrcgt.uconn.edu/wp-content/uploads/sites/953/2015/04/rm02170.pdf

Wheeler, D. J. (2006). *EMP III (Evaluating the measurement process III): Using imperfect data*. Knoxville, TN: SPC Press.

Wheeler, D. J., & Chambers, D. S. (1992). *Understanding statistical process control*. Knoxville, TN: SPC Press.

TPM®

UNEASE IN ORGANIZATIONS: A STUDY AMONG USERS OF AN ANTI-MOBBING WINDOW

ROBERTA MAERAN
MARCO MARCATI
UNIVERSITY OF PADOVA

MASSIMO DE FELICE
FEDERATION OF ITALIAN TRADE UNIONS IN VENEZIA MESTRE

This study analyzes mobbing, a particularly relevant phenomenon nowadays, as the attention given by media shows. The research involved the users of a trade union's anti-mobbing window set up in an Italian town in 2006, with the aim to provide assistance and support to the workers seeking help. After an examination of the literature, the models, causes, and effects of mobbing on the individuals, their families, and the organization, we presented the intervention methodology implemented at the window. The intervention envisaged an initial interview, administering the Leyman Inventory of Psychological Terror (LIPT; Leymann, 1990) adapted by Ege (2002; LIPT Ege), and determining the professional harm that mobbing causes. A qualitative and quantitative analysis provided a snapshot of mobbing regarding the data collected through the LIPT on a sample of 106 workers, and identified possible areas of intervention from a psychological and legal point of view.

Key words: Mobbing; Organizations; Anti-mobbing window; Psychological terror; Double mobbing.

Correspondence concerning this article should be addressed to Roberta Maeran, Department FISPPA – Section of Applied Psychology, University of Padova, Via Venezia 14, 35131 Padova (PD), Italy. Email: roberta.maeran@unipd.it

The phenomenon of mobbing has recently become relevant in different research fields, attracting the attention of many investigators and experts and of the media as well. Research previously conducted to register how and how often seven national and local newspapers discussed the phenomenon, revealed knots in the interpretations and an abiding shortage in definitions (Maeran & Bavarone, 2007).[1]

A more accurate review of the phenomenon was the subject of services covered by the national news and the economic pages of some newspapers.[2] These news reports presented real cases of mobbing victims in various work contexts (from public to private), placing importance on the experiences of unease caused and their impact on a personal, family, and professional level. Media's recent attention to mobbing underlines how the phenomenon has spread to various organizational contexts, allowing a broader public to understand this insidious form of organizational discomfort. This treatment, however, does not highlight that mobbing is mainly an organizational problem and that it concerns not only the people involved. According to Leymann (1996), mobbing is not to be intended as a disorder of interpersonal relationships, rather it has to be considered in the larger system of organizational factors. That said, one's workplace, living

together, can turn into a relational hell in a particularly significant sphere of human existence (Dejours, 1998; Williams, 2001; Zamperini & Menegatto, 2013).

The spreading of organizational models, in which the ability to win the competition between individuals is considered a value and a management style, generates a working condition that is increasingly marked by a limited sense of belonging; the latter is inevitably reflected in the organizational climate, interpersonal relationships and, possibly, individual and group performances. In these contexts, workers are not only deprived of their control over their workplace, but also of the chance to respond properly to the demands of everyday social life. An organization that forces teamwork without favoring participation and trust, instead encouraging internal competition with benefits and rewards, can be considered a context in which mobbing may arise (Zapf & Kulla, 1996). From these premises and the analyses of the main models and definitions of mobbing, this study aims to investigate episodes of unease and organizational discomfort through the anti-mobbing window activities set up by a trade union in a town in Northern Italy in 2006.

MOBBING: SOME DEFINITIONS

A definition, among those given for mobbing, is by Leymann (1996),

> mobbing or psychological terror on the workplace consists of a hostile and not ethical communication carried out consistently by one or more individuals usually against a single individual, who, being forced into that position by continuous harassing actions ends up being defenceless and unable to get any help. These actions occur with a standard high frequency and endure for a long time. The high frequency and the duration of a hostile behaviour cause serious psychological, psychosomatic and social problems (pp. 168).

According to Ege (2001),

> mobbing is a war on the job in which, through psychological, physical and/or moral violence, one or more victims are forced to fulfil the will of one or more attackers. This violence is expressed through frequent and persisting attacks that are intended to harm the victim's health, communication channels, information flows, reputation and/or professionalism. The psycho-physical consequences of such an aggressive behavior are unavoidable for the one being mobbed (p. 33).

Analyzing the Italian situation, with the specific aim of better defining mobbing events, Ege (2002) developed a model consisting of six stages that are logically related to one another: "condition zero" or generalized conflict; Stage 1: targeted conflict; Stage 2: mobbing start; Stage 3: first psychosomatic symptoms; Stage 4: management errors and abuse; Stage 5: serious worsening of mental and physical health of the victim; Stage 6: exclusion from the labor market. Mobbing is defined as an "organizational virus" for its ability to create vicious circles within the organization, poisoning the climate and work relations, and contributing to lower performance; it can also produce new cases of mobbing and generate other unfair actions (Giorgi & Majer, 2009; Spector & Fox, 2010). Beyond the motivations that can push individuals to harass a colleague or an employee (fear of losing their position, mutual intolerance, envy, career associated anxiety, sexual overture), the climate that dominates in the workplace and the quality of relationships are essential aspects (Arthur, 2011; Caiozzo & Vaccani, 2010; Duffy & Sperry, 2012). The cultural medium for mobbing is almost always a workplace where: a) there is no clear set of rules, on the

contrary changing and unpredictable standards of behavior apply; b) the rules are not complied with or compliance is expected for some but not for others; c) there are highly unsatisfactory work situations for which personal conflicts degenerate until up to the point of a lack of respect for people's right to their own dignity (Oliverio Ferraris & Oliverio, 2009). According to Leymann (1996), mobbing arises in organizations that do not know how to achieve a rational structure and do not ensure their employees' positive social relationships.

An organization is dysfunctional when conflicts erupt easily and their destructive force is not softened by any cushion. When conflicts are poorly managed, they steer toward courses of intensified violence and the psychological pressure that weighs on all workers can lead to identifying innocent victims (Verdarelli, 2013). The aspect of legitimizing violence was studied by Einarsen (2000): mobbing spreads more easily if tolerating harassment of any kind is part of the organizational culture. By not punishing violence, the organization is, in fact, legitimizing it, not sanctioning it, authorizing anyone to practice it, an extremely negative signal that can be picked up by anyone capable of reading it (Verdarelli, 2013). By not perceiving the risk of being condemned and punished, the attacker perpetrates negative actions more freely, while the victim may face persecution (Giorgi & Majer, 2009; Zapf & Einarsen, 2011). The "targeted ones" inevitably conclude that the mobbers do not value their relationship with them. Otherwise, they would not be systematically tortured. So, jointly to the feelings of fear and humiliation that accompany these victims, the perception of being rejected is also usually manifested (Zamperini & Menegatto, 2015).

According to Depolo (2003), the organizational climate (team size, communication, leadership, and freedom) has a substantial connection with the existence of episodes of mobbing. In other studies, some aspects in the job design, such as conflicts and role ambiguity, were found to have a direct effect on the perception of mobbing (Notelaers, Einarsen, Vermunt, & De Witte, 2005). As for conflicts, Tessarolo (2007) argues that possessing a "quarrel culture" does not mean developing new ways to react to conflict, rather picturing it better. Making it clear would allow us to manage it without ideologies, de-moving and facing it properly. The variables that are prerequisites for mobbing would be wiped out.

Mobbing is not a passing and temporary hostile situation or a short period of crisis, rather a long and painful experience that can last years, inevitably affecting the immediate family of the victim (Favretto, 2005). Since the family's stamina tends to run out (Ege, 2001), it gets tired, worn out by the hardship and suffering of the mobbing victim. When the family reaches saturation point and a crisis must be faced, the family ceases, more or less consciously and suddenly, to provide support to the victim, now seen mostly as a threat to the family balance and harmony. This may determine double mobbing (Favretto, 2004).

The consequences a case of mobbing can have on a personal level are somehow difficult to define. Given the clinical profile a victim may present, there is a variety of different symptoms and consequences on a personal level: anxiety problems, post-traumatic stress disorder, adjustment disorders, psycho-physiological balance disorders, behavioral disorders, and social balance disorders (De Carlo, Falco, & Capozza, 2013).

Mobbing, especially when perpetrated for a prolonged period of time, may also imply learnt helplessness, a condition that occurs when the subject is protractedly exposed to a stressful situation from which there is no way out (Alloy, Kelly, Mineka, & Clements, 1990). Following this line, Scott and Stradling (1994) proposed an additional diagnosis, Prolonged Duress Stress Disorder (PDSD), which could be considered for the symptoms some individuals exposed to

stress manifest, caused by one or more stressors of a relatively lower intensity in comparison to typical traumatic events.

From occupational stress and before culminating into mobbing, one can experience a condition of *straining,* that is, a type of forced stress, superior to that related to the nature of work and intentionally directed against a victim or group of victims in a discriminatory manner, so as to cause a permanent worsening in their work life, even before deteriorating their physical and mental condition (Gullotta, 2007). Typical straining actions are often the same as mobbing, yet without a strong harassing or vexatious content, rather they are aimed at establishing disparities in the workplace by means of systematic isolation, change of duties (resorting, in particular, to assigning meaningless or irrelevant tasks), demotion, confinement in remote work stations, removal of the working tools (Tronati, 2008).

In particular, Tronati (2011) argues that the problem must be tackled on a collective and trade union base and not alone, avoiding resignation. On the contrary, the victim must not feel guilty, remain passive, but rather react properly to mobbing from the very beginning, preferably by contacting the union. It is also important to record and collect any single evidence of oppressive behavior chronologically and treat any associated diseases. In this direction, support services for victims of mobbing have spread within local and institutional contexts, especially since the year 2000. Their core activity is to listen and support and they represent a response to mobbing actions in the workplace. This study analyzes the experience at the anti-mobbing window of a trade union in a Northern Italian city, where the issue of psychosocial discomfort in the workplace is central to their service. The window started operating in 2006 and over the years has responded to more than 300 calls for help from workers.

THE STUDY

Method: Assessing the Damage Caused by Mobbing

Participants and Procedure

Determining a mobbing experience involves two important aspects that are essential to one another: grasping the profound meaning of human experience and representing this experience in a clear and unequivocal manner. The procedure followed by the window consists of an interview and the administration of the Leymann Inventory of Psychological Terror (LIPT; Leymann, 1990), in Ege's (2002) adaptation (LIPT Ege). In this work, the responses given to LIPT Ege are analyzed.

The study involved 106 people who turned to the window between 2006 and 2015 for mobbing-related issues: 33 males (31%) and 73 females (69%); 51% were aged between 45 and 65 years, 35% between 35 and 45 years, 14% between 25 and 35 years. They were employed in the private (75%) and public sector.[3]

The interview was the first step for assessing the case, allowing us to grasp the profound meaning of the person's experience. Being the first form of contact, it also served the purpose of discriminating between possible situations of mobbing and other kinds of discomfort. Indeed, the interview sets up a relationship of "shared sense" that allows the person to share, even overcoming that sentiment of shame victims of harassment normally experience, not only the events that

have painfully marked their working life, but also the deep emotions underlying these events. One of the deep needs of the victims of harassment is to be heard, understood, and acknowledged. The listening part itself also allows us to grasp the profound meaning of this experience.

The next step is to compile the LIPT Ege.[4] The LIPT, elaborated by Leymann in the early '90s, is considered the precursor of all mobbing questionnaires. From the beginning, Ege (1997) considered it inapplicable to the Italian reality and brought some translation changes, resulting in the Italian version called "modified LIPT." Since 1998, the questionnaire has undergone a number of changes and subsequent extensions, due to the various needs arising as the research on mobbing evolved. A conversation with a psychologist may follow to integrate and clarify the responses collected and to acquire the elements of the subjective perception related to the actual job story. The LIPT Ege is the most suitable tool to check the seven basic parameters for identifying mobbing and its extent: it was, in fact, specially tailored to these two objectives.

The LIPT Ege instrument consisting of 30 questions is divided into three sections.[5] The first section concerns the framing of the subject at a personal and professional level (gender, date of birth, position within the company, sector, company type, professional qualification, annual income). The second section investigates the type of hostile action suffered (attacks on human contacts, systematic isolation, changes in work tasks, attacks on reputation, violence and threats, other hostile actions). The third section analyzes the consequences for the subject, as a result of the hostile treatment and persecution, the symptoms experienced, the impact on family life, self-confidence, and self-esteem. In the questionnaire, a specific question (number 21) covers the direct consequences on the worker's mental and physical health: the symptoms mentioned, mostly of a psychosomatic origin, are largely taken from the original LIPT (Leymann, 1990) and are systematically involved in mobbing. The additions relate mainly to sexual issues and allergic reactions and/or skin breakouts. In this third section, there is a new part in relation to the original LIPT, in which questions are developed to obtain specific information about the victim, on his/her psychological state and his/her private life. The purpose of these questions is twofold: on one hand to evaluate, if present, the extent of the individual's loss of self-esteem (his/her capacity to face new challenges, expectations for the future), on the other hand, to determine how rapidly mobbing at work has had an impact on private and family life (double mobbing). All the questions (from 22 to 30) are functional to both purposes.

Results

The answers given to the LIPT Ege were statistically processed using SPSS. In this contribution, we consider: a) the descriptive statistics related to the main questions in the three sections; b) the differences (ANOVA, *t*-test) among various subsamples referring to the total mobbing score; and c) the correlations (Pearson *r*) between the most salient items included in the tool. Results are presented and discussed referring to the three sections of LIPT Ege.

Section One

Confirming what Ege (1997) defined as an apparently typical "white collar plague" in Italy, also our sample is mostly made up of this category of workers (employees: 52.9%; execu-

tives: 4.7%; and blue-collar workers: 42.4%). The average score for mobbing[6] in our sample was 34.29 (*SD* = 20.70), the most affected areas being industries and the administrative sector, followed by schools, healthcare organizations, and banks. We also observed some significant differences among the various departments in companies, $F(4, 97) = 4.35$, $p = .003$. Particularly in the "services" department, mobbing scores were higher than in other departments, as already pointed out by Ege (1997). The workers belonging to a union who turned to the anti-mobbing window were 78% of our sample; members seem to have more information and access to the service offered by the union, while we cannot assert that being a union member can constitute a critical variable for a mobbing attack. Conflicts are developed mainly in private companies (44%) and in the goods and services sector (48.1%), with a total number of employees of less than 50 people.

Section Two

The results of the second section of the questionnaire showed that 97.2% of the participants claimed to have suffered attacks against their reputation; 77.4% reported problems in human contacts and communication at work; 76.4% reported changes in work tasks, while 55.7% reported violence or threats of violence. The high percentage of attacks against one's reputation confirms what is described in the literature (Ege, 2002).

The difficulties and obstacles related to humiliations even emerged from reports[7]: "when company memos arrive, my supervisor meets up with my colleagues to share their interpretations and I'm absolutely excluded . . ." Or as in another case "my colleagues make my workday hell, with their arrogant manners, shouting, words, insults, their way of discriminating against me in front of everyone, picking on me for anything, even something irrelevant, wearing me down . . . almost all my colleagues have no respect and consideration for what I do anymore, my decisions are constantly being criticized."

The frequency and duration of harassing actions were required afterward. About the frequency, 46.2% reported to have suffered hostile actions on a daily basis, 36.8% almost daily, 1% weekly, 8.5% rarely, while 7.5% a few times a month. The key element of the "repetitiveness of the attacks" and the fury with which the mobber seeks to eliminate the victim are the main parameters by which an action of mobbing is described. The criterion of the frequency of attacks marks the border between mobbing and other phenomena related to organizational discomfort such as, for example, straining. About the duration, data show that in 29.2% of the cases the actions occurred over a period of one and two years, in 27.4% over a period of two and five years, while in 11.3% for more than five years. Finally, 21.7% over a period of six and 12 months and 10.4% for less than six months.

With regard to the hierarchical level, in 74.5% of cases "vertical mobbing" occurred, that is, the mobber was a person in a higher hierarchical position than the victim, and, in 48.1% of cases, the mobbers were between two and four people. The reasons for which the person believed to have suffered mobbing actions can be summarized as follows: in 66.0% of the cases, the goal was to get rid of the worker by inducing him/her to resign; in 25.5% of the cases mobbing was expressed through attacks against the person (the victim believed that the cause of hostile actions was triggered by hostility, expressed in the form of envy, jealousy, and anger) or the organization's rules (the victim felt to have questioned some "unspoken" rules within the organization). Finally, 8.5% of the cases was attributed to issues related to injuries or illness.

Section Three

Section three concerns the consequences the person suffered as a result of hostile and harassing treatment at work. About looking for a new job after being mobbed, 62.3% of the people declared "the idea creates anxiety, fear, and concern," while 34% answered "no one would hire me." Following, responses such as "I would not stay too far from home" (31.1%), "I do not feel I can start all over again" (27.4%), "I do not know if I could do my best" (25.5%), are expressions of insecurity and fear of change.

Being mobbed pushes an individual to describe him/herself as "sad and depressed" (68.9%), "everything seems absurd" (47.2%), "I feel useless" (39.6%), "I have to try hard to do something" (35.8%), "I do not have much hope for the future" (34.9%), "I feel guilty" (29.2%). Specifically, for the answer "everything seems absurd" (M = 39.23, SD = 23.16) there was a significant difference, $t(89.76) = 2.34$, $p < .05$, compared to the average score of mobbing (M = 34.29, SD = 0.70). The fact that the answer to this item was linked to a higher rating can be explained in the light of the rapid change of life that mobbing inflicts on the victim, who cannot believe it is actually happening.

A second analysis shows some significant correlations between what the person thinks of him/herself and possible ways to find a new job. In detail, the following items are relevant: "I do not have much hope for the future" (Question 23) and "looking for a new job would be a disaster" (Question 22) ($r = .20$, $p < .04$). Another interesting correlation concerned the items "I am sad and depressed" (Question 23) and "I do not dedicate any time to family and/or friends anymore" (Question 28) ($r = .31$ $p = .001$), confirming that the vortex of mobbing can have an impact on the person and his/her family at the same time. As for the actions that the victim is willing to implement, 69% claimed to be likely to quit and 66% planned to start a lawsuit and press charges against the mobber.

Another worrying figure that results from the descriptive statistics is the high percentage of people who claimed to take sleeping pills (63.2%), anxiety pills (62.3%), and antidepressants (47.2%) as a result of conflicts at work: "I began not to sleep at night and to have actual nightmares in which I dreamt of being chased by my employer and my colleagues . . . the very idea of going to work made me feel very sick." Following this situation, 79.2% reported being absent from work for a state of anxiety, 52.8% for depression, while only 18% of the respondents did not avail themselves of any sick leave.

Conflict in the workplace linked to mobbing created a number of consequences on people that were manifested through a change in some behaviors: 69% said they smoked more, or had started smoking; 52% drank more alcohol; 40% said they had become more aggressive than usual. In addition, eating disorders, such as eating more than usual (34.9%) or eating too little (29.2%), occurred. As is clear from an excerpt, "I could not digest my lunch, I was so nervous, I alternated between moments of convulsive hunger to having my stomach in a knot."

Determining the Mobbing Score and the Ascertained Damage

Once the preliminary stage determined a typical case of mobbing based on the seven parameters, we proceeded by assessing the resulting damage. This second phase followed three

consecutive steps: a) the score for mobbing; b) the partial established damage for mobbing; c) the total permanent established damage for mobbing. Using the Ege Italian model, the work history of the subject was then examined to determine the stage of the case. After that, the partial percentage of the established damage for mobbing was calculated.

The partial damage for mobbing must be added, where feasible, to two percentage increases. The first increase relates to self-esteem, meaning the prejudice on the opinion that the individual has of him/herself and his/her skills. It also includes the sphere of expectations and mental attitudes toward the future, of great importance if compared to the possibility of outplacement. The second increase refers to the particular situation of double mobbing, that is, the negative impact of mobbing on the victim's private and family life, which is of a typically existential nature. The repercussions of work on the family emerge primarily through more frequent quarrels at home (45.3%) and the manifestation of sexual problems in married life (35.8%); 33% feel misunderstood at home, while 25.5% reported to have faced problems related to divorce/separation. Regarding conflict in the family and the search for a new job, a significant correlation between the items "I would not stay too far from home" and "Often quarrel with my family" ($r = .25$, $p < .02$) is found, as is clear from these answers: "Even my relations at home with my husband and my son have worsened" or "I had a tremendous psychological breakdown, I went through a very, very, very hard period of total and absolute apathy toward the world and toward everyone."

The total score is the sum of the two increases and the partial established damage for the previously calculated mobbing percentage, reaching the final percentage of the damage suffered by the subject due to mobbing. Having established the permanent total damage for mobbing, the quantification of compensation that will be relevant to the damage suffered by the subject is worked out. The result is called "damage for mobbing," which is equivalent to the economic loss resulting from the reduction of one's specific work skills, as well as any possibility of outplacement after judicial proceedings.

Discussion

Mobbed workers, whatever the harassing strategy is, will mainly suffer damage from the professional point of view, whether it is damage to their image, missed promotions, demotion or an improper removal. The damage will invest, above all, their economic security and will have asset and income effects (Ege, 2002). The close dependence between the damage from mobbing and the subject-victim's professionalism causes a potential financial loss that occurs through the loss of skills acquired or acquirable. A good proportion of mobbed workers were unwillingly forced to resign, to apply for mobility or early retirement with subsequent difficulties in re-inventing themselves in their profession, to cope with forced unemployment, to enter the labor market from square one again, or to wait to be eligible for an often still distant retirement.

The relationship between work and psychological distress has been investigated by several studies that have shown that working conditions can be considered among the main variables related to mental health (Wall et al., 1997) and that mobbing mainly affects supervisors and office staff, while executives appear to be less involved in the phenomenon. In most cases, there is

"vertical mobbing," where it is the leader or "the chief" who fills the role of the mobber toward his/her employees.

A very high percentage of the sample reveals to have experienced attacks on their reputation during the harassment. Undermining one's credibility and reputation seems to be the most frequent and sneaky behavior for a mobber. This study shows a detailed analysis of the explanation that victims give of the attacks. For our sample, harassment primarily originates from the desire to oust the victim from the workplace by trying to get him/her to quit. According to this interpretation, mobbing would be a behavior aimed at getting a result and would not merely consist of pure hostility and aggression poured onto the victim. It is otherwise thought that hostile manifestations may be caused by recent requests for sick, maternity, and parental leave. Other reasons mentioned by the victims refer to feelings of envy, jealousy, and anger that the mobber may hold against them.

The impact on the victim may be of different degrees involving various spheres of existence. However, what seems to emerge from the results is that mobbing can destroy the personal resources that are necessary to resolve the conflict. In addition, it may be noted that the effects of mobbing are not only limited to the negative consequences on the current state of the subject, but they may also affect his/her possibilities and future choices. It results from both descriptive statistics and, above all, from the correlations estimated with the final mobbing score. The damage that is reflected on the future following harassment is comprised in the methodology prepared by Ege to assess professional and financial damage. The research also found the effects of occupational stress on family life, which mainly occur in the form of difficulties in the couple, quarrels, and misunderstandings.

Finally, this research shows how the anti-mobbing window is a type of first aid in the territory for dealing with such forms of unease making use of qualified consultation, both psychological and legal. Being able to go to a window that first allows the person to share his or her discomfort, be listened to and understood and then subsequently supported, using specific mobbing assessment tools, has proven to be a precious opportunity for workers.

Overall, mobbing is configured as a complex phenomenon in which we can trace a number of variables related to the organization, to the characteristics of the victim and the mobber, as well as the social group. Because mobbing can be triggered by a large number of variables, it should be investigated with a multidisciplinary contribution by using a precautionary approach. Failure to promptly recognize and fight the contrast of variables that trigger mobbing can result in significant consequences on the welfare of workers in organizational contexts. In conclusion, we can affirm that the LIPT, in the version updated by Ege (2002) and adapted to Italy which we used in this research, allows us to identify a series of indices expressing the unease experienced by the subject, which will enable the consultant not only to identify the mobbing in a precise manner, but also to quantify the damage it caused.

NOTES

1. Newspapers show limited interest in the topic, whereas the judicial phase, compensations, and convictions are reported the most.
2. For instance: RAI Television, TG2 News dated February 17th 2015 and national newspaper – Il Sole 24 Ore – dated June 11th 2015 reported court sentence no. 22635/2015.

3. About 70% of the people who compiled LIPT then turned to the union's legal office or to a lawyer. They did not always opt for a lawsuit.
4. LIPT Ege was published for the first time for educational and explanatory purposes in "The expert assessment of the damage of mobbing" (Ege, 2002). Although both "LIPT Ege modified" and "LIPT Ege" clearly derive from Leymann's work, they are exclusive and original questionnaires, protected by copyright and subject to the Italian copyright law (LDA No. 633/41 and subsequent modifications). The tool can only be used by psychologists who have attended a specific training course with the author in order to learn the correct scoring method.
5. Each question has a multiple choice answer; more than one answer can be given.
6. To assess a mobbing score, the following four parameters derivable from LIPT Ege need to be considered: a) the number of categories of hostile actions against the victim (C, Question 13), b) the frequency rate and/or the systematic nature with which these hostile actions have been perpetrated over time (F, Question 15), c) the chronological data concerning the duration of mobbing (D, Question 16), d) the subject's income (R, Question 7). Once these four factors C, F, D, and R have been acquired, the mobbing score can be calculated. The mobbing score is the product of the four indexes.
7. We report some excerpts from the interviews.

REFERENCES

Alloy, L. B., Kelly, K. A., Mineka, S., & Clements, C. M. (1990). Comorbidity in anxiety and depressive disorders: A helplessness-hopelessness prospective. In J. D Maser & C. R. Cloninger (Eds.), *Comorbidity in anxiety and mood disorders* (pp. 499-544). Washington, DC: American Psychiatric Press.

Arthur, J. B. (2011). Do HR system characteristics affect the frequency of interpersonal deviance in organizations? The role of team autonomy and internal labor market practices. *Industrial Relations*, *50*, 30-56. doi:10.1111/j.1468-232X.2010.00624.x

Caiozzo, P., & Vaccani, F. (2010). *Le cause organizzative del mobbing. Se il malato fosse l'organizzazione*? [The organizational causes of mobbing. If a sick person is the organisation?]. Milano, Italy: FrancoAngeli.

De Carlo, N. A., Falco, A., & Capozza, D. (2013). *Stress, benessere organizzativo e performance. Valutazione & intervento per l'azienda positiva* [Stress, well-being and organizational performance. Assessment and intervention for positive company]. Milano, Italy: FrancoAngeli.

Dejours, C. (1998). *Souffrance en France* [Suffering in France]. Paris, France: Editions du Seuil.

Depolo, M. (2003). *Mobbing: quando la prevenzione è intervento. Aspetti giuridici e psicosociali del fenomeno* [Mobbing: When prevention is intervention. Legal and psychosocial aspects of the phenomenon]. Milano, Italy: FrancoAngeli.

Duffy, M., & Sperry, L. (2012). M*obbing: Causes, consequences, and solutions*. New York, NY: Oxford University Press.

Ege, H. (1997). *Il mobbing in Italia* [Mobbing in Italy]. Bologna, Italy: Pitagora.

Ege, H. (2001). *Mobbing. Conoscerlo per vincerlo* [Mobbing. Know it to win it]. Milano, Italy: FrancoAngeli.

Ege, H. (2002). *La valutazione peritale del danno da mobbing* [The expert assessment of the damage of mobbing]. Milano, Italy: Giuffrè.

Einarsen, S. (2000). Harassment and mobbing at work: A review of the Scandinavian approach, *Aggression and Violent Behaviour*, *5*(4), 379-401.

Favretto, G. (2004). *Lo stress nelle organizzazioni* [Stress in organizations]. Bologna, Italy: Il Mulino.

Favretto, G. (2005). *Le forme del mobbing, cause e conseguenze di dinamiche organizzative disfunzionali* [The forms of mobbing, causes and consequences of dysfunctional organizational dynamics]. Milano, Italy: Raffaello Cortina Editore.

Giorgi, G., & Majer, V. (2009). *Mobbing: virus organizzativo. Prevenire e contrastare il mobbing e i comportamenti negativi sul lavoro* [Mobbing: Organizational virus. Prevent and combat mobbing and negative behaviors at work]. Firenze, Italy: Giunti OS.

Gullotta, G. (2007). *Il vero e il falso mobbing* [The true and the false mobbing]. Milano, Italy: Giuffrè Editore.

Leymann, H. (1990). *Manual of the LIPT questionnaire for assessing the risk of psychological violence at work*. Stockholm, Sweden: Violen.

Leymann, H. (1996). The content and development of mobbing at work. *European Journal of Work and Organizational Psychology*, *5*(2), 165-184.

Maeran, R., & Bavarone, A. (2007). *Il mobbing: come il fenomeno viene proposto dai quotidiani* [Mobbing: How the phenomenon is brought by the newspapers]. Bologna, Italy: Pàtron.

Notelaers, G., Einarsen, S., Vermunt, J., & De Witte, H. (2005). Clinical definitions determining the size of bullied workers versus a data driven estimation with latent cluster analysis. In J. Houdmont & S. McIntyre (Eds.), *6th Annual Conference of the European Academy of Occupational Health Psychology: Proceedings. 24-26 November 2004* (pp. 325-334). Castêlo da Maia, Portugal: Publismai.

Oliverio Ferraris, A., & Oliverio, A. (2009). *Capire il comportamento* [Understand the behavior]. Bologna, Italy: Zanichelli.

Scott, M. J., & Stradling, S. G. (1994). Post-traumatic stress disorder without the trauma. *British Journal of Clinical Psychology*, *33*, 71-74. doi:10.1111/j.2044-8260.1994.tb01095.x

Spector, P. E., & Fox, S. (2010). Theorizing about the deviant citizen: An attributional explanation of the interplay of organizational citizenship and counterproductive work behavior. *Human Resource Management Review*, 20, 132-143. doi:10.1016/j.hrmr.2009.06.002

Tronati, B. (2008). *Mobbing e straining nel rapporto di lavoro* [Mobbing and straining in the employment relationship]. Roma, Italy: Ediesse.

Tronati, B. (2011). *Stop a mobbing, straining e stress lavoro-correlato. Guida teorico-pratica sul disagio lavorativo* [Stop mobbing, straining, and work-related stress. Theoretical and practical guidance on occupational stress]. Roma, Italy: Ediesse.

Tessarolo, M. (2007). La comunicazione interpersonale [Interpersonal communication]. Bari, Italy: Laterza.

Verdarelli, P. (2013). *Il rischio psicosociale nelle organizzazioni* [Psychosocial risk in organizations]. Roma, Italy: Carrocci.

Wall, T. D., Bolden, R. I., Borrill, C. S., Carter, A. J., Golya, D. A., Hardy, G. E., . . . West, M. A. (1997). Psychiatric disorder in NHS trust staff: Occupational and gender differences. *British Journal of Clinical Psychology*, *171*(6), 519-523.

Williams, K. D. (2001). *Emotions and social behavior. Ostracism: The power of silence*. New York, NY: Guilford Press.

Zamperini, A., & Menegatto, M. (2013). La violenza collettiva e il G8 di Genova. Trauma psicopolitico e terapia sociale della testimonianza [Collective violence and the Genoa G8 Summit. Psychopolitical trauma and social therapy of testimony]. *Psicoterapia e Scienze Umane*, XLVII, 3, 423-442. doi:10.3280/PU2013-003002

Zamperini, A., & Menegatto, M. (2015). Giving voice to silence: A study of state violence in Bolzaneto prison during the Genoa G8 Summit. In F. D'Errico, I. Poggi, A. Vinciarelli, & L. Vincze (Eds.), *Conflict and multimodal communication. Social research and machine intelligence* (pp. 185-205). Heidelberg, Germany: Springer.

Zapf, D., & Kulla, M. (1996). On the relationship between mobbing factors and job content, social work and health outcomes. *The European Work and Organizational Psychologist*, *5*(2), 215-237.

Zapf, D., & Einarsen, S. (2011) Individual antecedents of bullying: Victims and perpetrators. In Einarsen S. (Ed.), *Bullying and emotional abuse in the workplace: International perspectives in research and practice* (pp. 177-200). London, UK: Taylor & Francis.

TPM®

MEASURING FOUR-DIMENSIONAL ENGAGEMENT IN SCHOOL: A VALIDATION OF THE STUDENT ENGAGEMENT SCALE AND OF THE AGENTIC ENGAGEMENT SCALE

CONSUELO MAMELI
STEFANO PASSINI
UNIVERSITY OF BOLOGNA

Student engagement has traditionally been defined as a metaconstruct made up of three aspects: emotional, behavioral, and cognitive. Recently, however, a fourth component has been proposed, namely the agentic one. Notwithstanding the widely recognized importance of investigating student engagement, in Italy a validated scale suitable for this purpose does not exist. The present work represents a first contribution to the validation in Italy of a questionnaire designed to measure student engagement in high school. By combining two separate questionnaires (i.e., the Student Engagement Scale and the Agentic Engagement Scale) within a single instrument, this study focuses on the psychometric properties of a four-dimensional student engagement scale on 1,210 Italian secondary school students. Results confirm the robustness of the four-dimensional structure of the student engagement scale.

Key words: Student engagement; Agentic engagement; Validation; Measurement; Assessment instruments

Correspondence concerning this article should be addressed to Consuelo Mameli, Department of Education Studies, University of Bologna, Via Filippo Re 6, 40126 Bologna (BO), Italy. Email: consuelo.mameli@unibo.it

Over the past twenty years the concept of student engagement, recently defined as energy in action (Ainley, 2012; Skinner & Pitzer, 2012), has received increasing interest worldwide (Appleton, Christenson, & Furlong, 2008; Lawson & Lawson, 2013; Shernoff & Schmidt, 2008). A number of studies converge in identifying engagement as a crucial factor in predicting learning and academic success (e.g., National Research Council & Institute of Medicine, 2004), and some researchers indicate that it functions as a full mediator between intrinsic motivation and school achievement (Reeve, 2013). There is also evidence that adolescents showing low engagement levels tend to exhibit risky behaviors (e.g., a more frequent use of psychoactive substances), and they are more likely to dropout of school (Li & Lerner, 2011; Wang & Fredricks, 2014).

Besides a large amount of empirical evidence, there are other reasons — theoretical, practical, and historical — that have contributed to directing scholars' attention toward this issue (Fredricks, Blumenfeld, & Paris, 2004). From a theoretical point of view, student engagement is fascinating because, while still involving the individual, it has the nature of a dynamic, social, and synergistic process (Lawson & Lawson, 2013). In its most recent conceptualizations, engagement corresponds to a metaconstruct defined and incessantly redefined within social contexts and interpersonal relationships (Peck, Roeser, Zarrett, & Eccles, 2008; Wang & Eccles, 2013) which influence the degree to which a student is able to take action in terms of school commitment and effort (Pianta, Hamre, & Allen, 2012).

TPM Vol. 24, No. 4, December 2017 – 527-541 – doi:10.4473/TPM24.4.4 – © 2017 Cises

This characterization has important implications for professional practice because it qualifies engagement as a malleable and evolving dimension (Crick, 2012; Fredricks et al., 2004), liable to change by intervening in the learning environment and the relationships involved in it (Borman, Hewes, Overmann, & Brown, 2003). Furthermore, this ductility makes student engagement a relevant variable upon which to intervene especially in this historical moment, characterized by a general disaffection of young people vis-à-vis school (Crosnoe, 2002; Modell & Elder, 2002).

All these reasons signal the importance of thoroughly understanding and investigating student engagement. To do this, standardized instruments are fundamental to identify and possibly intervene in critical situations and contexts. Although in the literature there are several of these instruments, as far as we know in Italy a validated scale suitable for this purpose does not exist. This lack is worrisome considering the critical situation of this country from an educational point of view. The report by the Ministry of Education, University, and Research (2013) states that school dropout rate in this country corresponds to 19.2% (mostly males), and this datum places Italy in a significantly retarded position, specifically fourth from last with respect to the other European countries. In addition, the results of the international tests OECD-PISA (2012) on reading, writing, and mathematical skills showed that Italian students are in the lower positions of the ranking, although there have been some improvements with respect to the same survey conducted in 2009.

In light of these considerations, the present work represents, to our knowledge, a first contribution to the validation in Italy of a questionnaire appropriate for measuring student engagement in high school. In this study, we consider an engagement theoretical model comprising four dimensions: affective, behavioral, cognitive, and agentic.

Engagement as a Three-Dimensional Construct

Scholars' interest in student engagement has been proportional to the effort to clarify its definition and to design tools with which to evaluate it. The concentration of studies, however, has led to a variety of solutions, resulting in some confusion. Two areas result to be particularly critical (Fredricks & McColskey, 2012; Glanville & Wildhagen, 2007; Lam et al., 2014; Lawson & Lawson, 2013): the first concerns the number of dimensions which fall within the concept of engagement and their definition, while the second relates to the measurement of each one.

As for the number of dimensions, the most recent lines of research converge in defining student engagement as a metaconstruct made up of three main aspects (Fredricks et al., 2004; Jimerson, Campos, & Grief, 2003; Wang & Fredricks, 2014): emotional or affective, behavioral, and cognitive. Emotional engagement corresponds to student identification and the sense of belonging to school (Finn, 1989; Voelkl, 2012), and to the affective feelings about learning and educational activities (Finn & Zimmer, 2012). Other scholars (Fredricks & McColskey, 2012), however, also include in this dimension the set of positive or negative emotions that students experience with respect to their teachers and classmates. Behavioral engagement is defined as student participation and involvement in curricular (Fredricks et al., 2004) and extra-curricular (Finn, Pannozzo, & Voelkl, 1995) activities. Some researchers (e.g., Finn & Rock, 1997) also include student discipline in this component, meaning their ability to keep to school rules. Never-

theless, the latter aspect is controversial because it is unclear whether student conduct should be considered as an engagement indicator or an outcome (Lam et al., 2014). Finally, cognitive engagement — which represents the weakest and most elusive construct component (Wang & Fredricks, 2014) — is conceptualized as the students' degree of investment in learning processes and strategies. Although some scholars link this element to the ability to self-regulate learning processes (Cleary & Zimmerman, 2012; Walker, Greene, & Mansell, 2006), others argue that self-regulation strategies have a behavioral nature and should not therefore be included in this dimension (Lam et al., 2014).

As regard the student engagement measurement, self-report instruments are those most commonly used due to ease and speed of administration. As pointed out in some literature reviews (Fredricks et al., 2004; Reeve, 2012), however, most of the available tools present some drawbacks. Some instruments, for instance, assessed engagement by means of a unique and general scale, thus failing to distinguish between behavioral and psychological aspects (Marks, 2000). Other questionnaires were instead focused on a single component of engagement, such as the cognitive (Pintrich, Smith, Garcia, & McKeachie, 1993) or the emotional (Voelkl, 2012) one. Other instruments have attempted to measure all three construct components (e.g., Fredricks et al., 2004), although differences have been observed in the way similar items were adopted as indicators of different engagement elements. Finally, some scales confused engagement indicators with other variables that could rather be considered as antecedents, such as the quality of students' social relations (Appleton, Christenson, Kim, & Reschly, 2006), or outcomes, such as conduct problems (Wang, Willet, & Eccles, 2011), of engagement.

Seeking to overcome these limits, Lam and colleagues (2014) conducted an international study in twelve countries (Austria, Canada, China, Cyprus, Estonia, Greece, Malta, Portugal, Romania, South Korea, the United Kingdom, and the United States of America) on nearly 3,500 students from 7th to 9th grade in order to clarify the concept of student engagement and to build up a questionnaire appropriate for measuring its three components in different student populations. The developed 33-item scale presents several strengths. First of all, it includes items specifically selected by an international research team and based on an extensive review of relevant past studies. Second, the items were singled out according to a clear definition of the three engagement components: the affective engagement subscale assesses student liking for learning and school; the behavioral engagement subscale measures student effort in learning and participation in school and extrascholastic activities; the cognitive engagement subscale evaluates student use of meaningful information-processing strategies in learning. Third, the items were chosen with the specific aim of avoiding confusion between the actual engagement indicators and their antecedents or outcomes. Fourth, the scale showed good psychometric properties for the international samples where it was tested, and this makes the questionnaire potentially suitable for its use in different contexts, including the Italian one.

Agency as the Fourth Component of Student Engagement

As previously stated, the three-component nature of student engagement is largely shared in the literature. Recently, however, Reeve (2012, 2013; Reeve & Tseng, 2011) suggested adding

a fourth component to these three dimensions, namely agentic engagement, which has given rise to a certain amount of attention in the academic debate (Lawson & Lawson, 2013).

Within the theoretical framework of Self-Determination Theory (Ryan & Deci, 2000, 2002; Vansteenkiste, Niemiec, & Soenens, 2010), and in particular of the student-teacher dialectic framework, Reeve (2012) stressed the importance of considering engagement as a social process built into the interpersonal contexts in which the individual (i.e., the student) actively participates. Briefly, this theoretical framework emphasizes the circular and bidirectional nature of the factors related to the engagement construction. On the one hand, the learning environment may be more or less supportive and can contribute in different ways to shaping intrinsic student motivation, the result of which is reflected in the engagement dimension. For instance, an interpersonal context characterized by teachers who encourage student autonomy, competence, and positive relationships will facilitate high intrinsic motivation and may thus encourage a good level of engagement. On the other hand, Reeve and other authors (Ainley, 2012; Brooks, Brooks, & Goldstein, 2012; Crick, 2012; Mameli & Molinari, 2014) argue that a student does not simply *receive* these influences, but actively intervenes in the learning environment, thereby modifying it. For instance, students may ask questions if they did not understand a topic, they may express their own opinions or may ask to deepen an issue of particular interest to them. The agentic engagement is placed by Reeve in this action space, and is defined as "the process in which students proactively try to create, enhance, and personalize the conditions and circumstances under which they learn" (Reeve, 2012, p. 161).

According to this scholar, the behavioral, emotional, and cognitive engagement components so far conceptualized and measured in the literature are appropriate to assess the way in which students react to the activities and tasks proposed during the lessons, but fail to capture their active and transformative contribution. To understand and evaluate this aspect, Reeve and Tseng (2011) developed a 5-item scale (the Agentic Engagement Scale, AES), later modified and improved (Reeve, 2013). The scale used an agentic engagement definition based on five key points (Reeve & Tseng, 2011): it is proactive, intentional (deliberate and purposive), enriches and personalises the learning activity, contributes to the flow of the teacher's instructions, and does not indicate the teacher's ineffectiveness or incompetence. Despite its innovative and original potential, to date Reeve's scale has only been used in his works, based on moderate size samples of students from Taiwan and South Korea. To our knowledge, no studies have tested the validity of this scale on Western student populations.

THE PRESENT STUDY

The present study focuses on the psychometric properties of a four-dimensional student engagement scale, composed by combining two separate questionnaires into a single instrument. As for the first three dimensions — emotional, behavioral, and cognitive — we relied on the Student Engagement Scale as proposed by Lam and collaborators (2014), which showed good psychometric properties in a number of countries. As for the fourth dimension, that is agentic engagement, we used the Agentic Engagement Scale originally proposed by Reeve (Reeve, 2013; Reeve & Tseng, 2011).

Two specific objectives were pursued in this study. The first aim is to confirm the four-dimensional factor solution of the questionnaire. We therefore expect to find good reliabilities for

all the four dimensions as well as a confirmation of the enlarged structure of the concept of student engagement. The second goal is to analyze the concurrent validity of the questionnaire. Based on the vast existing literature, we chose three dimensions which were expected to show a high association with the four engagement aspects assessed. In particular, we expect student engagement to correlate positively with a good relationship with peers (Polychroni, Hatzichristou, & Sideridis, 2012; Ream & Rumberger, 2008) and academic achievement (Skinner & Pitzer, 2012; Wang & Holcombe, 2010), while we predict a negative association between engagement and psychological distress (Antaramian, Huebner, Hills, & Valois, 2010; Steele & Fullagar, 2009).

METHOD

Participants

Participants were 1,210 Italian secondary-school students (664 males, 543 females, and three persons who did not indicate their gender), coming mostly from middle-class families. They were enrolled in five academic and technical secondary schools located in Northern Italy. The average age of the participants was 16.08 (*SD* = 1.36, range 14-19) years. They were almost equally divided into students attending the first two years of compulsory schooling (aged 14-16, *n* = 540, 44.63%) and the last three years of high school (aged 17-19, *n* = 670, 55.37%). Almost all the participants were of Italian origin (*n* = 1,126, 93.1%), while the remaining students nonetheless spoke fluent Italian.

Procedure

For underage students, participation in the study was preceded by an informed-consent procedure that required active consent from both of the students' parents. Only the parents of two students denied their consent and were therefore excluded.

The questionnaires were distributed in an online version in the classrooms during school lab hours. In cases where this was not possible (*n* = 210), we proceeded with the administration of paper questionnaires that were then manually entered into the database. The research was introduced to the students as a survey on school experience and they were asked to verbally express their consent to take part in the study. The scholar explained the procedure and guaranteed confidentiality and anonymity. It took approximately 20 to 25 minutes to complete the questionnaire. This survey was approved by the Ethics Commission of the institution where the authors work and was conducted in agreement with the ethical norms laid down by the Italian National Psychological Association.

Measures

Affective, behavioral, and cognitive engagement. The questionnaire used is the one proposed by Lam and collaborators (2014). The 33 items making up the questionnaire were selected

by the authors from preexisting instruments widely used in the literature (e.g., Finn et al., 1995; Miller, Greene, Montalvo, Ravindran, & Nichols, 1996; Rao & Sachs, 1999; Skinner & Belmont, 1993).[1]

For the Italian version, the instrument was subjected to a back-translation (Brislin, 1970) by a native English speaker. The back-translated items were then reviewed by the authors and, where necessary, unclear statements were reformulated. In particular, the items that had ambivalent or confusing meanings in the Italian structure were reformulated paying attention to maintaining a conceptual equivalence (Herdman, Fox-Rushby, & Badia, 1997) with the original English item. For example, the original item "When I'm in class, my mind wanders" should be literally translated into Italian as "Quando sono in classe, la mia mente vaga." Since this formulation is somewhat confusing, this item was reformulated as "Quando sono in classe mi distraggo," which in English corresponds to "When I'm in class, I get distracted."

The questionnaire investigates student engagement by means of three scales: affective, behavioral, and cognitive. Affective engagement scale (nine items, one of which is reverse scored) measures students' liking for learning and school. Sample items are: "I am very interested in learning" and "I think learning is boring" (reverse). Behavioral engagement scale (12 items, three of which are reverse scored) assesses students' effort in learning and involvement in school and extrascholastic activities. Sample items are: "In class, I work as hard as I can" and "When I'm in class, I just act like I'm working" (reverse). The cognitive engagement scale (12 items) estimates students' use of significant information-processing strategies in learning. Sample items are: "When I study, I figure out how the information might come in useful in the real world" and "I make up my own examples to help me understand the important concepts I learn from school." For the first two subscales, students were asked to indicate their level of agreement on a 7-point Likert scale (from 1 = *strongly disagree* to 7 = *strongly agree*). For the cognitive engagement scale, a 7-point Likert scale of frequency was used (from 1 = *never* to 7 = *always*). The mean of the items on each subscale was used as an overall score on the corresponding dimension. Copies of the Italian questionnaire are available from the authors on request.

Agentic engagement. This dimension was measured with the Agentic Engagement Scale (AES; Reeve & Tseng, 2011) in its revised form (Reeve, 2013). As this instrument had never been used before in Italy, a back-translation procedure was adopted. The scale is made up of five items assessing students' contributions to education but also more transactional and dialectical inputs. Students were asked to indicate their grade of agreement on a 7-point Likert scale (range from 1 = *completely disagree* to 7 = *completely agree*). A sample item is "I let my teacher know what I need and want." Copies of the Italian version of this scale are available from the authors on request.

Connectedness among students. The students' perception of the connectedness among classmates was measured using the Connected Classroom Climate Inventory (CCCI; Dwyer et al., 2004). The students were asked to indicate to what extent they agreed with 18 statements (e.g., "The students in my class respect one another," "The students in my class are concerned about one another") on a 7-point Likert scale, from 1 = *strongly disagree* to 7 = *strongly agree*. The mean of the scores was used to indicate the students' perception of a classroom environment in which students feel socially connected through commonalities, a sense of community, and a mutual concern for each other. Cronbach's alpha for this scale was .93.

Psychological distress. The General Population Clinical Outcomes in Routine Evaluation measure (GP-CORE; Evans, Connell, Audin, Sinclair, & Barkham, 2005) was used to investigate nonclinical levels of distress. The GP-CORE is a 14-item instrument derived from the larger Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM; Evans et al., 2002). Items include statements such as "I have felt tense, anxious, or nervous" and "I have felt warmth or affection for someone" (reversed), with higher scores signifying higher levels of concern and distress. For this study, responses are evaluated on a 7-point Likert scale from 1 = *not at all* to 7 = *most or all of the time*. In the present work, Cronbach's alpha was .78.

School achievement. All students were asked to indicate their average school grade at the end of the first semester.

Analytical Procedures

First, the normality, the internal reliability, and the item analysis of each dimension of the student engagement scale were examined. In particular, as concerns normality of the scale, values of skewness and kurtosis were considered. Normality of the data is considered acceptable when skewness < |3.0| and kurtosis < |8.0| (Kline, 2011). Concerning the other psychometric properties, internal reliability > .70 (Cronbach & Meehl, 1955) and item-total correlations > .30 (Green & Lewis, 1986) are considered acceptable. Second, confirmatory factor analysis (CFA) was performed to confirm the structure of the scale. As suggested by Hu and Bentler (1999), model fit of CFA was assessed using the comparative fit index (CFI, cutoff value close to .90), the Tucker-Lewis index (TLI, cutoff value close to .90), and root mean square error of approximation (RMSEA, cutoff value close to .06). In particular, we examined two different multidimensional structures: the three-dimensional (affective, behavioral, and cognitive) and the four-dimensional (affective, behavioral, cognitive, and agentic) solutions. These structures were each time compared with the corresponding one-dimensional structure. To test significant improvement in model fit, the chi-square difference test was used to compare nested models. Finally, correlations of the dimensions of the student engagement scale were computed with the other variables in order to examine the concurrent validity.

Results

As can be seen in Table 1, the three original dimensions of the student engagement scale identified by Lam and colleagues (2014) showed good psychometric properties. Internal reliabilities of these three dimensions did not increase with the elimination of any item. Considering the agentic dimension, the analysis showed an acceptable internal reliability, although this dimension is composed by fewer items. Moreover, in both cases, the normality of the scale and interitem correlations were statistically acceptable.

The means of the dimensions showed that participants gave high scores to the affective and cognitive dimensions and medium scores to the behavioral and agentic ones. As concerns gender differences (see Table 1), in line with the literature (e.g., Skinner, Kindermann, & Furrer, 2009; Wang & Eccles, 2013), *t*-test analysis showed that girls had higher scores on the behavioral dimension.

TABLE 1

Means, standard deviations, psychometric properties, and gender differences on each dimension of the Student Engagement Scale

				Psychometric properties				*t*-test for gender		
	n item	*M*	*SD*	Skewness	Kurtosis	α	Inter-item *r* (range)	*M* girls	*M* boys	*t*
Affective	9	4.71	1.07	–.66	.26	.87	.32-.70	4.77	4.66	1.82
Behavioral	12	4.39	1.01	–.37	–.01	.87	.37-.68	4.59	4.22	6.47***
Cognitive	12	5.10	0.99	–.71	.64	.90	.44-.79	5.03	5.13	–1.82
Agentic	5	4.06	1.20	–.26	–.38	.78	.47-.65	3.93	4.17	–3.54***

*** $p \leq .001$.

Differently from other studies reporting no discrepancy between males and females in respect to the agentic engagement (Reeve, 2013; Reeve & Tseng, 2011), a difference in favor of males was found on this dimension. No differences were found on the affective and cognitive dimensions.

Then, in order to assess the structure of the entire scale, two distinct confirmatory factor analyses (three- and four-dimensional structures) were performed on the scale items. As can be seen in Table 2, these analyses confirmed the robustness of all the multidimensional structures. All the factor loadings were significant at $p < .001$ (see Table 3). In all the models, the same correlations between error terms were included.[2] These correlations were all between error terms of items loading on the same dimension and were all theoretically plausible given the very similar meaning and formulation of the associated items. For instance, the error term of the item co9[3] correlated with the error term of the item co10, with both items referring to the cognitive process of connecting new school information with past subjective experiences. Or else, the error term of the item ag3 correlated with the error term of the item ag1, with both items referring to a request for clarification to the teacher. In specific, four correlations were allowed on the affective dimension: af4-af6; af2-af6; af4-af7; af6-af7. Nine correlations were allowed on the behavioral dimension: be1-be5; be3-be6; be3-be12; be6-be12; be2-be7; be2-be4; be9-be11; be10-be11; be7-be11. Eight correlations were allowed on the cognitive dimension: co1-co2; co9-co10; co3-co11; co5-co11; co6-co11; co8-co12; co8-co9; co5-co6. Two correlations were allowed on the agentic dimension: ag1-ag3; ag4-ag5. Moreover, one correlation between error terms was estimated between two items of distinct dimensions, that is, be8 (behavioral) with ag1 (agentic). Also in this case, the theoretical closeness is plausible given that both the items refer to the active participation of the student in the classroom.

In general, the three-dimensional structure was the analysis with a modestly better fit. However, considering that the four-dimensional structure also had acceptable fit, and considering the opportunity of having a fourth dimension which also considers the aspects related to unilateral and original student contributions, the four-dimensional one was considered the best solution. In both cases, chi-square difference tests indicated a significantly better fit over the one-factor model: $\Delta\chi^2(3) = 2568.03$, $p < .001$ for the three-dimensional structure; $\Delta\chi^2(6) = 3009.02$, $p < .001$ for the four-dimensional structure.

In order to test how much each dimension is related to engagement in school, a second-order model was computed, with the four dimensions as first-order factors, and a student engagement latent variable as the higher-order factor. The contribution of each dimension on the student engagement factor was significant, $\chi^2(634) = 2315.32$; CFI = .90; TLI = .90; RMSEA = .047; affective = .85, $p < .001$; behavioral = .81, $p < .001$; cognitive = .60, $p < .001$; and agentic = .50, $p < .001$.

TABLE 2
Confirmatory factor analyses on one-, three-, and four-dimensional structures

Measures	χ^2	*df*	CFI	TLI	RMSEA
One dimension (33 items)	4180.39	474	.81	.79	.080
Three dimensions (33 items)	1612.36	471	.94	.93	.045
One dimension (38 items)	5354.73	641	.79	.77	.078
Four dimensions (38 items)	2345.71	635	.92	.91	.047

Note. CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation.

TABLE 3
Standardized correlations between factors and factor loadings for the four-dimensional structure of the Student Engagement Scale

	Affective	Behavioral	Cognitive	Agentic
Correlations				
Affective	–			
Behavioral	.71	–		
Cognitive	.50	.47	–	
Agentic	.40	.35	.43	–
	af1 = .73	be1 = .60	co1 = .68	ag1 = .42
	af2 = .68	be2 = .62	co2 = .77	ag2 = .69
	af3 = .75	be3 = .66	co3 = .58	ag3 = .51
	af4 = .55	be4 = .28	co4 = .57	ag4 = .68
	af5 = .54	be5 = .62	co5 = .42	ag5 = .75
	af6 = .55	be6 = .28	co6 = .51	
	af7 = .52	be7 = .70	co7 = .75	
	af8 = .74	be8 = .62	co8 = .72	
	af9 = .43	be9 = .54	co9 = .83	
		be10 = .76	co10 = .81	
		be11 = .73	co11 = .47	
		be12 = .29	co12 = .65	

Note. af = affective; be = behavioral; co = cognitive; ag = agentic. All the correlations and loadings have $p < .001$.

Finally, concurrent validity of the student engagement scale was analyzed by inspecting the correlations with the CCCI, the GP-CORE, and school achievement. In line with the hypotheses (see Table 4), affective engagement showed moderate positive correlations with CCCI and academic achievement, and a high negative correlation with psychological distress index (GP-CORE). Behavioral engagement showed a high positive association with academic achievement, a moderate positive correlation with the CCCI, and a moderate negative association with GP-CORE. Cognitive engagement was confirmed as the weakest component of the construct, showing on the whole, the lowest correlations with the dimensions of school achievement, connectedness among students, and psychological distress (the latter in a negative direction). Finally, the agentic engagement, to date the most unexplored component of the construct, showed the same directions of association as the other three dimensions. In particular, this aspect exhibited a moderate positive correlation with CCCI and a low positive association with school achievement. Furthermore, it presented a moderate negative correlation with the GP-CORE.

DISCUSSION

With this study, we presented a first contribution to investigate the psychometric properties of a four-dimensional student engagement scale. As shown by results, our attempt to refer to a four-dimensional student engagement scale combining the Student Engagement Scale and the Agentic Engagement Scale within a single instrument produced good results. Indeed, the findings

TABLE 4
Pearson correlation coefficients among all the variables

Measures	1	2	3	4	5	6	7
1. Affective	—						
2. Behavioral	.57***	—					
3. Cognitive	.39***	.45***	—				
4. Agentic	.33***	.38***	.37***	—			
5. GP-CORE	–.33***	–.16***	–.15***	–.22***	—		
6. CCCI	.27***	.14***	.12***	.27***	–.34***	—	
7. School achievement	.28***	.39***	.21***	.14***	–.16***	.01	—
8. Age	–.22***	–.04	–.01	.00	.06*	–.08**	.02

Note. GP-CORE = general population clinical outcomes in routine evaluation; CCCI = connected classroom climate inventory.
* $p < .05$. ** $p < .01$. *** $p < .001$.

from the confirmatory factor analyses show that both the three- and the four-dimensional structures have good fits. We think that this is promising for considering all the aspects connected to students' engagement, including the agentic one. In fact, as Reeve and Tseng (2011) pointed out, "recognizing that students constructively contribute into the instruction they receive clarifies the picture of how students learn and profit from potential learning opportunities" (p. 263).

Second, the correlations between the four engagement subscales and CCCI, GP-CORE, and school achievement are in the hypothesized directions as a support to the concurrent validity of the scale. All four engagement dimensions are positively correlated with school grades, confirming the fact that commitment and involvement play a critical role in achievement and learning (Kahu, 2013). Even student engagement and classmates' relationships are positively associated. Peers are an important part of school, and it is reasonable to assume that when students are socially connected and reciprocally supporting, they feel positively motivated toward academic work and school activities (Juvonen, Espinoza, & Knifsend, 2012). Finally, the degree of engagement is negatively associated with psychological distress, and this is consistent with other research suggesting that, besides multiple factors including family and peer relationships, personal school commitment affects, and is affected by, student's psychological distress (DeSantis-King, Huebner, Suldo, & Valois, 2006; Ma & Huebner, 2008). In addition, the negative association between agentic engagement and GP-CORE is particularly significant because it highlights the importance for students to be recognized as legitimated and competent actors within the school context for their personal well-being (Marginson, 2014; Ryan, Deci, & Vansteenkiste, 2016).

LIMITS AND CONCLUSIONS

This study has some limitations that need to be taken into account and that leave some questions unanswered. First, the results are based on a single sample. Moreover, CFAs fit the data well only after allowing correlations between many error terms. Future studies should

replicate these results in other schools and in other contexts to enhance the validity of the findings of the present study. Second, concurrent validity should be investigated using other variables. For instance, it could be interesting to see the relationships between the four dimensions of school engagement and family or teacher-student relationships. Third, and related to the last point, future studies should also consider variables that differentiate the four dimensions, in order to confirm their distinctiveness.

However, despite these limitations, the results presented in this article are promising. Indeed, our work extends current research by offering a comprehensive engagement scale which includes affective, behavioral, cognitive, and agentic components. Moreover, our study has tested the agentic dimension in a Western student population for the first time. The final scale displays good psychometric properties and can therefore be considered as a valid choice among the many tools existing on this subject. Moreover, the Italian validation of this instrument, to our knowledge, provides for the first time in this country the chance to use a questionnaire directly linked to the most recent international literature. Information about student engagement may be useful to define and evaluate prevention and intervention programs aimed at having an impact on students' liking for school and improving their school pathways.

NOTES

1. For a full discussion, please refer to the original article (Lam et al., 2014).
2. The model without the inclusion of correlated error terms did not fit the data well, $\chi^2(659) = 5597.83$; CFI = .78; TLI = .76; RMSEA = .079. However, as some scholars (see, Beckstead, 2002) have pointed out, the inclusion of correlated error terms in the CFA models does not undermine the factorial validity, whereas they are theoretically plausible. Rather, it provides a factorial representation of the observed data structure more appropriate and realistic in terms of real data.
3. The precise formulation of the Italian items composing the questionnaire is available from the authors upon request.

REFERENCES

Ainley, M. (2012). Students' interest and engagement in classroom activities. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 283-302). New York, NY: Springer.

Antaramian, S. P., Huebner, E. S., Hills, K. J., & Valois, R. F. (2010). A dual-factor model of mental health: Toward a more comprehensive understanding of youth functioning. *American Journal of Orthopsychiatry*, *80*, 462-472. doi:10.1111/j.1939-0025.2010.01049.x

Appleton, J. J., Christenson, S. L., & Furlong, M. J. (2008). Student engagement with school: Critical conceptual and methodological issues of the construct. *Psychology in the Schools*, *45*, 369-386. doi:10.1002/pits.20303

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the student engagement instrument. *Journal of School Psychology*, *44*, 427-445. doi:10.1016/j.jsp.2006.04.002

Beckstead, J. W. (2002). Confirmatory factor analysis of the Maslach Burnout Inventory among Florida nurses. *International Journal of Nursing Studies*, *39*(8), 785-792.

Borman, G. D., Hewes, G. M., Overmann, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, *73*, 125-230. doi:10.3102/00346543073002125

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Dissertation Abstracts International*, *3*, 185-216. doi:10.1177/135910457000100301

Brooks, R., Brooks, S., & Goldstein, S. (2012). The power of mindsets: Nurturing engagement, motivation, and resilience in students. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 541-562). New York, NY: Springer.

Cleary, T. J., & Zimmerman, B. J. (2012). A cyclical self-regulatory account of student engagement: Theoretical foundations and applications. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 237-257). New York, NY: Springer.

Crick, R. D. (2012). Deep engagement as a complex system: Identity, learning power and authentic enquiry. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 675-694). New York, NY: Springer.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281-302. doi:10.1037/h0040957

Crosnoe, R. (2002). High school curriculum track and adolescent association with delinquent friends. *Journal of Adolescent Research*, *17*, 144-168. doi:10.1177/0743558402172003

DeSantis-King, A. L. D., Huebner, S., Suldo, S. M., & Valois, R. F. (2006). An ecological view of school satisfaction in adolescence: Linkages between social support and behavior problems. *Applied Research in Quality of Life*, *1*, 279-295. doi:10.1007/s11482-007-9021-7

Dwyer, K. K., Bingham, S. G., Carlson, R. E., Prisbell, M., Cruz, A. M., & Fus, D. A. (2004). Communication and connectedness in the classroom: Development of the connected classroom climate inventory. *Communication Research Reports*, *21*, 264-272. doi:10.1080/08824090409359988

Evans, C., Connell, J., Audin, K., Sinclair, A., & Barkham, M. (2005). Rationale and development of a general population well-being measure: Psychometric status of the GP-CORE in a student sample. *British Journal of Guidance & Counselling*, *33*, 153-173. doi:10.1080/03069880500132581

Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE-OM. *The British Journal of Psychiatry*, *180*, 51-60. doi:10.1192/bjp.180.1.51

Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research*, *59*, 117-142. doi:10.3102/00346543059002117

Finn, J. D., Pannozzo, G. M., & Voelkl, K. E. (1995). Disruptive and inattentive-withdrawn behavior and achievement among fourth graders. *The Elementary School Journal*, *95*, 421-434. doi:10.1086/461853

Finn, J. D., & Rock, D. A. (1997). Academic success among students at risk for school failure. *Journal of Applied Psychology*, *82*, 221-234. doi:10.1037/0021-9010.82.2.221

Finn, J. D., & Zimmer, K. S. (2012). Student engagement: What is it? Why does it matter? In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 97-131). New York, NY: Springer.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*, 59-109. doi:10.3102/00346543074001059

Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 763-782). New York, NY: Springer.

Glanville, J. L., & Wildhagen, T. (2007). The measurement of school engagement: Assessing dimensionality and measurement invariance across race and ethnicity. *Educational and Psychological Measurement*, *67*, 1019-1041. doi:10.1177/0013164406299126

Green, L. W., & Lewis, F. M. (1986). *Measurement and evaluation in health education and health promotion*. Palo Alto, CA: Mayfield Publishing Company.

Herdman, M., Fox-Rushby, J., & Badia, X. (1997). "Equivalence" and the translation and adaptation of health-related quality of life questionnaires. *Quality of Life Research*, *6*. doi:10.1023/A:1026410721664

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*, 1-55. doi:10.1080/10705519909540118

Jimerson, S. R., Campos, E., & Grief, J. L. (2003). Toward an understanding of definitions and measures of school engagement and related terms. *California School Psychologist*, *8*, 7-27. doi:10.1007/bf03340893

Juvonen, J., Espinoza, G., & Knifsend, C. (2012). The role of peer relationships in student academic and extracurricular engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 387-401). New York, NY: Springer.

Kahu, E. R. (2013). Framing student engagement in higher education. *Studies in Higher Education*, *38*, 758-773. doi:10.1080/03075079.2011.598505

Kline, R. B. (2011). *Principles and practice of structural equation modelling* (3rd ed.). New York, NY: Guilford Press.

Lam, S., Jimerson, S., Wong, B. P. H., Kikas, E., Shin, H., Veiga, F. H., . . . Zollneritsch, J. (2014). Understanding and measuring student engagement in school: The results of an international study from 12 countries. *School Psychology Quarterly: The Official Journal of the Division of School Psychology, American Psychological Association*, *29*, 213-232. doi:10.1037/spq0000057

Lawson, M. A., & Lawson, H. A. (2013). New conceptual frameworks for student engagement research, policy, and practice. *Review of Educational Research*, *83*, 432-479.doi:10.3102/0034654313480891

Li, Y., & Lerner, R. M. (2011). Trajectories of school engagement during adolescence: Implications for grades, depression, delinquency, and substance use. *Developmental Psychology*, *47*, 233-247. doi:10.1037/a0021307

Ma, C. Q., & Huebner, E. S. (2008). Attachment relationships and adolescents' life satisfaction: Some relationships matter more to girls than boys. *Psychology in the Schools*, *45*, 177-190. doi:10.1002/pits.20288

Mameli, C., & Molinari, L. (2014). Seeking educational quality in the unfolding of classroom discourse: A focus on microtransitions. *Language and Education*, *28*, 103-119. doi:10.1080/09500782.2013.771654

Marginson, S. (2014). Student self-formation in international education. *Journal of Studies in International Education*, *18*, 6-22. doi:10.1177/1028315313513036

Marks, H. M. (2000). Student engagement in instructional activity: Patterns in elementary, middle and high school years. *American Educational Research Journal*, *37*,153-184. doi:10.3102/00028312037001153

Miller, R. B., Greene, B. A., Montalvo, G. P., Ravindran, B., & Nichols, J. D. (1996). Engagement in academic work: The role of learning goals, future consequences, pleasing others, and perceived ability. *Contemporary Educational Psychology*, *21*, 388-422. doi:10.1006/ceps.1996.0028

Ministry of Education, University, and Research (2013). Dati sul Sistema Istruzione nel Mezzogiorno [Data on the Education System in the South]. Retrieved from http://archivio.pubblica.istruzione.it/ fondistrutturali/valutazione/bussola_riferimento/dsi2.htm

Modell, J., & Elder, G. H. (2002). Children develop in history: So what's new? In W. Hartup & R. Weinberg (Eds.), *Child psychology in retrospect and prospect: In celebration of the 75th anniversary of the Institute of Child Development. The Minnesota Symposia on Child Psychology* (Vol. 32, pp. 173-205). Mahwah, NJ: Lawrence Erlbaum.

National Research Council & Institute of Medicine. (2004). *Engaging schools: Fostering high school students' motivation to learn*. Washington, DC: National Academy Press.

OECD-PISA. (2012). What students know and can do: Student performance in mathematics, reading and science. Retrieved from http://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf

Peck, S. C., Roeser, R. W., Zarrett, N., & Eccles, J. S. (2008). Exploring the roles of extracurricular activity quantity and quality in the educational resilience of vulnerable adolescents: Variable- and pattern-centered approaches. *Journal of Social Issues*, *42*, 135-156. doi:10.1111/j.1540-4560.2008.00552.x

Pianta, R. C., Hamre, B. K., & Allen, J. P. (2012). Teacher-student relationships and engagement: Conceptualizing, measuring, and improving the capacity of classroom interactions. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 365-386). New York, NY: Springer.

Pintrich, P. R., Smith, D., Garcia, T., & McKeachie, W. (1993). Predictive validity and reliability of the motivated strategies for learning questionnaire. *Educational and Psychological Measurement*, *53*, 801-813. doi:10.1177/0013164493053003024

Polychroni, F., Hatzichristou, C., & Sideridis, G. (2012). The role of goal orientations and goal structures in explaining classroom social and affective characteristics. *Learning and Individual Differences*, *22*, 207-217. doi:10.1016/j.lindif.2011.10.005

Rao, N., & Sachs, J. (1999). Confirmatory factor analysis of the Chinese version of the motivated strategies for learning questionnaire. *Educational and Psychological Measurement*, *59*, 1016-1029. doi:10.1177/00131649921970206

Ream, R. K., & Rumberger, R. W. (2008). Student engagement, peer social capital, and school dropout among Mexican American and non-Latino white students. *Sociology of Education*, *81*, 109-139. doi:10.1177/003804070808100201

Reeve, J. (2012). A self-determination theory perspective on student engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 149-172). New York, NY: Springer.

Reeve, J. (2013). How students create motivationally supportive learning environments for themselves: The concept of agentic engagement. *Journal of Educational Psychology*, *105*, 579-595. doi:10.1037/a0032690

Reeve, J., & Tseng, C. M. (2011). Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology*, *36*, 257-267. doi:10.1016/j.cedpsych.2011.05.002

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*, 68-78. doi:10.1037//0003-066x.55.1.68

Ryan, R. M., & Deci, E. L. (2002). An overview of self-determination theory: An organismic-dialectical perspective. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 3-33). Rochester, NY: University of Rochester Press.

Ryan, R. M., Deci, E. L., & Vansteenkiste, M. (2016). Autonomy and autonomy disturbances in self-development and psychopathology: Research on motivation, attachment, and clinical process. *Developmental Psychopathology*, One:9, 1-54. doi:10.1002/9781119125556.devpsy109

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*, 507-514. doi:10.1007/bf02296192

Shernoff, D. J., & Schmidt, J. A. (2008). Further evidence of an engagement-achievement paradox among US high school students. *Journal of Youth and Adolescence*, *37*, 564-580. doi:10.1007/s10964-007-9241-z

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, *85*, 571-581. doi:10.1037//0022-0663.85.4.571

Skinner, E. A., Kindermann, T. A., & Furrer, C. (2009). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement*, *69*, 493-525. doi:10.1177/0013164408323233

Skinner, E. A., & Pitzer, J. R. (2012). Developmental dynamics of student engagement, coping, and everyday resilience. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 283-302). New York, NY: Springer.

Steele, J. P., & Fullagar, C. J. (2009). Facilitators and outcomes of student engagement in a college setting. *The Journal of Psychology*, *143*, 5-27. doi:10.3200/jrlp.143.1.5-27

Vansteenkiste, M., Niemiec, C. P., & Soenens, B. (2010). The development of the five mini-theories of self-determination theory: An historical overview, emerging trends, and future directions. *Advances in Motivation and Achievement*, *16*, 105-165. doi:10.1108/s0749-7423(2010)000016a007

Voelkl, K. E. (2012). School identification. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 193-218). New York, NY: Springer.

Walker, C. O., Greene, B. A., & Mansell, R. A. (2006). Identification with academics, intrinsic/extrinsic motivation, and self-efficacy as predictors of cognitive engagement. *Learning and Individual Differences*, *16*, 1-12. doi:10.1016/j.lindif.2005.06.004

Wang, M. T., & Eccles, J. S. (2013). School context, achievement motivation, and academic engagement: A longitudinal study of school engagement using a multidimensional perspective. *Learning and Instruction*, *28*, 12-23. doi:10.1016/j.learninstruc.2013.04.002

Wang, M. T., & Fredricks, J. A. (2014). The reciprocal links between school engagement, youth problem behaviors, and school dropout during adolescence. *Child Development*, *85*, 722-737. doi:10.1111/cdev.12138

Wang, M. T., & Holcombe, R. (2010). Adolescents' perceptions of school environment, engagement, and academic achievement in middle school. *American Educational Research Journal*, *47*, 633-662. doi:10.3102/0002831209361209

Wang, M. T., Willet, J. B., & Eccles, J. S. (2011). The assessment of school engagement: Examining dimensionality and measurement invariance by gender and race/ethnicity. *Journal of School Psychology*, *49*, 465-480. doi:10.1016/j.jsp.2011.04.001

TPM®

WHO'S MOST LIKELY TO GET STRESSED AND LEAVE THE COMPANY? EFFECTS OF REGULATORY MODE ON WORK STRESS AND TURNOVER INTENTIONS

CALOGERO LO DESTRO
SAPIENZA UNIVERSITY OF ROMA

MARINA CHERNIKOVA
UNIVERSITY OF MARYLAND, COLLEGE PARK

ANTONIO AIELLO
UNIVERSITY OF PISA

ANTONIO PIERRO
SAPIENZA UNIVERSITY OF ROMA

Regulatory mode theory (Kruglanski et al., 2000) posits the existence of two independent regulatory mode orientations: locomotion and assessment. Locomotion reflects the tendency to move from one state to another, whereas assessment reflects the tendency to evaluate means and goals in an attempt to choose the best possible option. Past research has sought to identify the relationship between regulatory mode and well-being, however it has not been fully disclosed. To our knowledge, the present work is the first attempt to investigate the relationship between locomotion, assessment, and two important organizational outcomes: work stress and turnover intentions. We recruited employees from 24 Italian organizations, and obtained their individual scores on: the Regulatory Mode Scale (Kruglanski et al., 2000); ratings of work stress, and turnover intentions. Two separate moderated multiple regression analyses were run to test the main effects and the interactions of regulatory mode on the two outcome measures. In line with our predictions, the results revealed that assessment was positively associated with both work stress and turnover intentions, while locomotion was negatively associated with those variables. Using a mediated moderation analysis, we also found that the combination of high assessment and low locomotion was the best predictor of turnover intentions, and this relationship was mediated by work stress. We discuss implications, limitations, and future directions for these findings.

Key words: Regulatory mode; Locomotion; Assessment; Work stress; Turnover intentions.

Correspondence concerning this article should be addressed to Calogero Lo Destro, Department of Social and Developmental Psychology, Sapienza University of Roma, Via dei Marsi 78, 00185 Roma (RM), Italy. Email: calogero.lodestro@uniroma1.it

The main goal of the present work is to investigate the interactive effect of regulatory mode (Kruglanski et al., 2000) on factors related to well-being in organizations, such as stress and turnover intentions. Work-related stress — defined as "the subjective feeling that work demands exceed the individual's belief in his or her capacity to cope" (Cropanzano, Howes,

Grandey, & Toth, 1997, p. 164) — can have a variety of negative effects on both employees and organizations. For instance, such stress is associated with increased employee health problems (Ganster & Schaubroeck, 1991), reduced productivity (Joure, Leon, Simpson, Holley, & Frye, 1989), more workplace accidents (Kirkcaldy, Trimpop, & Cooper, 1997), and more lost work time (Cartwright & Cooper, 1997). As work stress increases, an individual is more likely to form a turnover intention — in other words, decide to leave the organization he or she is working for (Fisher & Gitelson, 1983; Griffeth, Hom, & Gaertner, 2000). Much like work stress, turnover intentions also have many negative consequences, such as decreased team performance (Ton & Huckman, 2008), lower company profits (Cascio, 1982, 1986; Heavey, Holwerda, & Hausknecht, 2013), loss of job knowledge (Trevor, Gerhart, & Boudreau, 1997), and lower likelihood of company survival (Phillips, 2002). Clearly, work stress and turnover intentions can have a strong impact on both employees' and organizations' well-being; as such, it is important to identify variables that can predict these factors.

Past research has shown that individual differences can play a major role in individuals' likelihood of experiencing organizational outcomes such as stress and turnover intentions (Fogarty et al., 1999; Girardi, Falco, Dal Corso, Kravina, & De Carlo, 2011; Swider & Zimmerman, 2010). Two individual difference factors that are particularly relevant to these outcomes are locomotion and assessment regulatory mode (Kruglanski et al., 2000). These are described in more detail below.

According to regulatory mode theory, locomotion regulatory mode is the aspect of self-regulation concerned with motion and progress from state to state. Assessment regulatory mode, on the other hand, is the facet of self-regulation that allows one to make comparisons between different entities (e.g., means or goals; Kruglanski et al., 2000). In line with these notions, locomotion is associated with swift action and the desire to maintain uninterrupted motion, while assessment is related to the preference for methodical and thoughtful analysis of every available option. The two regulatory modes can be measured as chronic individual differences (Kruglanski et al., 2000) or induced as state variables (Avnet & Higgins, 2003). They are orthogonal, so an individual can be low or high on both dimensions, or low on one and high on the other.

Assessment is relevant to organizational outcomes such as stress because it is generally associated with more negative affect and depression. This association between assessment and negative affective outcomes is likely due to assessors' tendency to continuously engage in critical evaluations of their own and others' behavior, which is not conducive to psychological well-being (Hong, Tan, & Chan, 2004; Kruglanski et al., 2000). Another essential aspect of assessment, the propensity to focus on the discrepancies between one's current self and one's ideal self, has also been shown to lead to significant increases in negative affect (Higgins, 1987). In contrast, locomotion is relevant to organizational outcomes such as stress because it is generally associated with positive affect, optimism, and increased self-esteem. This link between locomotion and positive affective outcomes can be explained by locomotors' focus on forward motion and progress, which causes them to avoid dwelling on negative aspects of their current or past states (Kruglanski, Pierro, & Higgins, 2016; Kruglanski et al., 2000).

Based on the logic above, assessment should be linked to lower organizational well-being because of assessors' tendency to dwell on discrepancies, which exacerbates negative affect. Locomotion, meanwhile, should be linked to higher organizational well-being because high locomotors devote little time to dwelling on their own or others' shortcomings (Kruglanski et al., 2016, 2000). Prior research has found some support for these notions. Pierro, Giacomantonio, Pica,

Kruglanski, and Higgins (2013, Study 2) observed that locomotion predicted job satisfaction and was related to lower work stress and turnover intentions; that study, however, did not test the relationship between assessment and the aforementioned variables. Bélanger et al. (2016) revealed that locomotion predicted withdrawal behaviors at work. More specifically, locomotion was negatively related to absenteeism, lateness, and early departures from work. De Carlo et al. (2014) found that assessment was positively associated with psychological strain and burnout, while locomotion was negatively associated with both. Moreover, workaholism was found to mediate the relationship between both assessment and locomotion and the dependent variables. At the same time, high locomotors experienced more work engagement and consequently less psychological strain and burnout, while high assessors showed less work engagement, which resulted in more burnout and psychological strain.

Considering passion, Bélanger et al. (2014) evidenced that the negative association between locomotion and work stress and burnout, was mediated by harmonious[1] passion. In contrast, assessment had a positive direct and indirect (via obsessive passion) effect on work stress. There were no significant direct effects on turnover intentions, but the indirect effect (although not significant) followed the same trend as work stress. Thus, multiple studies have corroborated the idea that locomotion leads to more positive organizational outcomes, and assessment leads to more negative organizational outcomes.

Nonetheless, the interaction between assessment and locomotion in predicting stress and turnover intentions has not been fully explored. One study did investigate the possible interactive effects of the regulatory mode on subjective well-being (Hong et al., 2004). In a student sample, Hong et al. noticed that individuals low on locomotion but high on assessment experienced more depressive moods, while individuals high on locomotion but low in assessment exhibited higher life satisfaction. Interestingly, they found that locomotion-assessment complementarity (i.e., the combination of high locomotion and high assessment) did not increase life satisfaction, though it increases performance in a variety of domains (Hamstra, Orehek, & Holleman, 2014; Pierro, Kruglanski, & Higgins, 2006; Pierro, Pica, Mauro, Kruglanski, & Higgins, 2012). Importantly, however, the above authors did not investigate the interactive effects of locomotion and assessment on stress and turnover intentions in the organization. The goal of the present research was to fill this gap.

HYPOTHESES

Based upon the previous reasoning, we hypothesized that:

H1. Assessment should be positively associated with both work stress and turnover intentions, while locomotion should be negatively associated with them.

More importantly, we were interested in testing the interactive effects of assessment and locomotion on these two organizational outcomes. Specifically, we hypothesized that:

H2. Employees high on assessment and low on locomotion should experience the highest work stress and the greatest turnover intentions. We also expected that high locomotion would serve as a buffer against the effects of assessment: the presence of locomotion should attenuate the negative effects of assessment.

Lastly, we were interested in testing a mediated moderation model. In other words, we expected that:

H3. When locomotion is low, assessment is associated with greater turnover intentions through the mediation of work stress.

METHOD

Participants

One thousand eight hundred and eighty-six employees (894 females) from 24 public and private Italian organizations participated in this research on a voluntary basis. Employees' mean age was 41.65 (*SD* = 10.92) and their mean job tenure was 14.17 years (*SD* = 10.63). Among participants 25.3% had a university degree, 50.1% a high-school degree, 21.7 % a middle-school diploma, and 2.9% an elementary-school diploma. The study complied with the Declaration of Helsinki and was approved by the local ethics committee.

Procedure and Materials

Participants were given information about the experimental procedures and provided written consent. At the workplace, participants filled out the Regulatory Mode Scale (Kruglanski et al., 2000), which was followed by a measure of stress at work (Cohen, Kamarck, & Mermelstein, 1983), and a measure of turnover intentions (Mobley, 1977). The paper-and-pencil questionnaire administered to participants included an introductory letter in which the purpose of the study was explained and anonymity was guaranteed.

Regulatory mode orientations. Participants completed the Italian version of the Regulatory Mode Scale (Kruglanski et al., 2000), which is composed of two separate 12-item self-report measures designed to tap individual differences in locomotion and assessment. Specifically, respondents rated the extent to which they agree with self-descriptive statements reflecting *locomotion* (e.g., "By the time I accomplish a task, I already have the next one in mind") or *assessment* (e.g., "I spend a great deal of time taking inventory of my positive and negative characteristics"). Ratings are made on a 6-point Likert scale ranging from 1 (*strongly disagree*) to 6 (*strongly agree*). Previous studies with Italian samples (Kruglanski et al., 2000) demonstrated that the locomotion and assessment scales have satisfactory reliability (α = .82 for the locomotion scale and .78 for the assessment scale). In the present sample, the alpha for the locomotion and assessment scales was .80 and .69, respectively.

Stress. Six items from the Perceived Stress Scale (Cohen et al., 1983) were adapted to measure participants' stress at work (e.g., "In the last month, I often felt nervous and stressed at work"; "In the last month, I often felt unable to control important things at work"; see also Bélanger et al., 2014). The six items were translated into Italian and then translated back into English. Participants responded on a 6-point Likert scale ranging from 1 (*strongly disagree*) to 6 (*strongly agree*). A score of perceived stress was computed by averaging across responses to each item (α = .89).

Turnover intentions. These were assessed via three items adapted from a turnover intention measure developed by Mobley (1977; e.g., "I have often seriously considered finding a job

elsewhere"). The items of the scale were translated into Italian and then translated back into English. Participants' responses were recorded on a 6-point Likert scale ranging from 1 (*strongly disagree*) to 6 (*strongly agree*). A composite turnover intentions score was computed by averaging across responses to each item (α = .91).

Data Analysis

Descriptive statistics and correlations between variables are presented in Table 1. Predictions regarding the main effects and the interaction of locomotion and assessment on the two outcome measures (i.e., stress and turnover intentions) were tested by means of two separate moderated multiple regression analyses. In these moderated multiple regression analyses, we entered the main effects of locomotion and assessment orientations and the interaction between them. Following the recommendation by Aiken and West (1991), predictor variables were centered, and the interaction term was based on these centered scores. Gender (dummy coded as male = 0 and female = 1), age, job tenure, and education were entered as control variables. A summary of the results of these analyses is reported in Table 2.

In addition to our hypothesis concerning the effects of regulatory mode on stress and turnover intentions, we also tested the mediating role of stress in the relationship between regulatory modes and turnover, using Model 8 (a mediated moderation model) proposed by Preacher and Hayes (2008). A summary of estimated direct and indirect effects of mode on turnover intentions through stress is reported in Table 3 and Figure 2.

Results

As can be seen (Table 1) in this sample, the correlation between assessment and locomotion scales was nonsignificant. This is consistent with prior research, which has generally found low or no correlation between the two regulatory modes (e.g., Kruglanski et al., 2000). Moreover, as expected, stress and turnover intentions were strongly and positively correlated, and both exhibited a significant and positive correlation with assessment and a significant and negative correlation with locomotion.

TABLE 1
Descriptive statistics and correlations between variables (N = 1886)

	M (*SD*)	1	2	3	4
1. Locomotion	4.60 (0.67)	(.80)			
2. Assessment	3.25 (0.70)	–.03	(.69)		
3. Stress	3.05 (1.22)	–.24***	.24***	(.89)	
4. Turnover intentions	2.61 (1.52)	–.28***	.21***	.38***	(.91)

Note. Standard deviations and Cronbach's alphas are reported in parentheses.
*** p < .001.

As can be seen in Table 2, the results of the moderated multiple regression analyses showed: a) a significant and positive main effect of assessment on stress (b = .47, SE = .04, p < .001) and turnover intentions (b = .42, SE = .05, p < .001), indicating that stress and turnover intentions were higher for employees high in assessment; b) a significant and negative main effect of locomotion on stress (b = –.44, SE = .04, p < .001) and turnover intentions (b = –.60, SE = .05, p < .001), suggesting that stress and turnover intentions were lower for employees high in locomotion. Of greater interest is that the hypothesized interaction between locomotion and assessment was significant and negative for both criterion variables (stress: b = –.23, SE = .06, p = .001; turnover intentions: b = –.25, SE = .07, p = .004).[2]

TABLE 2
Summary of results of moderated multiple regression analyses, unstandardized coefficients

Predictors	Criteria					
	Stress			Turnover intentions		
	b	*SE*	*p* =	*b*	*SE*	*p* =
Control variables						
Gender	.23	.05	.000	–.14	.07	.034
Age	.00	.00	.796	–.00	.01	.346
Education	.00	.03	.972	.30	.04	.000
Job tenure	–.00	.00	.817	–.01	.01	.046
Main predictors						
Locomotion	–.44	.04	.000	–.60	.05	.000
Assessment	.47	.04	.000	.42	.05	.000
Locomotion × Assessment	–.23	.06	.001	–.25	.07	.004
R^2	.132		.000	.161		.000
ΔR^2	.007		.000	.006		.000

Note. *SE* = standard error; R^2 = the overall explained variance for the model including all predictors; ΔR^2 = the increase in explained variance due to the addition of the interaction terms.

To further illustrate the nature of these interaction effects, we performed simple slopes analyses for low (1 *SD* below the mean) and high (1 *SD* above the mean) levels of locomotion, following the recommendations of Aiken and West (1991). These analyses revealed that the relationship between assessment and both stress and turnover was stronger for participants relatively low in locomotion (stress: b = .62, SE = .06, p < .001; turnover intentions: b = .59, SE = .08, p < .001) than for participants relatively high in locomotion (stress: b = .32, SE = .05, p < .001; turnover: b = .25, SE = .06, p < .001). Overall, these results suggest that the relations between assessment and the criterion variables were weakened for high (vs. low) locomotors. These findings are illustrated in Figure 1, a and b.

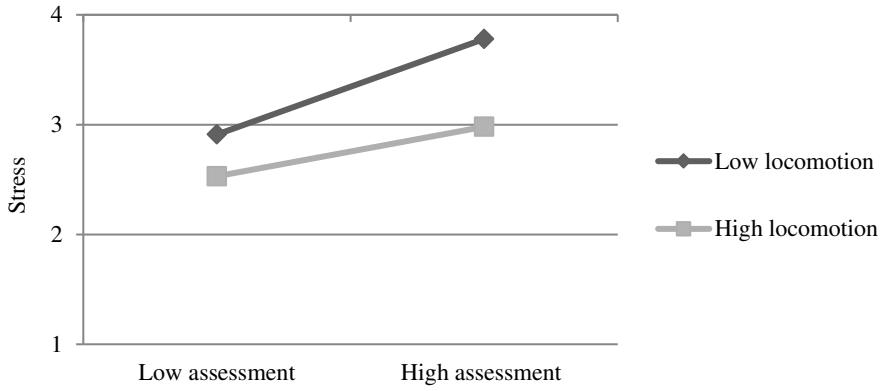

Figure 1a. Stress as a function of the interaction between locomotion and assessment.

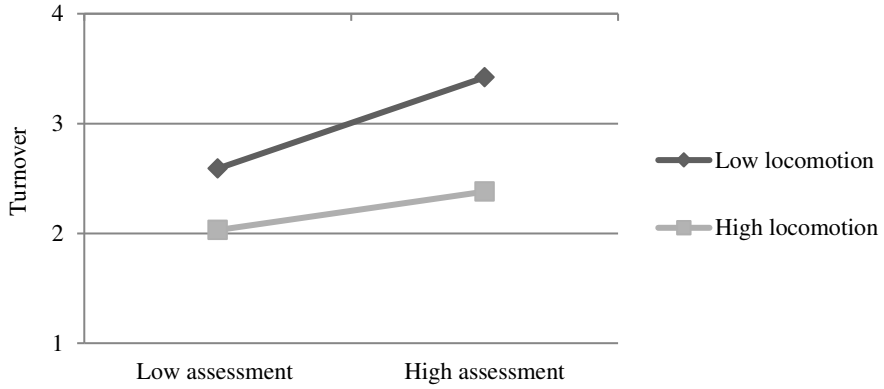

Figure 1b. Turnover intentions as a function of the interaction between locomotion and assessment.

FIGURE 1
Simple slopes for the interaction between locomotion and assessment.

Finally, as expected, for the mediated moderation analysis we found (see Table 3 and Figure 2) that the direct conditional effect of assessment on turnover was positive and significant at low (–1*SD*), mean (*M*), and high (+1*SD*) levels of locomotion; however, the strength of the effect increased as locomotion levels decreased (b = .36, SE = .08, p < .001 at –1*SD*; b = .24, SE = .05, p < .001 at *M*; b = .13, SE = .06, p < .05 at +1*SD*). More importantly, consistently with our mediated moderation hypothesis, stress mediated the effect of assessment on turnover intentions at low (–1*SD*), mean (*M*), and high (+1*SD*) levels of locomotion. As for conditional direct effects, the strength of conditional indirect effect increased with a decreasing in locomotion levels (b = .23, SE = .03 at –1*SD*; b = .18 SE = .02 at *M*; b = .12 SE = .02 at +1*SD*). Moreover, as predicted, the negative relation between the highest order interaction (i.e., the interaction between locomotion and assessment) and turnover intentions was mediated by stress (b = –.08, SE = .02). Bootstrapped CIs corroborated the reliability of the indirect effects.

TABLE 3
Conditional effects (unstandardized estimates) of assessment on turnover intentions
at specific values of the moderator (locomotion)

	Indirect effects via stress				Direct effects				
Locomotion	*b*	*SE*	BootLLCI	BootULCI	*b*	*SE*	*p* =	LLCI	ULCI
–1*SD*	.23	.03	.17	.30	.36	.08	.000	.21	.50
Mean	.18	.02	.14	.22	.24	.05	.000	.15	.34
+1*SD*	.12	.02	.07	.17	.13	.06	.018	.02	.24

Indirect effects of the highest order interaction			
b	*SE*	BootLLCI	BootULCI
.08	.02	–.13	–.04

Note. *SE* = standard error; BootLLCI = Bootstrap lower level confidence interval; BootULCI = Bootstrap upper level confidence interval ; LLCI = Lower level confidence interval; ULCI = Upper level confidence interval.

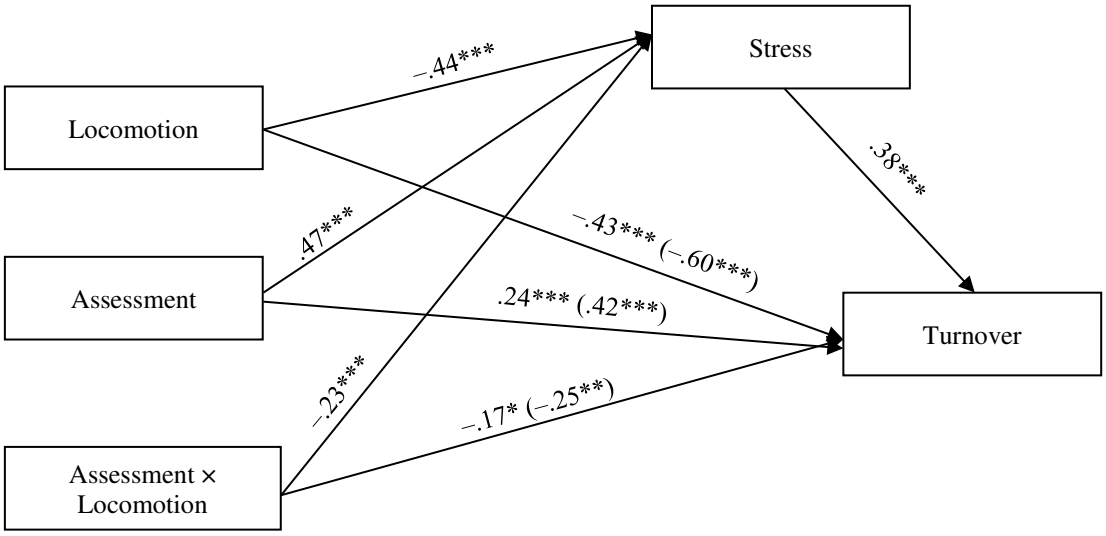

FIGURE 2
Mediated moderation model of the relationship between assessment and turnover intentions,
unstandardized coefficients.
The regression coefficients of the predictors when the mediator was not included
in the model are reported in brackets.
* $p < .05$. ** $p < .01$. ***$p < .001$.

DISCUSSION

In the present study, we investigated the role of locomotion and assessment regulatory modes in predicting the organizational phenomena of work-related stress and turnover intentions. We found that the two regulatory modes had opposite effects on these outcomes: assessment pos-

itively predicted both variables, while locomotion negatively predicted them. We also observed a mediated moderation pattern: the relationship between assessment and turnover intentions was mediated by stress, this mediation being moderated by locomotion. More specifically, the direct conditional effect of assessment on turnover intentions was positive for every locomotion level; however, the lower the locomotion level, the stronger the effect. We also found that stress mediated the effect of assessment on turnover intentions at any locomotion level, and the conditional indirect effect increased when locomotion levels decreased. It therefore appears that locomotion was serving as a sort of buffer for the negative effects of assessment: when locomotion was high, the negative effects of assessment (on both stress and turnover intentions) were weaker. Finally, we found that stress negatively mediated the highest order interaction between the combination of regulatory modes and turnover intentions. These results support the idea that the combination of high assessment and low locomotion is the best predictor of increased work stress, and that this combination is subsequently related to higher turnover intentions.

A wide variety of past research has shown that being high on both assessment and locomotion (i.e., locomotion-assessment complementarity) generally leads to the best performance-related outcomes (Hamstra et al., 2014; Pierro et al., 2006, 2012). Furthermore, it has been found that for simple (vs. complex), familiar (vs. unfamiliar), or for tasks requiring low interdependence (vs. high interdependence) individuals high on locomotion, but low on assessment, perform better (Chernikova et al., 2016; Chernikova, Lo Destro, Pierro, Higgins, & Kruglanski, 2017; Lo Destro, Chernikova, Pierro, Kruglanski, & Higgins, 2016, 2017). Nonetheless, in the present work, we found yet another pattern for affective states. Indeed, the highest stress levels, as well as the highest turnover intentions, were observed in employees high on assessment and low on locomotion. This was likely due to high assessors' focus on discrepancies, combined with low locomotors' lack of progress toward addressing those discrepancies. On the other hand, all other combinations of regulatory modes led to better outcomes. In fact, in the presence of high locomotion, stress and turnover intentions were weakened regardless of assessment level, suggesting that locomotion can have a protective "buffer" effect against some negative outcomes.

With the necessary caution, we can assume that employees who show this self-regulation pattern (i.e., the combination of low locomotion and high assessment) are at higher risk of encountering difficulties in coping with organizational stressors and consequently developing turnover intentions. In line with this notion, employers should carefully take into account potential employees' regulatory mode during the recruitment process, since it seems employees high on locomotion and low on assessment show a lower predisposition to experience work stress. Furthermore, employees who experience a higher well-being level are less likely to leave the company, and as a consequence the costs related to recruiting and training new workers are reduced and resources are optimized. In a prevention perspective, managers should adapt task requests to employees' capabilities and should clarify employees' roles and responsibilities. The organization, moreover, should take care of employees' well-being, especially of those employees who they know to be at the highest risk of experiencing stress during the course of their daily work (low locomotors and high assessors), furnishing adequate training about the causes of stress and the way to keep it away. Finally, employers may also wish to tailor well-being interventions (e.g., mindfulness training; Mackenzie, Poulin, & Seidman-Carlson, 2006).

Several limitations of the present research should be noted. One limitation is that we used a self-report measure of stress. In future research, it would be useful to have some physiological

assessments of stress (e.g., salivary cortisol levels; Pruessner, Hellhammer, Pruessner, & Lupien, 2003; Schulz, Kirschbaum, Prüßner, & Hellhammer, 1998). Another limitation is that we used a self-report measure of turnover intentions, which may not be highly correlated with actual turnover. Though a large amount of studies have shown that behavioral intentions are a strong predictor of behavior (e.g., Ajzen, 1985, 1991; Schifter & Ajzen, 1985), future research should nonetheless attempt to replicate our findings with an actual measure of turnover. A third limitation is that these data are derived from the same source, and thus could potentially be susceptible to common method/source bias. It must be noted, however, that, although common method/source biases can inflate the relationship between variables, they normally lead to the underestimation of interaction effects (Evans, 1985; McClelland & Judd, 1993). Yet another limitation is that we did not take organizational performance into account in the present research. One possibility is that employees who are more stressed perform worse (Lang, Thomas, Bliese, & Adler, 2007; Taris, 2006), but it is also plausible that a moderate amount of stress could actually enhance performance. Future studies would do well to examine the relationship between regulatory mode, organizational well-being, performance, and turnover intentions.

Other directions for future research involve testing whether team composition matters. For instance, does within-group regulatory mode complementarity have a positive effect on employee well-being? Or do employees feel better when they work in a team of individuals who are similar to them, thus experiencing a fit effect? These questions can be fruitfully explored in further studies on this topic.

NOTES

1. In the dualistic model of passion (Vallerand et al., 2003) passion is defined as a strong inclination toward an activity that individuals find important and in which they invest energy and time. The authors distinguished two types of passion: (1) harmonious, which derives from an autonomous internalization of the activity into one's identity; (2) obsessive, stemming from a controlled internalization of the activity into one's identity.
2. We acknowledge that the structure of the data is nested (i.e., individual stress and turnover intentions ratings are nested within organizations) and that this may raise the concern of non-independent data. Thus, we calculated the intraclass correlation coefficient (ICC). Results yielded a coefficient of .089 for stress and .173 for turnover, suggesting that only a small proportion of the variance in stress and turnover ratings was between organizations. Furthermore, we applied a multilevel modeling approach to the data, using restricted maximum likelihood (REML) estimation. In the two analyses (one for each criterion variable), we entered all our level-one control variables and the main predictor variables as fixed; only the intercepts (entered at the organization level) were a random effect. Consistent with our results, the analysis showed a significant and positive effect of assessment, a significant and negative effect of locomotion, and a significant and negative two-way interaction between assessment and locomotion on both stress and turnover intentions, confirming that our conclusions are not compromised by the potential dependency of observations. (Data of multilevel analysis are available upon request from the corresponding author.)

REFERENCES

Aiken, L. S., & West, S. G., (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage Publications.

Ajzen, I. (1985). From intentions to actions: A Theory of Planned Behavior. In J. Kuhl & J. Beckmann (Eds.), *Action control* (pp. 11-39). Berlin-Heidelberg, Germany: SSSP Springer Series in Social Psychology.

Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*, *50*, 179-211. doi:10.4135/9781412952576.n208

Avnet, T., & Higgins, E. T. (2003). Locomotion, assessment, and regulatory fit: Value transfer from "How" to "What." *Journal of Experimental Social Psychology*, *39*, 525-530. doi:10.1016/s0022-1031(03)00027-1

Bélanger, J. J., Pierro, A., Kruglanski, A. W., Vallerand, R. J., De Carlo, N. A., & Falco, A. (2014). On feeling good at work: The role of regulatory mode and passion in psychological adjustment. *Journal of Applied Social Psychology*, *45*, 319-329. doi:10.1111/jasp.12298.

Bélanger, J. J., Pierro, A., Mauro, R., Falco, A., De Carlo, N. A., & Kruglanski, A. W. (2016). It's about time: The role of locomotion in withdrawal behavior. *Journal of Business and Psychology*, *31*, 265-278. doi:10.1007/s10869-015-9409-6

Cartwright, S., & Cooper, C. L. (1997). *Managing workplace stress.* London, UK: Sage Publications.

Cascio, W. F. (1982). *Costing human resources: The financial impact of behavior in organizations*. Boston, MA: Kent.

Cascio, W. F. (1986). *Managing human resources: Productivity, quality of work life, profits*. New York, NY: McGraw-Hill.

Chernikova, M., Lo Destro, C., Mauro, R., Pierro, A., Kruglanski, A. W., & Higgins, E. T. (2016). Different strokes for different folks: Effects of regulatory mode complementarity and task complexity on performance. *Personality and Individual Differences*, *89*, 134-142. doi:10.1016/j.paid.2015.10.011

Chernikova, M., Lo Destro, C., Pierro, A., Higgins, E. T., & Kruglanski, A. W. (2017). A multilevel analysis of person-group regulatory-mode complementarity: The moderating role of group-task interdependence. *Group Dynamics: Theory, Research, and Practice*, *21*(2), 108-120.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, *24*, 385-396. doi:10.2307/2136404

Cropanzano, R., Howes, J. C., Grandey, A. A., & Toth, P. (1997). The relationship of organizational politics and support to work behaviors, attitudes, and stress. *Journal of Organizational Behavior*, *18*, 159-180. doi:10.1002/(SICI)1099-1379(199703)18:2<159::AID-JOB795>3.0.CO;2-D

De Carlo, N. A., Falco, A., Pierro, A., Dugas, M., Kruglanski, A. W., & Higgins, E. T. (2014). Regulatory mode orientations and well-being in an organizational setting: The differential mediating roles of workaholism and work engagement. *Journal of Applied Social Psychology*, *44*, 725-738. doi:10.1111/jasp.12263

Evans, M. G. (1985). A Monte Carlo study of the effects of correlated method variance in moderated regression analysis. *Organizational Behavior and Human Decision Processes*, *36*, 305-323. doi:10.1016/0749-5978(85)90002-0

Fisher, C. D., & Gitelson, R. (1983). A meta-analysis of the correlates of role conflict and ambiguity. *Journal of Applied Psychology*, *68*, 320-333. doi:10.1037/0021-9010.68.2.320

Fogarty, G. J., Machin, M. A., Albion, M. J., Sutherland, L. F., Lalor, G. I., & Revitt, S. (1999). Predicting occupational strain and job satisfaction: The role of stress, coping, personality, and affectivity variables. *Journal of Vocational Behavior*, *54*, 429-452. doi:10.1006/jvbe.1998.1670

Ganster, D. C., & Schaubroeck, J. (1991). Work stress and employee health. *Journal of Management*, *17*, 235-271. doi:10.1177/014920639101700202

Girardi, D., Falco, A., Dal Corso, L., Kravina, L., & De Carlo, A. (2011). Interpersonal conflict and perceived work stress: The role of negative affectivity. *TPM – Testing, Psychometrics, Methodology in Applied Psychology*, *18*, 257-273. doi:10.4473/TPM.18.4.5

Griffeth, R. W., Hom, P. W., & Gaertner, S. (2000). A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. *Journal of Management*, *26*, 463-488. doi:10.1177/014920630002600305

Hamstra, M. R., Orehek, E., & Holleman, M. (2014). Subordinate regulatory mode and leader power: Interpersonal regulatory complementarity predicts task performance. *European Journal of Social Psychology*, *44*, 1-6. doi:10.1002/ejsp.1992

Heavey, A. L., Holwerda, J. A., & Hausknecht, J. P. (2013). Causes and consequences of collective turnover: A meta-analytic review. *Journal of Applied Psychology*, *98*, 412-453. doi:10.1037/a0032380

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review*, *94*, 319-340. doi:10.1037/0033-295x.94.3.319

Hong, R. Y., Tan, M. S., & Chang, W. C. (2004). Locomotion and assessment: Self-regulation and subjective well-being. *Personality and Individual Differences*, *37*, 325-332. doi:10.1016/j.paid.2003.09.006

Joure, S. A., Leon, J. S., Simpson, D. B., Holley, C. H., & Frye, R. L. (1989). Stress: The pressure cooker of work. *The Personnel Administrator*, *34*(11), 92-95.

Kirkcaldy, B. D., Trimpop, R., & Cooper, C. L. (1997). Working hours, job stress, work satisfaction, and accident rates among medical practitioners and allied personnel. *International Journal of Stress Management*, *4*, 79-87. doi:10.1007/BF02765302

Kruglanski, A. W., Pierro, A., & Higgins, E. T. (2016). Experience of time by people on the go: A theory of the locomotion-temporality interface. *Personality and Social Psychology Review*, *20*, 100-117. doi:10.1177/1088868315581120

Kruglanski, A. W., Thompson, E. P., Higgins, E. T., Atash, M., Pierro, A., Shah, J. Y., & Spiegel, S. (2000). To "Do the right thing" or to "Just do it": Locomotion and assessment as distinct self regulatory imperatives. *Journal of Personality and Social Psychology*, *79*, 793-815. doi:10.1037/0022-3514.79.5.793

Lang, J., Thomas, J. L., Bliese, P. D., & Adler, A. B. (2007). Job demands and job performance: The mediating effect of psychological and physical strain and the moderating effect of role clarity. *Journal of Occupational Health Psychology*, *12*, 116-124. doi:10.1037/1076-8998.12.2.116

Lo Destro, C., Chernikova, M., Pierro, A., Kruglanski, A. W., & Higgins, E. T. (2016). Practice benefits locomotors: Regulatory mode complementarity and task performance. *Social Psychological and Personality Science*, *7*, 358-365. doi:10.1177/1948550615616171

Lo Destro, C., Chernikova, M., Pierro, A., Kruglanski, A. W., & Higgins, E. T. (2017). *Effect of regulatory mode on work performance: The moderating roles of job complexity and familiarity*. Unpublished manuscript, Department of Social and Developmental Psychology, Sapienza University of Rome.

Mackenzie, C. S., Poulin, P. A., & Seidman-Carlson, R. (2006). A brief mindfulness-based stress reduction intervention for nurses and nurse aides. *Applied Nursing Research*, *19*, 105-109. doi:10.1016/j.apnr.2005.08.002

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, *114*, 376-390. doi:10.1037/0033-2909.114.2.376

Mobley, W. H. (1977). Intermediate linkages in the relationship between job satisfaction and employee turnover. *Journal of Applied Psychology*, *62*, 237-240. doi:10.1037/0021-9010.62.2.237

Phillips, D. J. (2002). A genealogical approach to organizational life chances: The parent-progeny transfer among Silicon Valley Law firms, 1946-1996. *Administrative Science Quarterly*, *47*, 474-506. doi:10.2307/3094848

Pierro, A., Giacomantonio, M., Pica, G., Kruglanski, A. W., & Higgins, E. T. (2013). Locomotion and the preference for multitasking: Implications for well-being. *Motivation and Emotion*, *37*, 213-223. doi:10.1007/s11031-012-9300-y

Pierro, A., Kruglanski, A. W., & Higgins, E. T. (2006). Regulatory mode and the joys of doing: Effects of "Locomotion" and "Assessment" on intrinsic and extrinsic task motivation. *European Journal of Personality*, *20*, 355-375. doi:10.1002/per.600

Pierro, A., Pica, G., Mauro, R., Kruglanski, A. W., & Higgins, E. T. (2012). How regulatory modes work together: Locomotion-assessment complementarity in work performance. *TPM – Testing, Psychometrics, Methodology in Applied Psychology*, *19*, 247-262. doi:10.4473/TPM19.4.1

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879-891.

Pruessner, M., Hellhammer, D. H., Pruessner, J. C., & Lupien, S. J. (2003). Self-reported depressive symptoms and stress levels in healthy young men: Associations with the cortisol response to awakening. *Psychosomatic Medicine*, *65*, 92-99. doi:10.1097/01.psy.0000040950.22044.10

Schifter, D. E., & Ajzen, I. (1985). Intention, perceived control, and weight loss: An application of the Theory of Planned Behavior. *Journal of Personality and Social Psychology*, *49*, 843-851. doi:10.1037/0022-3514.49.3.843

Schulz, P., Kirschbaum, C., Prüßner, J., & Hellhammer, D. (1998). Increased free cortisol secretion after awakening in chronically stressed individuals due to work overload. *Stress and Health*, *14*, 91-97. doi:10.1002/(SICI)1099-1700(199804)14:2<91::AID-SMI765>3.0.CO;2-S

Swider, B. W., & Zimmerman, R. D. (2010). Born to burnout: A meta-analytic path model of personality, job burnout, and work outcomes. *Journal of Vocational Behavior*, *76*, 487-506. doi:10.1016/j.jvb.2010.01.003

Taris, T. W. (2006). Is there a relationship between burnout and objective performance? A critical review of 16 studies. *Work & Stress*, *20*, 316-334. doi:10.1080/02678370601065893

Ton, Z., & Huckman, R. S. (2008). Managing the impact of employee turnover on performance: The role of process conformance. *Organization Science*, *19*, 56-68. doi:10.1287/orsc.1070.0294

Trevor, C. O., Gerhart, B., & Boudreau, J. W. (1997). Voluntary turnover and job performance: Curvilinearity and the moderating influences of salary growth and promotions. *Journal of Applied Psychology*, *82*, 44-61. doi:10.1037/0021-9010.82.1.44

Vallerand, R. J., Blanchard, C., Mageau, G. A., Koestner, R., Ratelle, C., Léonard, M., . . . Marsolais, J. (2003). Les passions de l'ame: On obsessive and harmonious passion. *Journal of Personality and Social Psychology*, *85*(4), 756-767.

TPM®

THE ROLE OF PSYCHOLOGICAL OWNERSHIP IN THE RELATIONSHIP BETWEEN LEADER-MEMBER EXCHANGE AND JOB SATISFACTION

MIRNA MATIC
MARCO GIOVANNI MARIANI
UNIVERSITY OF BOLOGNA

MATTEO CURCURUTO
LEEDS BECKETT UNIVERSITY

PILAR GONZÁLEZ-NAVARRO
ROSARIO ZURRIAGA
UNIVERSITY OF VALENCIA

The purpose of this research is to expand previous work in the field of psychological ownership (PO) by introducing a model where PO plays a mediator role between leader-member exchange (LMX) and job satisfaction. The data were collected in a furniture manufacturing company located in Italy. Sample size was 442 participants. Results showed a positive relationship between LMX and psychological ownership and between job satisfaction and psychological ownership in an organizational context. The results are stable with respect to various tenure levels, but different depending on the job position in the organization: for staff employees, PO has a fully mediating role, whereas for line employees, it is a partial mediator.

Key words: Psychological ownership; Leader-member exchange; Job satisfaction; Organizational tenure; Job position in organization.

Correspondence concerning this article should be addressed to Marco Giovanni Mariani, Department of Psychology, University of Bologna, Viale Berti Pichat 5, 40100 Bologna (BO), Italy. Email: marcogiovanni.mariani@unibo.it

When ownership is mentioned, society immediately thinks about formal ownership, which is specified and protected by the legal system (Pierce, Kostova, & Dirks, 2003). The focus of this research was to understand the construct of ownership as a psychological process and explain its association with the organizational context in a non-family organizational setting. Psychological ownership (PO) can be described as a feeling recognized by the individual who holds it. It can be defined as a state in which an individual feels as though the target of ownership, or a piece of that target, is "his or hers" (Pierce et al., 2003). The main question PO answers is "What do I feel is mine?" (Pierce & Jussila, 2011, p.16).

Psychological ownership is distinct from legal ownership in several ways. Specifically, a person can believe and feel that something is his/hers even though there is no legal right to the

material or non-material object in question (Pierce & Jussila, 2011). Furthermore, in the absence of legal ownership, the role of PO becomes important in creating a positive relationship related to the financial performance of the organization (Wagner, Parker, & Christiansen, 2003). In an organizational context, PO is a psychologically experienced phenomenon where an employee develops possessive feelings toward the target (Van Dyne & Pierce, 2004). The concept is important because it is a form of emotional attachment to the organization that exceeds the simple cognitive evaluation of the firm and can have both positive and negative consequences in the organizational context.

This research aims to expand on previous research in the field of PO. Based on the study by Bernhard and O'Driscoll (2011), who examined PO's mediation between the leadership style and different organizational attitudes and behaviors, this research explores new antecedents of their model. Thus, leader-member exchange (LMX) is added as a new antecedent in the mediation model, with one organizational outcome, affective job satisfaction. Moreover, it adds two moderators to the mediation model: tenure and the job position in the organization. The paper is structured in the following way: first, the literature review describes the construct of PO and its relationships with LMX and job satisfaction. In addition, the literature review contains the mediated model we propose. Then, we focus on the method and the results, followed by the discussion.

THEORETICAL BACKGROUND AND HYPOTHESIS

Psychological Ownership

Psychological ownership has its origins in biological, social, and cultural factors (Dittmar, 1992). It develops to satisfy some basic human needs, such as efficacy and effectance, self-identity, the feeling of home, and a need for stimulation (Pierce & Jussila, 2011). Other factors that influence the development of PO are target factors, individual factors, and situational forces (Pierce et al., 2003). Employees who exert high levels of ownership will promote good citizenship behavior, the voluntary behavior that contributes to the community's well-being and is intended to be positive in nature and has no promised retribution (Organ, 1988). The sense of responsibility, stewardship, personal sacrifice, and risk acceptance will also be higher in employees with high levels of PO because they will see the organization as something that partly belongs to them. By contrast, several negative aspects, such as alienation, frustration, and stress, can arise from PO (Pierce et al., 2003).

Psychological ownership develops through three main paths: control of the target, intimate knowledge about the target, and investment of self. The feeling of control over things has been extensively researched in relation to the feeling of possession, which is the central feeling of PO (Pierce & Jussila, 2011); these things can include ideas, equipment, work, and the organization (Hall, 1966). Coming to know something intimately is a state that develops when things are in the individual's possession over an extended period of time and psychologically become his/hers. Moreover, the person develops a feeling of effectance, motivation, efficacy, and competence, and starts to consider his/her work as home, developing a sense of PO (Pierce & Jussila, 2011). In addition, investment of self is considered an important path to developing PO. It can be defined as an "investment of individual psychology, energy, time, effort, and attention in an ob-

ject that causes the self to become one with the object. This union between self and object emerges because the object has emerged from the self" (p. 84). All these paths have been empirically investigated and confirmed in relation to PO in the organizational context.

Leader-Member Exchange and Psychological Ownership

In a recent review of leadership theory and research, Avolio, Walumbwa, and Weber (2009) stated that it is necessary to carry out studies to research the causal mechanism linking leadership to work-related outcomes. Because most of the research related to antecedents of PO is concerned with work and job design (Pierce, Jussila, & Cummings, 2009), this study aimed to enrich the field by adding a new antecedent, leader-member exchange (LMX). LMX is a leadership framework that assesses the quality of the relationship between a leader and a subordinate, based on the dimensions of respect, trust, and obligation (Graen & Uhl-Bien, 1995). It is based on social exchange theory, which claims that in a high-quality exchange, parties provide valuable assets to each other (Blau, 1964). High satisfaction of the follower in this relationship will have many positive impacts, such as the overall satisfaction of the follower with the leader, increased follower performance, and followers' positive organizational citizenship behavior (OCB; Zacher, Rosing, Henning, & Frese, 2011). By contrast, if there is low-quality LMX, where the leader only gives the follower basic information necessary for job performance and fulfillment, the follower's performance and organizational citizenship will be lower (Zacher et al., 2011).

Leader-member exchange was found to have a positive relationship with job satisfaction in several studies (see Baranik, Roling, & Eby, 2010; Larsen, Marnburg, & Øgaard, 2012; Scandura, 1999). In the present study, we wanted to find out whether this direct positive relationship is mediated by the intervening process of PO.

The explanation of the connection between LMX and PO comes from the PO theory by Pierce and colleagues (2003), who argue that investment of self in a target is one of the paths that develops ownership feelings. However, the investment of self takes a long time to develop. For a follower, it will take a long time to trust the leader and develop a sense of the way the leader invests in him or her. When this happens, the follower will tend to develop a sense of investment of self while building a better relationship with the leader. Moreover, in the absence of legal ownership, a sense of PO is more likely to develop, establishing the basis for an LMX-PO relationship. On the other hand, employees who receive less attention and fewer rewards are managed by formal rules and policies, and not through direct communication with the leader (Lunenburg, 2010). Thus, they do not feel that the leader is investing in them, or feel the obligation to give something back. In fact, it is quite unlikely that they will invest in the relationship or develop a sense of PO. Based on the literature review, we formed our first hypothesis.

Hypothesis 1: the quality of the LMX relationship will have a positive effect on PO.

Job Satisfaction and Psychological Ownership

In 1976, Locke defined job satisfaction as "a pleasurable or positive emotional state resulting from the appraisal of one's job or job experience" (p. 1300). This response can be related

to the job (organization) as a whole or to any job/organizational context (job security, supervision, amount of work, etc.). In other words, if any aspect of the work provides an employee with a feeling of pleasure or displeasure, he/she will consequently be satisfied or dissatisfied with that job facet. This research is particularly interested in the affective component of job satisfaction and its relationship with PO. If the job as a whole provides an employee with a feeling of pleasure or displeasure, he/she will be satisfied or dissatisfied with the job in general (Pierce & Jussila, 2011). In fact, research has shown that a high level of job satisfaction will produce self-identification with the organization (e.g., Marletta et al., 2014). According to this argument, the theoretical basis for the positive relationship between PO and job satisfaction is built on the sense of possessiveness that develops toward the organization. Supporting this, a corpus of research states that it is logical to expect employees to like or feel satisfaction with their jobs when they feel PO over them (e.g., Mayhew, Ashkansay, Bramble, & Gardner, 2007). This relationship was also found in Southeast Asian countries (e.g., Md-Sidin, Sambasivan, & Muniandy, 2010). Based on the literature presented, the second hypothesis was formed.

Hypothesis 2: there will be a positive effect of PO on job satisfaction.

The Mediation Role of Psychological Ownership

The basis of our model and the most similar research to date comes from the research conducted by Bernhard and O'Driscoll (2011). In their research, the two authors proposed a model consisting of two mediators (organization-based and job-based PO), three types of leadership styles (transformational, transactional, and passive), and several outcomes (extra-role behavior, in-role behavior, affective commitment, job satisfaction, and turnover intention).

Bernhard and O'Driscoll (2011) proposed an explanation for the relationship between leadership style and job satisfaction, indicating that this association operates through the mediating effect of PO. The researchers found that leadership styles (transformational and transactional) had a significant positive relationship with organization-based PO. Their finding confirmed the role of transformational and transactional leadership as contextual factors in the development of PO. Passive leadership did not show a positive relationship with PO; therefore, the aim of the present study is to find out whether the quality of the relationship can provide additional insight into the antecedents of the emergence of a state of PO. The second important finding relevant to the current research is that organization-based PO mediated the relationship between transformational and transactional leadership and job satisfaction. Based on their research, the authors concluded that family-owned businesses can benefit from the recognition of the favorable effects of PO on individuals' attitudes and behaviors.

One of the main reasons for looking at PO as a mediator stems from Avolio et al. (2009), who stated that "determining the causal mechanisms that link leadership to outcomes will be a priority" (p. 442) for future research. We also believe that a positive relationship between LMX and job satisfaction fosters a sense of attachment toward the job and, consequently, tends to reinforce the development of a sense of PO toward the organization.

Hypothesis 3: PO will mediate the relationship between LMX and job satisfaction (see Figure 1).

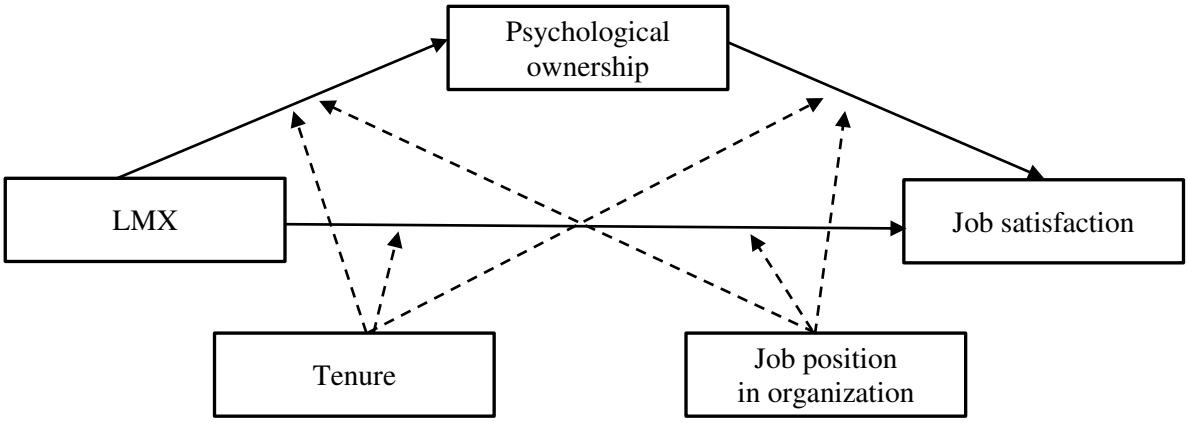

FIGURE 1
The hypothesized model.

An additional research question in this study is whether the tested mediation model is stable when adding variables such as tenure and the job position in the organization. The literature shows that tenure and working hours have an influence on the path of investment of self and, therefore, on PO (e.g., O'Driscoll, Pierce, & Coghlan, 2006; Pierce et al., 2003). Furthermore, the job position in the organization is also found to be associated with PO. Generally, senior employees or those with high positions may experience a strong fit and PO because of long-term integration of organizational and personal values (Han, Chiang, & Chiang, 2013). So, the research questions are: is the mediation role of ownership depending on the participants' tenure and job position in the organization?

METHOD

Participants and Procedure

The data were collected using a questionnaire in a furniture manufacturing company located in the northeastern part of Italy. The sample contained 442 participants. A group of participants came from staff management, with a percentage of 28.3%. The second group was classified together, and they were all in charge of production, as part of line management, with a percentage of 71.7%. Regarding the average time working in the organization, 53.3% had worked in the organization for more than six years, and 46.7% had worked there from zero to six years. A structured anonymous questionnaire was used to collect the data. The study guaranteed respondents' anonymity and confidentiality. The questionnaire included a statement about personal data treatment, in accordance with the Italian privacy law (Legislat. Decree DL-196/2003). The workers authorized and approved the use of anonymous/collective data for possible future scientific publications. After completion, the questionnaires were handed in at "ballot boxes." All questionnaires were filled in and collected in one week in order to control threats to validity from external events that may affect participants' responses. Age and gender were not requested in order to guarantee the anonymity of the participants.

Instruments

To measure ownership, the Psychological Ownership Scale, developed by Van Dyne and Pierce (2004), was used. There are two versions of the scale, one with seven items and one with four items. The second version was adopted because it shows better psychometric quality (Mariani, Martuscelli, & Curcuruto, 2015; Van Dyne & Pierce, 2004). Therefore, the Italian version of the 4-item scale, validated by Mariani et al. (2015), was used. Items such as "This is MY organization" or "I feel that this is MY company" are rated on a 5-point Likert scale with responses ranging from *strongly disagree* to *strongly agree*. Cronbach's alpha was .84 for the entire scale.

The LMX-7 scale for subordinates (Scandura & Graen, 1984) was used to evaluate the subordinates' relationship with their supervisor (dyadic exchange). According to Graen and Uhl-Bien (1995), the scale was developed from several levels of analysis: "from focus on differences within groups (group-level effect), the focus on dyads regardless of groups (dyad-level effect), to the focus on the combination of dyads into groups and networks (dyads within group effect)" (pp. 220). The scale consists of seven questions rated on a 5-point Likert scale with responses ranging from *not at all* to *a great deal*. LMX is both transactional and transformational, and the scale was used to assess the quality of the relationship between the leader and the follower from the follower's point of view. Some of the questions were: "How well does your leader understand your job problems and needs?"; "I have enough confidence in my leader that I would defend and justify his/her decision if he/she were not present to do so"; and "How would you characterize your working relationship with your leader?". Cronbach's alpha was .92.

The Michigan Organizational Assessment Questionnaire-Job Satisfaction Subscale (MOAQ-JSS; Cammann, Fichman, Jenkins, & Klesh, 1979) was used to measure employees' job satisfaction. The subscale consists of three items rated on a 5-point Likert scale with a response range from *strongly disagree* to *strongly agree*. The scores on the MOAQ-JSS are computed using the average scores on the following three items: "All in all, I am satisfied with my job"; "In general, I don't like my job"; and "In general, I like working here." Cronbach's alpha was .92.

An exploratory factor analysis (EFA) with principal axis factoring was performed on all 14 item-variables on the three scales to test discriminant and convergent validity. The Kaiser method suggested a 3-factor solution that explained 70% of the variance and showed, using oblimin rotation, that all the loadings were above .64 in the original factor (convergent validity). All loadings in different factors were lower than .15 (discriminant validity).

The adaptation of the LMX-7 scale for subordinates and the Michigan Organizational Assessment Questionnaire to Italian was performed taking into account the international methodological standards recommended by the International Test Commission (ITC) when adapting an instrument to a foreign language (Hambleton, 2005). However, a qualitative pilot study was carried out with six employees from the company to evaluate the language forms and ensure proper understanding of all the scales. We added some socio-organizational variables such as the work sector and the years an employee had worked in the organization.

Data Analysis

The statistical analysis plan consisted of the following steps: 1) calculation of the descriptive statistics and correlation indexes of the variables, 2) examination of the common meth-

od effect, and 3) mediation analysis and moderated mediation analysis. The degree to which common-method variance could be a threat to our analyses was analyzed because a one-wave self-report design was used. Harman's single-factor test through confirmatory factor analysis (CFA) was performed to check the hypothesis that a single factor can account for all of the variance in our data (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Moreover, the latent method factor approach recommended by Podsdakoff et al. (2003) was adopted to further analyze this issue. Hence, a CFA was conducted in which an unmeasured latent method construct (ULMC) was added to the measurement model. Items were allowed to load on their theoretical constructs and on a latent common method variance factor. The structural parameters were examined, both with and without the latent common method variance factor. This further analysis made it possible to isolate the variance of the measures into theoretical, methodological, and random error components.

Before we could analyze the mediating role of ownership, the three conditions suggested by Baron and Kenny (1986) were tested. Mediation analyses were performed on SPSS 22 using PROCESS (Model 4), provided by Hayes (2013). PROCESS was conducted using one independent variable (LMX), one mediator (PO), and one dependent variable (job satisfaction). The variables in the proposed model were mean-centered to minimize multicollinearity. Mediation was tested using the contemporary bootstrapping technique described by Hayes, 5,000 resamples with replacement. Bootstrapping was used because it provides a more reliable estimate of indirect effects, and it does not make the often unrealistic assumption of normality in the sampling distribution. Therefore, it is appropriate for data with skewness and kurtosis values greater than two standard errors, probably showing a significant difference compared to normal distribution (Table 1). In addition, the bootstrapping method also has higher power and better Type I error control than other mediation analyses (Preacher & Hayes, 2008). Significance was determined by examining the 95% confidence interval produced by the bootstrapping mediation analyses. In order to consider the mediation model significant, the confidence interval must not include zero. Furthermore, the indirect effect size was estimated by Fairchild, MacKinnon, Taborga, and Taylor's (2009) R^2_{med} formula, which measures the proportion of variance in Y attributable to the indirect effect of X on Y through M. Finally, the Kappa-squared index by Preacher and Kelly (2011) was used to compute the ratio of the indirect effect, compared to its maximum possible value in the data. K^2 is bound between 0 and 1, with a value closer to 1 representing a greater indirect effect. Finally, a moderation-mediation model was tested with the PROCESS procedure (Hayes, 2013), considering two control variables: participants' tenure and job position in the organization. In this case, the 76° PROCESS model was adopted.

RESULTS

Table 1 shows descriptive statistics and correlation indexes for LMX, PO, and job satisfaction. Results show strong positive relationships among the three variables. Before testing our hypotheses, and considering the one-wave self-report study design, common-method variance bias was analyzed. Harman's single-factor test of common-method variance showed fit indexes that were not adequate for a one-factor model (CFI = .68; RMSEA = .16). Then, the ULMC method was used to analyze common-method bias. Two CFAs were performed, one where the CFA broke down the variance of the measures into theoretical, methodological, and random error

components. The results showed that 54% of the variance was explained by the three theoretical constructs, 22% by the ULMC, and the remaining 24% by random errors. Both model fits were adequate (model with ULMC: CFI = .96; RMSEA = .06; model without ULMC: CFI = .95; RMSEA = .07). Thus, the hypothesis that common method variance could explain a substantial amount of covariance among the variables was rejected.

TABLE 1
Descriptive statistics, reliability, and intercorrelations for scores on three scales (N = 442)

Scales	Items	Min-Max	*M*	*SD*	Skewness	Kurtosis	Alpha	1	2
LMX	7	1-5	3.56	1.10	–0.63	–0.49	.92	–	
PO	4	1-5	3.16	1.04	–0.20	–0.71	.84	.47	–
Job satisfaction	3	1-5	4.16	0.92	–0.16	–0.45	.82	.39	.44

Note. LMX = leader-member exchange; PO = psychological ownership; Standard error (*SE*) is .12 for skewness and .23 for kurtosis; all correlations are significant at $p < .001$.

The correlations met the three conditions for mediation analysis (Baron & Kenny, 1986). There was a significant association between the predictor and the dependent variable (r = .39), the predictor and the hypothesized mediator (r = .47), and the hypothesized mediator and the dependent variable (r = .44). According to these prerequisites, the mediating role of PO could be studied. We tested whether PO mediated the effect between LMX and job satisfaction (see Table 2).

TABLE 2
Total, direct, and indirect effects of mediation model (N = 442)

Effects	*B*	*SE*	*t*	*p*
Effect of LMX on PO	.482	0.043	11.283	.000
Effect of PO on job satisfaction	.270	0.039	6.961	.000
Total effect LMX on job satisfaction (path c)	.322	0.037	8.798	.000
Direct effect LMX on job satisfaction controlling for PO	.192	0.039	4.854	.000

Note. LMX = leader-member exchange; PO = psychological ownership; B = unstandardized regression coefficient; SE = standard error.

The total effect of LMX on job satisfaction was significant (B =.322, SE= 0.037, $p < .001$, R^2 = .14), as was the effect of LMX on PO (B = .482, SE = 0.043, $p < .001$, R^2 = .24) and the effect of PO on job satisfaction (B =.270, SE = 0.039, $p < .001$). The effect of LMX on job satisfaction remained significant when PO was included in the model (B =.192, SE = 0.039, $p < .001$). The Sobel test of the indirect effect, which assesses whether the total effect of LMX on job satisfaction is reduced by the addition of PO to the model, was significant and indicated mediation (z = 5.908, $p < .001$).

The bootstrapping method by Hayes (2013), previously presented, was used to compute a confidence interval around each estimate of the indirect effect. Using 5,000 bootstrapped samples, the estimate of the indirect effect again indicated mediation, with a point estimate of 0.130, Boot *SE* = 0.023; 95% CI [.089, .176]. Evidence for mediation is present because these confidence intervals do not include zero. Results showed an R^2_{med} of .11 and a K^2 of .16. The formula by Fairchild et al. (2009) estimates that the proportion of variance in Y attributable to the indirect effect is about 11%. These are medium-effect sizes according to Cohen's (1988) guidelines.

Finally, we tested the model considering two control variables: participants' tenure and job position in the organization. A moderated mediation model was performed, precisely the PROCESS 76° procedure (Hayes, 2013), which considered LMX as the independent variable, job satisfaction as the dependent variable, PO as the mediation variable, and both tenure and job position in the organization as moderators in every relationship. Results showed that a mediation effect is almost always constant, whereas the direct effect is verified only for production line sectors (Table 3). For staff employees, PO plays a more important role in this model: it is a full mediator between LMX and job satisfaction.

TABLE 3
Moderated mediation analysis with control variables (*N* = 442)

Tenure	Job position	*B*	*SE*	Boot *SE*	*t*	*p*	LLCI	ULCI
Conditional direct effect(s) of X on Y at values of the moderator(s):								
0-6 years	Staff	.173	0.103		1.683	.093	–.029	.375
0-6 years	Line	.210	0.073		2.879	.004	.067	.354
7-32 years	Staff	.156	0.116		1.353	.177	–.071	.383
7-32 years	Line	.194	0.058		3.317	.001	.079	.308
Conditional indirect effect(s) of X on Y at values of the moderator(s):								
0-6 years	Staff	.177		0.065			.084	.343
0-6 years	Line	.128		0.039			.063	.216
7-32 years	Staff	.174		0.065			.068	.331
7-32 years	Line	.126		0.039			.063	.214

Note. SE = standard error; LLCI = lower level confidence interval; ULCI = upper level confidence interval.

In conclusion, the results found that PO fully mediated the relationship between LMX and job satisfaction in staff job position and partially in line job position. Thus, Hypotheses 1, 2, and 3 were supported.

DISCUSSION

In today's organizational settings, companies are concerned with new ways and constructs that can enhance organizational performance and increase financial revenue. Therefore, we believe that exploring new antecedents and expanding the research on the PO construct can provide additional insight into this relevant question.

In fact, the objective of this research was to extend the work done in the field of PO by exploring a new antecedent of PO, LMX. An additional research question was whether the relationship between LMX and job satisfaction is mediated by PO, and if this model remains stable when adding control variables such as tenure and job position in the organization.

Therefore, we tested a new mediation model in which PO mediates the relationship between LMX and job satisfaction. We also tested the model considering two control variables: participants' tenure and job position in the organization (staff and line). Results showed that a mediation effect is almost always constant, whereas the direct effect is found only for line job position.

For staff job position, PO plays a more important role in this model: it is a full mediator between LMX and job satisfaction. The results show that PO fully mediated the relationship between LMX and job satisfaction in staff job position, and partially in line job position.

Theoretical Implication

Stringer (2006) thought that the greater the level of mutual respect, trust, and obligation between a supervisor and a subordinate (LMX), the greater the level of job satisfaction perceived by the subordinate would be. In addition, Bernhard and O'Driscoll (2011) proposed and found that PO also emerges from leadership styles and influences job satisfaction. Therefore, PO can play a mediation role between leadership style and job satisfaction. Moreover, as mentioned above, according to Graen and Uhl-Bien (1995), LMX incorporates both transformational and transactional leadership. Our research presents evidence supporting this theoretical framework.

Practical Implication

Practical evidence shows that when employees work to support their colleagues, and not only directly on the core activities, the quality of the relationship between leader and members does not directly influence job satisfaction, but does through members' feelings of PO. The development of PO seems to be more important for staff job position than for line job position. From a practical perspective, this is an interesting finding because it sheds light on the importance of the development of PO in the organizational context, especially in staff job position. The literature reports practical ways to improve ownership. For instance, Pierce and Jussila (2011) suggest enhancing job control and knowledge about the organization because they are related to the feeling of possession. Another way to increase the level of ownership is to support the employee's perception of the workplace as his/her home, for example, by encouraging the placement of a photo of his/her family in the office.

Finally, it is in the interest of an organization to have its employees score high on job satisfaction because it produces increased motivation and commitment to their jobs (Guglielmi et al., 2016). Based on the findings from the proposed mediation model, more importance should be placed on the development of ownership, as well as on developing a high degree of respect, trust, and obligation between employees and their leader.

Limitations

There are some limitations in this research. First, self-report data was used to measure all the variables. The literature provides support for doing so; for example, Spector (1997) argued that when measuring constructs of emotional and psychological states, self-report methods are appropriate because they describe individuals' subjective states, which shape their behaviors.

Self-report data raised the possibility of common method bias. In order to control for potential common method bias, we followed the recommendations of Podsakoff, Mackenzie, & Podsakoff (2012). First, we assured respondents that their participation was anonymous and that only the researchers would have access to their answers. Second, we adopted reliable measurement scales with reversed items; third, proximal separation between the predictor and criterion variables was adopted in the questionnaire. Moreover, the fact that these dimensions are evidently dissimilar, both conceptually and in terms of their underlying factors, also decreases the risk of common method bias (Brannick, Chan, Conway, Lance, & Spector, 2010). Finally, we used two statistical strategies to control common method bias: Harman's single factor test and the latent method factor approach recommended by Podsakoff et al. (2003). Both analyses found consistent findings, and so we consider common method variance bias to be a limited problem in this study.

The data were collected using a cross-sectional survey. In our case a longitudinal survey would be more interesting because we are hypothesizing causal relationships among the dimensions of LMX, PO, and job satisfaction.

Another drawback is that we did not add additional moderators such as age and gender that could affect the mediation. It could be important to investigate this in the future. In addition, it would be interesting to conduct research in companies located in other countries in Europe that are developing forms of legal ownership in order to see how legal ownership is related to the level of PO and whether our mediation remains the same in organizations where legal ownership is developed.

Conclusion

We can conclude that this research is a good starting point for some further investigation, and that it also has potential practical implications. In fact, if organizations decide to promote PO they can obtain many benefits, such as: enhancement of the overall quality of organizational performance, productivity, and increased satisfaction and performance among employees.

Taking this into consideration, on a practical note, companies could promote the development of PO in order to achieve beneficial outcomes for the organization. However, we have to be cautious about the negative aspects of PO, such as overall possessiveness, alienation, frustration, and stress (Pierce et al., 2003). In conclusion, further investigation should be carried out in order to explore other antecedents of PO and its potential in the organizational context.

Funding

The contribution of Pilar González-Navarro and Rosario Zurriaga was supported by funding from the Generalitat Valenciana (Spain) for research groups of excellence (GVPROMETEO2012-048).

REFERENCES

Avolio, B. J., Walumbwa, F. O., & Weber T. J. (2009). Leadership: Current theories, research, and future directions. *Annual Review of Psychology*, *60*, 421-449. doi:10.1146/ 60.110707.163621

Baranik, L. E., Roling, E. A., & Eby, L. T. (2010). Why does mentoring work? The role of perceived organizational support. *Journal of Vocational Behavior*, *76*, 366-373. doi:10.1016/j.jvb.2009.07.004

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173-1182. doi:10.1037/0022-3514.51.6.1173

Bernhard, F., & O'Driscoll, M. P. (2011). Psychological ownership in small family-owned business: Leadership style and nonfamily-employees' work attitude and behaviors. *Group & Organization Management*, *36*, 345-384. doi:10.1177/1059601111402684

Blau, P. (1964). *Exchange and power in social life*. New York, NY: Wiley.

Brannick, M. T., Chan, D., Conway, J. M., Lance, C. E., & Spector, P. E. (2010). What is method variance and how can we cope with it? A panel discussion. *Organizational Research Methods*, *13*(3), 407-420.

Cammann, C., Fichman, M., Jenkins, D., & Klesh, J. (1979). *The Michigan Organizational Assessment Questionnaire*. Unpublished manuscript, University of Michigan, Ann Arbor, MI.

Cohen, J. (1988). *Statistical power analysis for the behavioral science*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Dittmar, H. (1992). *The social psychology of material possessions: To have is to be.* New York, NY: St.Martin's Press.

Fairchild, A. J., MacKinnon, D. P., Taborga, M. P., & Taylor, A. B. (2009). R2 effect sizes measures for mediation analysis. *Behavior Research Methods*, *41*, 486-498. doi:10.3758/BRM.41.2.486

Graen, G. B., & Uhl-Bien, M. (1995). Relationship-based approach to leadership: Development of leader-member exchange (LMX) theory of leadership over 25 years: Applying a multi-level multi-domain perspective. *Leadership Quarterly*, *6*, 219-247. doi:10.1016/1048-9843(95)90036-5

Guglielmi, D., Avanzi, L., Chiesa, R., Mariani, M. G., Bruni, I., & Depolo, M. (2016). Positive aging in demanding workplaces: The gain cycle between job satisfaction and work engagement. *Frontiers in Psychology*, *7*, 1-10. doi:10.3389/fpsyg.2016.01224.

Hall, D. T. (1966). *The hidden dimension*. Garden City, NY: Doubleday.

Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3-38). Mahwah, NJ: Lawrence Erlbaum.

Han, T. S., Chiang, H. H., & Chiang, C. L. (2013, September). *The investigation of the relation between person-organization fit, person-job fit, psychological ownership, and contextual performance: A longitudinal approach*. Proceedings of International Conference on Information and Social Science (ISS) & International Symposium on Marketing, Logistics, and Business (MLB), Nagoya, Japan.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Publications.

Larsen, S., Marnburg, E., & Øgaard, T. (2012). Working onboard-job perception, organizational commitment and job satisfaction in the cruise sector. *Tourism Management*, *33*, 592-597. doi:10.1016/j.tourman.2011.06.014

Legislative Decree DL-196/2003. *Personal data protection code.* Retrieved from http://194.242.234.211/documents/10160/2012405/Personal+Data+Protection+Code+-+Legislat.+Decree+no.196+of+30+June+2003.pdf

Locke, E. A. (1976). The nature and causes of job satisfaction. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 1297-1349). Chicago, IL: Rand-McNally.

Lunenburg, F. C. (2010). Leader-member exchange theory: Another perspective on the leadership process. *International Journal of Management, Business, and Administration*, *13*(1), 1-5.

Marletta, G., Caricati, L., Mancini, T., La Sala, R., Pelosi, G., Ampollini, M., . . . Sarli, L. (2014). Professione infermieristica: Stati dell'identità e soddisfazione lavorativa [Nursing profession: Identity's transitions and job satisfaction]. *Psicologia della Salute*, *2*(1), 139-157.

Mariani, M. G., Martuscelli, C., & Curcuruto, M. (2015). Ownership psicologica nelle organizzazioni: contributo alla validazione italiana della scala di Pierce e colleghi [Psychological ownership in organization: A contribution to the Italian validation of Pierce and colleagues' scale]. *Giornale Italiano di Psicologia*, *XLII*, 329-349. doi:10.1421/79852

Mayhew, M., Ashkansay, N. M., Bramble, T., & Gardner, J. (2007). A study of the antecedents and consequences of psychological ownership in organizational settings. *The Journal of Social Psychology*, *147*, 477-500. doi:10.3200/SOCP.147.5.477-500

Md-Sidin, S., Sambasivan, M., & Muniandy, N. (2010). Impact of psychological ownership on the performance of business school lecturers. *Journal of Education for Business*, *85*, 50-56. doi:10.1080/08832320903217903

O'Driscoll, M. P., Pierce, J. L., & Coghlan, A. (2006). Work environment structure, organizational commitment, and citizenship behaviors. *Group & Organization Management*, *31*, 388-416. doi:10.3200/SOCP.144.5.507-534

Organ, D. W. (1988). *Organizational citizenship behavior: The good soldier syndrome*. Lexington, MA: Lexington Books.

Pierce, J. L., & Jussila, I. (2011). *Psychological ownership and the organizational context: Theory, research evidence, and application.* Northampton, MA: Edward Elgar Publishing.

Pierce, J. L., Jussila, I., & Cummings, A. (2009). Psychological ownership within the job design context: Revision of the job characteristics model. *Journal of Organizational Behavior*, *30*, 477-496. doi:10.1002/job.550

Pierce, J. L., Kostova, T., & Dirks, K. T. (2003). The state of psychological ownership: Integrating and extending a century of research. *Review of General Psychology*, 84-107. doi:10.1037/1089-2680.7.1.84

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*(5), 878-903.

Podsakoff, P. M., Mackenzie, S. B., & Podsakoff, N. P. (2012). Sources/causes of common method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, *63*, 539-569. doi:10.1146/annurev-psych-120710-100452

Preacher, J. K., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*, 879-891. doi:10.3758/BRM.40.3.879

Preacher, K. J., & Kelly, K. (2011). Effect size measures for mediation models: Quantitative strategies for communication indirect effects. *Psychological Methods*, *16*, 93-115. doi:10.1037/a0022658

Scandura, T. A. (1999). Rethinking leader member exchange: An organizational justice perspective. *Leadership Quarterly*, *10*, 25-40. doi:10.1016/S1048-9843(99)80007-1

Scandura, T. A., & Graen, G. B. (1984). Moderating effects of initial leader-member exchange status on the effects of a leadership intervention. *Journal of Applied Psychology*, *69*, 428-436. doi:10.1037/0021-9010.69.3.428

Spector, P. E. (1997). *Job satisfaction: Application, assessment, causes, and consequences*. Thousand Oaks, CA: Sage.

Stringer, L. (2006). The link between the quality of the supervisor-employee relationship and the level of the employees job satisfaction. *Public Organization Review*, *6*, 125-142. doi:10.1007/s11115-006-0005-0

Van Dyne, L., & Pierce, J. L. (2004). Psychological ownership and feelings of possession: Three filed studies predicting attitudes and organizational citizenship behavior. *Journal of Organizational Behavior*, *25*, 439-459. doi:10.1002/job.249

Wagner, S. H., Parker, C. P., & Christiansen, N. D. (2003). Employees that think and act like owners: Effect of ownership beliefs and behaviors on organizational effectiveness. *Personnel Psychology*, *56*(4), 847-871.

Zacher, H., Rosing, K., Henning, T., & Frese, M. (2011). Establishing the next generation at work: Leader generatively as a moderator of the relationship between leader age, leader-member-exchange, and leadership success. *Psychology and Aging*, *26*, 241-252. doi:10.1037/a0021429

TPM®

SELF-REPRESENTATIONS, BURNOUT SYNDROME, AND JOB SATISFACTION AMONG CORRECTIONAL OFFICERS

CLAUDIA CASTIGLIONE
UNIVERSITY OF MESSINA

ALBERTO RAMPULLO
UNIVERSITY OF CATANIA

CHIARA GIOVINCO
UNIVERSITY OF MESSINA

The present study examines the relationship between self-representation discrepancies (future self vs. correctional officer), the burnout syndrome, and job satisfaction among a sample of correctional officers. Results showed that burnout decreased job satisfaction. Furthermore, self-representation discrepancies between future self and correctional officer decreased job satisfaction and increased burnout levels. Finally, burnout mediated the relationship between self-representation discrepancies and job satisfaction. Higher discrepancies decreased satisfaction with one's job through increasing burnout levels.

Key words: Prison staff; Correctional officers; Possible selves; Job satisfaction; Job burnout.

Correspondence concerning this article should be addressed to Claudia Castiglione, Department of Clinical and Experimental Medicine, University of Messina, Via Bivona, 98125 Messina (ME), Italy. Email: ccastiglione@unime.it

Correctional officers work daily to guarantee not only the safety of inmates but also to promote their rehabilitation through treatment activities. These conflicting roles and goals might expose correctional officers to highly stressful situations (Blau, Light, & Chamlin, 1986; Cheeck & Miller 1983; Prati & Boldrin, 2011; Tewksbury & Higgins, 2006). Indeed, correctional workers experience high levels of stress and burnout due to several factors. Main factors are prison overcrowding, lack of shift flexibility, role conflicts, and conflicts with administrators (Keinan & Malach-Pines, 2007; Schaufeli & Peeters, 2000; Walters, 1996). High levels of stress and burnout lead to job dissatisfaction, physical diseases, or family problems, which make prison officers unable to perform effectively their functions (Cheek & Miller, 1983; Lindquist & Whitehead, 1986; Schaufeli & Peeters, 2000). While most of the studies on burnout among correctional workers focused on organizational and environmental factors (Schaufeli & Peeters, 2000), literature on burnout syndrome among helping professions (e.g., teachers and nurses) showed that some individual factors (e.g., self-representations) play an important role on burnout development (Brouwers & Tomic, 2000; VanYperen, 1998). Self-representations drive cognitions, emotions, and behaviors (Markus & Nurius, 1986; Markus & Ruvolo, 1989; Ruvolo & Markus, 1992) and can have positive effects on burnout (Cao, Chen, Tian, Diao, & Hu, 2015) and job satisfaction levels (Cowin, Johnson, Craven,

& Marsh, 2008). Specifically, distance among self-representations affected emotions (Hart, Field, Garfinkle, & Singer, 1997; Heppen & Ogilvie, 2003; Higgins, 1989), and behaviors (Hart & Fegly, 1995). Few studies focused on the relationship between self-representation distance and job-related factors (Reich & Rosenberg, 2004; Sheldon, Ryan, Rawsthorne, & Ilardi, 1997).

Previous research on burnout mainly focused on helping professions (e.g., teachers and nurses), and the few studies related to correctional officers focused mainly on environmental factors (Schaufeli & Peeters, 2000). Thus, the present study aimed to extend previous literature, examining the role of individual factors and considering correctional officers, that is, a sample which is quite difficult to reach. Specifically, the present study explored, for the first time to our knowledge, the role of self-representation discrepancies (future self vs. correctional officer representation) on burnout syndrome and job satisfaction, among a sample of Italian correctional officers.

THE BURNOUT SYNDROME AMONG CORRECTIONAL OFFICERS

Freudenberger (1975) used for the first time the term burnout, referring to a sense of exhaustion and incapacity to deal with work demands. Maslach and Jackson (1981) defined burnout as "a syndrome of emotional exhaustion and cynicism" (p. 99) with psychological, behavioral, and physically negative outcomes for employees. The burnout syndrome is a reaction to continuing interpersonal and emotional job stress and is characterized by emotional exhaustion, depersonalization, reduced personal and professional accomplishment (Maslach, 1981; Maslach & Jackson, 1981; Maslach, Schaufeli, & Leiter, 2001), and finally, a sense of disillusionment, as a loss of enthusiasm and passion (Santinello, 2007). Burnout syndrome has been found especially in the context of helping professions, like nurses, teachers, and among correctional officers (Garland, 2002; Griffin, Hogan, & Lambert, 2012; Pines & Aronson, 1988; Prati & Boldrin, 2011).

Scholars have studied both consequences and antecedents of the burnout syndrome. Focusing on its consequences, researchers found that burnout is linked to both psychological and physical health symptoms. Regarding physical health problems, burnout has been found to be related to symptoms, such as headaches, insomnia, fatigue, and nightmares (Belcastro, 1982; Elman & Dowd, 1997; Maslach, 1981). Burnout negatively affects also psychological well-being. In particular, the burnout syndrome has been related to feelings of guilt, depression, anxiety, negative self-esteem, irritability (Belcastro, Gold, & Grant, 1982; Elman & Dowd, 1997; Honkonen et al., 2006; Maslach et al., 2001; Schulz et al., 2011).

Besides psychological and physical consequences of the burnout syndrome, researchers found that burnout has several negative effects on job quality, too. Indeed, the burnout syndrome negatively affects job satisfaction (Lambert, Hogan, & Altheimer, 2010; Shanafelt et al., 2009; Soler et al., 2008; Wolpin, Burke, & Greenglass, 1991; Ybema, Smulders, & Bongers, 2010), work performance (Shanafelt et al., 2009; West et al., 2006), absenteeism (Borritz et al., 2006; Duijts, Kant, Swaen, van den Brandt, & Zeegers, 2007; Maslach et al., 2001), intentions to leave the job, and turnover (Leiter & Maslach, 2009; Soler et al., 2008).

Fewer studies explored the consequences of burnout among prison staff. Burnout among correctional officers increases turnover (Belcastro et al., 1982; Carlson & Thomas, 2006; Lambert, Barton-Bellessa, & Hogan, 2015; Lambert, Hogan, & Altheimer, 2010), absenteeism (Lambert, Hogan, & Altheimer, 2010), physical symptoms (Belcastro et al., 1982), and decreases satisfaction with one's own life (Lambert, Hogan, & Altheimer, 2010; Lambert et al., 2015). Several research-

ers instead focused mainly on the antecedents of the burnout syndrome among correctional workers, linking it to several work stressors. Main factors were prison overcrowding, lack of shift flexibility, role conflicts, and conflicts with administrators (Cheek & Miller, 1983; Lindquist & Whitehead, 1986; Schaufeli & Peeters, 2000). Furthermore, burnout among correctional workers has been linked to organizational and work environment factors, such as promotional opportunity, formalization, and instrumental communication (Dollard & Winefield, 1998; Lambert, Hogan, & Jiang, 2010). Lack of social support both from managers and from coworkers showed to be another important antecedent of the burnout syndrome (Cieslak, Korczynska, Strelau, & Kaczmarek, 2008; Garland, 2004; Neveu, 2007; Lambert, Hogan, & Jiang, 2010). Finally, some characteristics of the job, like variety and autonomy, increased burnout levels (Griffin et al., 2012; Lambert, Hogan, Dial, Jiang, & Khondaker, 2012).

Although most of the antecedents of burnout syndrome have been linked to organizational and environmental factors (Dollard & Winefield, 1998; Lambert, Hogan, & Jiang, 2010), some researchers explored how some individual factors (Cowin et al., 2008; Thomsen, Soares, Nolan, Dallender, & Arnetz, 1999; Villa & Calvete, 2001) might affect positively the burnout syndrome by carrying out the function of coping factors. In particular, an interesting and promising field of research focused on the role of the self-representation. How professionals perceive and represent themselves can affect several work-related outcomes; for example, a positive self-representation increased job satisfaction levels, job performance, and decreased both turnover and intentions to quit among nurses (Cowin et al., 2008; Takase, Maude, & Manias, 2006). Positive self-representation was also related to less burnout in both nurses (Cao et al., 2015; Cao, Lu, & Liu 2010; Consiglio, Borgogni, Vecchione, & Maslach, 2014; Thomsen et al., 1999) and teachers (Liu & Qin, 2005; Villa & Calvete 2001).

Self-Representation Discrepancies

Behaviors, thoughts, and emotions, are affected and shaped by the representation that each individual elaborates about him/herself. Representations may refer to several domains, for instance, beliefs about ideal self, ought self, undesired selves, or future selves (Hart et al., 1997; Hewitt & Genest, 1990; Higgins, Klein, & Strauman, 1985; Markus & Nurius, 1986; Ogilvie, 1987; Strauman & Higgins, 1988). Representations about one's own future self refer to a set of knowledge related to what one might become. When people think of or describe themselves, they will use not only knowledge about traits and characteristics they think they actually possess, but are also able to project themselves into a future dimension. In this sense, the possible selves (Markus & Nurius, 1986) represent that set of representations about how someone will/might become in the future, or is afraid of becoming in the future. Self-representations guide thoughts, emotions, and actions, and facilitate performance through focusing on specific aims and implementing relevant strategies and plans (Castiglione, Licciardello, & Rampullo, 2015; Markus & Nurius, 1986; Markus & Ruvolo, 1989).

Based on the semantic space approach (Hart et al., 1997), multiple self-representations are organized in a semantic space. One way to define the location of self-representations in the individual semantic space is based on the similarity of descriptors of self-representations. The more similar they are, the closer two representations will be in individual semantic space (Hart et al., 1997; Higgins et al., 1985).

Proximity between actual and ideal self may positively affect emotions (Hart et al., 1997; Heppen & Ogilvie, 2003; Higgins, 1989), behaviors (Hart & Fegly, 1995), and self-esteem (Hart et al., 1997), whereas proximity between self- and other-representations may positively affect social relationships (Field, Hart, & Horowitz, 1999). Instead, discrepancies (Higgins et al., 1985) and incongruences (Markus & Nurius, 1986; Rogers, 1961; Turner, 1978) between self-representations increase negative emotions, like disappointment, sadness, stress, anxiety, and depression (Hart et al., 1997; Heppen & Ogilvie, 2003; Higgins, 1989), and decrease life satisfaction and commitment to one's occupational role (Reich & Rosenberg, 2004; Sheldon et al., 1997).

Aim and Hypotheses

The present study focuses on the relationship between discrepancies regarding self-representations, the burnout syndrome, and job satisfaction, among Italian correctional officers. Longitudinal studies on the relationship between burnout and job satisfaction have shown that the burnout syndrome is an antecedent of job satisfaction (Wolpin et al., 1991; Ybema et al., 2010). Thus, our first hypothesis was that burnout would decrease job satisfaction levels.

Another important predictor of job satisfaction is related to the self-representations. Indeed, the self-representation guides thoughts, emotions, and actions (Markus & Nurius, 1986; Markus & Ruvolo, 1989; Ruvolo & Markus, 1992), and may have positive effects on burnout (Cao et al., 2015) and job satisfaction (Cowin et al., 2008), carrying out the function of a coping factor. Self-representations closeness increases positive emotions (Hart et al., 1997; Heppen & Ogilvie, 2003; Higgins, 1989) and prosocial behaviors (Hart & Fegly, 1995), whereas distance (Higgins et al., 1985) increases negative emotions (Hart et al., 1997; Heppen & Ogilvie, 2003; Higgins, 1989). Thus, our second hypothesis was that discrepancies between self-representations (future self vs. correctional officer representation) should decrease job satisfaction and increase burnout levels. Lastly, our third hypothesis was that the relationship between self-representation discrepancies and job satisfaction is mediated by burnout levels. In particular, we predicted that the discrepancy between self-representations should decrease job satisfaction by increasing burnout levels (see Figure 1).

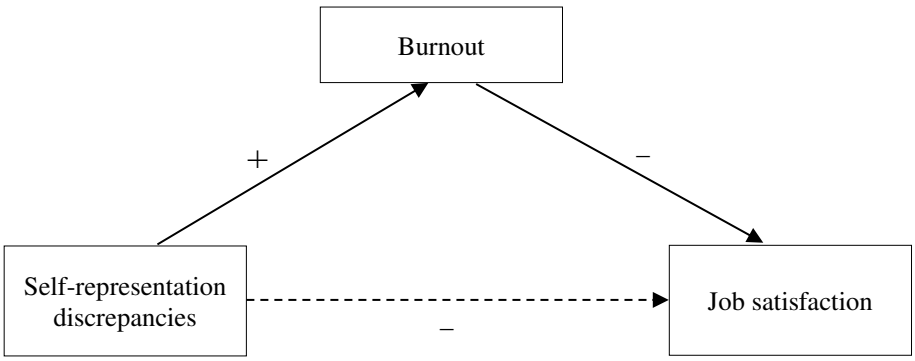

Figure 1
Model in which the discrepancies between future self and correctional officer predict job satisfaction through the mediation of burnout.

METHOD

Participants

Participants were recruited within an Italian correctional facility; a psychology student, under supervision, collected data using a questionnaire administered in individual sessions. Participants compiled an anonymous questionnaire with informed consent presented in the first page. The average time of completion was 15 minutes. The sample consisted of 87 correctional officers. Participants were 79 males and 8 females, with an average age of 44.06 years (*SD* = 7.58; range 25-59) and an average of 19.81 years of work (*SD* = 7.33; range 2-34).

Measures

For data collection, the instruments used were the following.

Link Burnout Questionnaire (LBQ; Santinello, 2007) to measure burnout levels (α = .96). The scale was composed of 24 items about psychophysical exhaustion (e.g., "I feel physically exhausted from my work"), relationship deterioration (e.g., "My inmates seem thankless"), professional inefficacy (e.g., "I feel inadequate to face the problems of my inmates"), and disillusion (e.g., "I doubt that what I do has any value"). Items were rated on a 6-point scale from 1 (*never*) to 6 (*everyday*). Higher scores refer to higher burnout levels.

Self-representation discrepancies (future self vs. correctional officer). Two semantic differentials were used to measure the representations of future self and correctional officer (Osgood, Suci, & Tannenbaum, 1978). They were made up of 34 7-point bipolar scales (e.g., strong-weak) (Castiglione, Licciardello, Mauceri, & Rampullo, 2012; De Caroli & Sagone, 2012). The concepts were "Future self" (as I will be) (α = .95) and "Correctional officer" (α = .88). Based on the semantic space approach (Hart et al., 1997), the location of self-representations in the individual semantic space is based on the similarity of self-representation attributes (Hart et al., 1997; Higgins et al., 1985). Thus, discrepancy was operationalized as Euclidean distance (Hafdahl, Panter, Gramzow, Sedikides, & Insko, 2000; Kirchler, Palmonari, & Pombeni, 1994) between descriptors of the future self and the correctional officer. A value of zero indicates identical descriptions of future self and correctional officer, higher values refer to higher discrepancies between them.

Job satisfaction was measured using a single item (Wanous, Reichers, & Hudy, 1997), self-report measure. Participants were asked how much they were satisfied with their job (i.e., "I am satisfied with my job"); answers were given on a 7-point scale from 1 (*not at all*) to 7 (*completely*).

Background questionnaire. Questions were used to collect information about age, gender, years of work.

RESULTS

Zero-order correlations, means, and standard deviations for discrepancies between self-representations (future self vs. correctional officer), burnout, and job satisfaction are shown in Table 1.

TABLE 1
Means, standard deviations, and correlations

	1	2	3
1. Future self vs. correctional officer discrepancies	1		
2. Burnout	.56**	1	
3. Job satisfaction	–.44**	–.76**	1
M	2.07	3.39	3.75
SD	0.85	1.40	2.42

Note. Burnout, range 1-6; Job satisfaction, range 1-7.
**$p < .001$.

To test our hypotheses we conducted a series of regression analyses. Finally, we tested mediation using Preacher-Hayes bootstrap approach (Preacher & Hayes, 2004); nonparametric bootstrap analyses are actually suitable for small samples (see also Fritz & MacKinnon, 2007; Hayes, 2009). Firstly, we predicted that burnout would decrease job satisfaction. In line with our hypothesis, burnout significantly decreased job satisfaction for our participants, $b = –1.32$, $SE = .12$, $t = –10.805$, $p < .001$, $R^2_{adj} = .57$, $F(1, 85) = 116.747$, $p < .001$. We also predicted that discrepancies between self-representations (future self vs. correctional officer) would decrease job satisfaction and increase burnout levels. In line with our predictions, self-representation discrepancies significantly decreased job satisfaction, $b = –1.26$, $SE = .26$, $t = –4.534$, $p < .001$, $R^2_{adj} = .20$, $F(1, 85) = 20.557$, $p < .001$, and increased burnout levels, $b = .91$, $SE = .17$, $t = 6.181$, $p < .001$, $R^2_{adj} = .31$, $F(1, 85) = 38.201$, $p < .001$.

When self-representation discrepancies (future self vs. correctional officer) and burnout were entered simultaneously in the model, $R^2_{adj} = .58$, $F(2, 84) = 197.464$, $p < .001$, the effect of burnout on job satisfaction was still significant, $b = –1.29$, $SE = .12$, $t(2.84) = –10.462$, $p < .001$, indicating that burnout significantly decreased job satisfaction after controlling for self-representation discrepancies. After controlling for burnout, self-representation discrepancies were no longer a significant predictor, $b = –.07$, $SE = .24$, $t < 1$, indicating that burnout mediates the relationship between self-representation discrepancies and job satisfaction (our third hypothesis).

To test mediation, we used bootstrapping with 10,000 resamples (Preacher & Hayes, 2004). The indirect effect (–1.19, $SE = .23$) of self-representation discrepancies on job satisfaction did not include zero, 95% CI [–1.68, –.75], thus confirming that burnout mediates the relationship between representation discrepancies (future self vs. correctional officer) and job satisfaction (see Figure 2).

DISCUSSION AND CONCLUSION

Correctional officers may experience high levels of burnout with several individual and organizational negative outcomes affecting their performance (Cheek & Miller, 1983; Lindquist & Whitehead, 1986; Schaufeli & Peeters, 2000). While most of studies among correctional workers

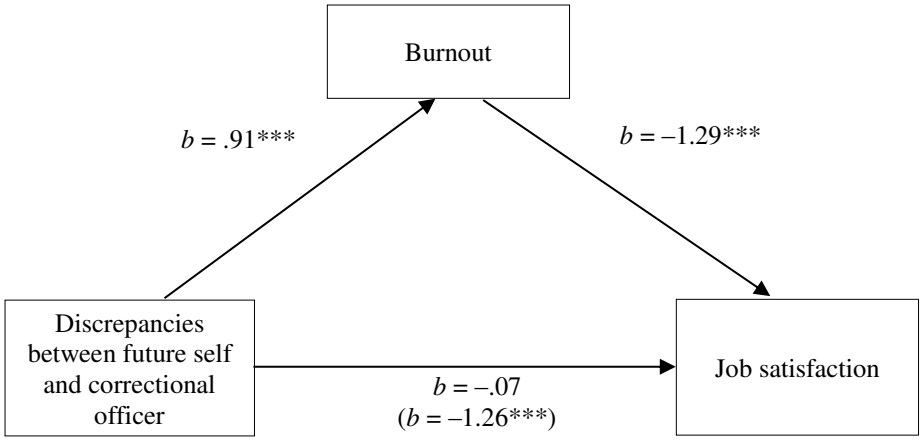

FIGURE 2
Model in which burnout mediates the relationship between self-representation discrepancies and job satisfaction.
The coefficient in parentheses refers to the total effect of representations on job satisfaction. ***$p < .001$.

focused on organizational and environmental dimensions (Schaufeli & Peeters, 2000), some individual factors (e.g., self-representations) might play an important role (Brouwers & Tomic, 2000; VanYperen, 1998). Self-representations guide thoughts, emotions, and actions (Markus & Nurius, 1986; Markus & Ruvolo, 1989). Furthermore, higher closeness between self-representations showed to affect positively individual at an emotional and behavioral level (Hart et al., 1997; Heppen & Ogilvie, 2003), while higher distance (Higgins et al., 1985) has negative effects on individuals through increasing the experience of negative emotions (Hart et al., 1997; Heppen & Ogilvie, 2003; Higgins, 1989), and decreasing life satisfaction (Sheldon et al., 1997). Thus, the present study focused on the relationship between self-representation discrepancies, burnout syndrome, and job satisfaction among a sample of Italian correctional officers.

We predicted that burnout would decrease job satisfaction levels. We also predicted that discrepancies between self-representations (future self vs. correctional officer) would decrease job satisfaction and increase burnout levels. Lastly, we predicted that the relationship between self-representation discrepancies and job satisfaction would be mediated by burnout levels.

Confirming our first hypothesis, burnout significantly decreased job satisfaction levels. This result is in line with literature on the relationship between burnout and job satisfaction, showing that the burnout syndrome may be an antecedent of job satisfaction (Wolpin et al., 1991; Ybema et al., 2010). In line with our prediction, self-representation discrepancies between future self and correctional officer decreased job satisfaction and increased burnout levels. Both results are coherent with literature on distance between self-representations (Hart et al., 1997; Higgins et al., 1985; Markus & Nurius, 1986). Previous researchers indeed showed that self-representation discrepancies are related to less satisfaction with one's own life and with negative emotions, like anxiety and depression (Hart et al., 1997; Heppen & Ogilvie, 2003; Higgins, 1989; Sheldon et al., 1997). Finally, results confirmed our third hypothesis that burnout mediates the relationship between self-representation discrepancies (future self vs. correctional officer) and job satisfaction.

Higher discrepancies between the two representations decreased satisfaction with one's job by increasing burnout levels.

A first limitation of the present study is that job satisfaction was measured using a single item. Even if the single item approach is a robust method (Nagy, 2002; Wanous et al., 1997), future studies should use other scales to deepen the relationship between self-representation discrepancies, burnout, and job satisfaction. Another limitation refers to the correlational design. Previous research addressed the causal paths from self-representation discrepancies to negative emotions and satisfaction (Hart et al., 1997; Heppen & Ogilvie, 2003; Higgins, 1989; Sheldon et al., 1997), and from burnout to satisfaction (Wolpin et al., 1991; Ybema et al., 2010), implementing other research designs (e.g., longitudinal). However, to deepen causal relationships, researchers should also consider experimental designs. Finally, our sample was mostly composed of male correctional workers. Considering the role of gender on burnout levels (Purvanova & Muros, 2010), and job dimensions (Castiglione, Licciardello, Sánchez, Rampullo, & Campione, 2013), future studies should evaluate the effect of gender on the relationship between self-representation discrepancies, burnout, and job satisfaction.

Our findings show that self-representation discrepancies increase burnout levels and decrease job satisfaction. More importantly, our results support the hypothesis that burnout mediates the relationship between self-representation discrepancies and job satisfaction, extending previous knowledge on the relationship between self-representation and burnout, and on discrepancies related to the self. These results also seem important for literature on burnout among correctional workers which has so far focused on consequences and organizational antecedents. Indeed, interventions based only on organizational factors should be extended to include individual difference variables and strategies should be implemented aimed at reducing self-representation discrepancies. Professional psychological counseling could be focused on both improving one's self-representations and supporting a positive representation of correctional officers as a group.

REFERENCES

Belcastro, P. A. (1982). Burnout and its relationship to teachers' somatic complaints and illnesses. *Psychological Reports*, *50*(3), 1045-1046.

Belcastro, P. A., Gold, R. S., & Grant, J. (1982). Stress and burnout: Physiologic effects on correctional teachers. *Criminal Justice and Behavior*, *9*(4), 387-395.

Blau, J. R., Light, S. C., & Chamlin, M. (1986). Individual and contextual effects on stress and job satisfaction: A study of prison staff. *Work and Occupations*, *13*(1), 131-156.

Borritz, M., Rugulies, R., Bjorner, J. B., Villadsen, E., Mikkelsen, O. A., & Kristensen, T. S. (2006). Burnout among employees in human service work: Design and baseline findings of the PUMA study. *Scandinavian Journal of Public Health*, *34*, 49-58. doi:10.1080/14034940510032275

Brouwers, A., & Tomic, W. (2000). A longitudinal study of teacher burnout and perceived self-efficacy in classroom management. *Teaching and Teacher Education*, *16*, 239-253. doi:10.1016/S0742-051X(99)00057-8

Cao, X., Chen, L., Tian, L., Diao, Y., & Hu, X. (2015). Effect of professional self-concept on burnout among community health nurses in Chengdu, China: The mediator role of organizational commitment. *Journal of Clinical Nursing*, *24*, 2907-2915. doi:10.1111/jocn.12915

Cao, X., Lu, L. Q., & Liu, X. H. (2010). The mediating effect of professional self-concept between nurses' professional identification and job burnout. *Chinese Journal of Nursing*, *45*(1), 965-968.

Carlson, J. R., & Thomas, G. (2006). Burnout among prison caseworkers and corrections officers. *Journal of Offender Rehabilitation*, *43*, 19-34. doi:10.1300/J076v43n03_02

Castiglione, C., Licciardello, O., Mauceri, M., & Rampullo, A. (2012). Current experience and future plans in Italian adolescents of secondary schools. *Procedia – Social and Behavioral Sciences*, *46*, 5593-5602. doi:10.1016/j.sbspro.2012.06.481

Castiglione, C., Licciardello, O., & Rampullo, A. (2015). Possible future changes in menopausal women. *Life Span and Disability*, *18*(2), 189-205.

Castiglione, C., Licciardello, O., Sánchez, J. C., Rampullo, A., & Campione, C. (2013). Liquid modernity and entrepreneurship orientation in university students. *Procedia – Social and Behavioral Sciences*, *84*, 1250-1254. doi:10.1016/j.sbspro.2013.06.738

Cheek, F. E., & Miller, M. (1983). The experience of stress for correction officers: A double-bind theory of correctional stress. *Journal of Criminal Justice*, *11*, 105-120. doi:10.1016/0047-2352(83)90046-6

Cieslak, R., Korczynska, J., Strelau, J., & Kaczmarek, M. (2008). Burnout predictors among prison officers: The moderating effect of temperamental endurance. *Personality and Individual Differences*, *45*, 666-672. doi:10.1016/j.paid.2008.07.012

Consiglio, C., Borgogni, L., Vecchione, M., & Maslach, C. (2014). Self-efficacy, perceptions of context, and burnout: A multilevel study on nurses. *Occupational Medicine*, *105*(4), 255-268.

Cowin, L. S., Johnson, M., Craven, R. G., & Marsh, H. W. (2008). Causal modeling of self-concept, job satisfaction, and retention of nurses. *International Journal of Nursing Studies*, *45*, 1449-1459. doi:10.1016/j.ijnurstu.2007.10.009

De Caroli, M. E., & Sagone, E. (2012). Professional self-representation and risk of burnout in school teachers. *Procedia – Social and Behavioral Sciences*, *46*, 5509-5515. doi:10.1016/j.sbspro.2012.06.466

Dollard, M. F., & Winefield, A. H. (1998). A test of the demand-control/support model of work stress in correctional officers. *Journal of Occupational Health Psychology*, *3*, 243-264. doi:10.1037/1076-8998.3.3.243

Duijts, S. F., Kant, I., Swaen, G. M., van den Brandt, P. A., & Zeegers, M. P. (2007). A meta-analysis of observational studies identifies predictors of sickness absence. *Journal of Clinical Epidemiology*, *60*(11), 1105-1115.

Elman, B. D., & Dowd, E. T. (1997). Correlates of burnout in inpatient substance abuse treatment therapists. *Journal of Addictions & Offender Counseling*, *17*, 56-65. doi:10.1002/j.2161-1874.1997.tb00114.x

Field, N. P., Hart, D., & Horowitz, M. J. (1999). Representations of self and other in conjugal bereavement. *Journal of Social and Personal Relationships*, *16*, 407-414. doi:10.1177/0265407599163006

Freudenberger, H. (1975). The staff burn-out syndrome in alternative institutions. *Psychotherapy: Theory, Research, and Practice*, *12*(1), 73-82.

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, *18*, 233-239. doi:10.1111/j.1467-9280.2007.01882.x

Garland, B. (2002). Prison treatment staff burnout: Consequences, causes and prevention. *Corrections Today*, *64*(7), 116-121.

Garland, B. (2004). The impact of administrative support on prison treatment staff burnout: An exploratory study. *The Prison Journal*, *84*, 452-471. doi:10.1177/0032885504269343

Griffin, M. L., Hogan, N. L., & Lambert, E. G. (2012). Doing "people work" in the prison setting: An examination of the job characteristics model and correctional staff burnout. *Criminal Justice and Behavior*, *39*, 1131-1147. doi:10.1177/0093854812442358

Hafdahl, A. R., Panter, A. T., Gramzow, R. H., Sedikides, C., & Insko, C. A. (2000). Free-response self-discrepancies across, among, and within FFM personality dimensions. *Journal of Personality*, *68*, 111-151. doi:10.1111/1467-6494.00093

Hart, D., & Fegley, S. (1995). Prosocial behavior and caring in adolescence: Relations to self-understanding and social judgment. *Child development*, *66*, 1346-1359. doi:10.1111/j.1467-8624.1995.tb00939.x

Hart, D., Field, N. P., Garfinkle, J. R., & Singer, J. L. (1997). Representations of self and other: A semantic space model. *Journal of Personality*, *65*, 77-105.doi:10.1111/j.1467-6494.1997.tb00530.x

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, *76*, 408-420. doi:10.1080/03637750903310360

Heppen, J. B., & Ogilvie, D. M. (2003). Predicting affect from global self-discrepancies: The dual role of the undesired self. *Journal of Social and Clinical Psychology*, *22*, 347-368. doi:10.1521/jscp.22.4.347.22898

Hewitt, P. L., & Genest, M. (1990). The ideal self: Schematic processing of perfectionistic content in dysphoric university students. *Journal of Personality and Social Psychology*, *59*, 802-808. doi:10.1037/0022-3514.59.4.802

Higgins, E. T. (1989). Self-discrepancy theory: What patterns of self-beliefs cause people to suffer? *Advances in Experimental Social Psychology*, *22*, 93-136. doi:10.1016/S0065-2601(08)60306-8

Higgins, E. T., Klein, R., & Strauman, T. (1985). Self-concept discrepancy theory: A psychological model for distinguishing among different aspects of depression and anxiety. *Social Cognition*, *3*, 51-76. doi:10.1521/soco.1985.3.1.51

Honkonen, T., Ahola, K., Pertovaara, M., Isometsä, E., Kalimo, R., Nykyri, E., . . . & Lönnqvist, J. (2006). The association between burnout and physical illness in the general population results from the Finnish Health 2000 Study. *Journal of Psychosomatic Research*, *61*, 59-66. doi:10.1016/j.jpsychores.2005.10. 002

Keinan, G., & Malach-Pines, A. (2007). Stress and burnout among prison personnel: Sources, outcomes, and intervention strategies. *Criminal Justice and Behavior*, *34*, 380-398. doi:10.1177/0093854806290007

Kirchler, E., Palmonari, A., & Pombeni, M. L. (1994). Social categorization processes as dependent on status differences between groups: A step into adolescents' peer-groups. *European Journal of Social Psychology*, *24*, 541-563. doi:10.1002/ejsp.2420240503

Lambert, E. G., Barton-Bellessa, S. M., & Hogan, N. L. (2015). The consequences of emotional burnout among correctional staff. *Sage Open*, *5*, 1-15. doi:10.1177/2158244015590444

Lambert, E. G., Hogan, N. L., & Altheimer, I. (2010). An exploratory examination of the consequences of burnout in terms of life satisfaction, turnover intent, and absenteeism among private correctional staff. *The Prison Journal*, *90*, 94-114. doi:10.1177/0032885509357586

Lambert, E. G., Hogan, N. L., Dial, K. C., Jiang, S., & Khondaker, M. I. (2012). Is the job burning me out? An exploratory test of the job characteristics model on the emotional burnout of prison staff. *The Prison Journal*, *92*, 3-23. doi:10.1177/0032885511428794

Lambert, E. G., Hogan, N. L., & Jiang, S. (2010). A preliminary examination of the relationship between organisational structure and emotional burnout among correctional staff. *The Howard Journal of Criminal Justice*, *49*, 125-146. doi:10.1111/j.1468-2311.2010.00606.x

Leiter, M. P., & Maslach, C. (2009). Nurse turnover: The mediating role of burnout. *Journal of Nursing Management*, *17*, 331-339. doi:10.1111/j.1365-2834.2009.01004.x

Lindquist, C. A., & Whitehead, J. T. (1986). Burnout, job stress and job satisfaction among Southern correctional officers: Perceptions and causal factors. *Journal of Offender Counseling Services Rehabilitation*, *10*(4), 5-26.

Liu, X. M., & Qin, H. F. (2005). The relationship between professional self-concept and burnout in Chinese teachers. *Chinese Journal of Clinical Psychology*, *2*(3), 150-152.

Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist*, *41*(9), 954-969.

Markus, H., & Ruvolo, A. (1989). Possible selves: Personalized representations of goals. In L. A. Pervin (Ed.), *Goal concepts in personality and social psychology* (pp. 211-241). Hillsdale, NJ: Lawrence Erlbaum.

Maslach, C. (1981). Burnout: A social psychological analysis. In J. W. Jones (Ed.), *The burnout syndrome* (pp. 30-53). Park Ridge, IL: London House.

Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. *Journal of Organizational Behavior*, *2*, 99-113. doi:10.1002/job.4030020205

Maslach, C., Schaufeli, W. B., & Leiter, M. P. (2001). Job burnout. *Annual Review of Psychology*, *52*(1), 397-422.

Nagy, M. S. (2002). Using a single-item approach to measure facet job satisfaction. *Journal of Occupational and Organizational Psychology*, *75*, 77-86. doi:10.1146/annurev.psych.52.1.397

Neveu, J. P. (2007). Jailed resources: Conservation of resources theory as applied to burnout among prison guards. *Journal of Organizational Behavior*, *28*, 21-42. doi:10.1002/job.393

Ogilvie, D. M. (1987). The undesired self: A neglected variable in personality research. *Journal of Personality and Social Psychology*, *52*, 379-385. doi:10.1037/0022-3514.52.2.379

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1978). *The measurement of meaning*. Urbana, IL: University of Illinois Press.

Pines, A., & Aronson, E. (1988). *Career burnout: Causes and cures*. New York, NY: Free press.

Prati, G., & Boldrin, S. (2011). Fattori di stress e benessere organizzativo negli operatori di polizia penitenziaria [Organizational stressors, exposure to critical incidents and organizational well-being among correctional officers]. *Giornale Italiano di Medicina del Lavoro ed Ergonomia*, *33*, 33-39.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 717-731.

Purvanova, R. K., & Muros, J. P. (2010). Gender differences in burnout: A meta-analysis. *Journal of Vocational Behavior*, *77*, 168-185. doi:10.1016/j.jvb.2010.04.006

Reich, W. A., & Rosenberg, S. (2004). Reflected self-image and commitment to a career role. *Self and Identity*, *3*, 115-123. doi:10.1080/13576500342000086

Rogers, C. R. (1961). The characteristics of a helping relationship. In C. R. Rogers (Ed.), *On becoming a person: A therapist's view of psychotherapy* (pp. 39-58). Boston, MA: Houghton Mifflin.

Ruvolo, A. P., & Markus, H. R. (1992). Possible selves and performance: The power of self-relevant imagery. *Social Cognition*, *10*, 95-124. doi:10.1521/soco.1992.10.1.95

Santinello, M. (2007). *LBQ. Link Burnout Questionnaire*. Firenze, Italy: Giunti O. S.

Schaufeli, W. B., & Peeters, M. C. (2000). Job stress and burnout among correctional officers: A literature review. *International Journal of Stress Management*, *7*(1), 19-48.

Schulz, M., Damkröger, A., Voltmer, E., Löwe, B., Driessen, M., Ward, M., & Wingenfeld, K. (2011). Work-related behaviour and experience pattern in nurses: Impact on physical and mental health. *Journal of Psychiatric and Mental Health Nursing*, *18*, 411-417. doi:10.1111/j.1365-2850.2011.01691.x

Shanafelt, T. D., Balch, C. M., Bechamps, G. J., Russell, T., Dyrbye, L., Satele, D., . . . Freischlag, J. A. (2009). Burnout and career satisfaction among American surgeons. *Annals of Surgery*, *250*, 463-471. doi:10.1097/SLA.0b013e3181ac4dfd

Sheldon, K. M., Ryan, R. M., Rawsthorne, L. J., & Ilardi, B. (1997). Trait self and true self: Cross-role variation in the Big-Five personality traits and its relations with psychological authenticity and subjective well-being. *Journal of Personality and Social Psychology*, *73*, 1380-1393. doi:10.1037/0022-3514.73.6.1380

Strauman, T. J., & Higgins, E. T. (1988). Self-discrepancies as predictors of vulnerability to distinct syndromes of chronic emotional distress. *Journal of Personality*, *56*, 685-707. doi:10.1111/j.1467-6494.1988.tb00472.x

Soler, J. K., Yaman, H., Esteva, M., Dobbs, F., Asenova, R. S., Katic, M., . . . Kotányi, P. (2008). Burnout in European family doctors: The EGPRN study. *Family Practice*, *25*, 245-265. doi:10.1093/fampra/cmn038

Takase, M., Maude, P., & Manias, E. (2006). The impact of role discrepancy on nurses' intention to quit their jobs. *Journal of Clinical Nursing*, *15*, 1071-1080. doi:10.1111/j.1365-2702.2005.01508.x

Tewksbury, R., & Higgins, G. E. (2006). Prison staff and work stress: The role of organizational and emotional influences. *American Journal of Criminal Justice*, *30*(2), 247-266.

Thomsen, S., Soares, J., Nolan, P., Dallender, J., & Arnetz, B. (1999). Feelings of professional fulfilment and exhaustion in mental health personnel: The importance of organizational and individual factors. *Psychotherapy and Psychosomatics*, *68*, 157-164. doi:10.1159/000012325

Turner, R. G. (1978). Consistency, self-consciousness, and the predictive validity of typical and maximal personality measures. *Journal of Research in Personality*, *12*, 117-132. doi:10.1016/0092-6566(78)90088-0

VanYperen, N. W. (1998). Informational support, equity and burnout: The moderating effect of self-efficacy. *Journal of Occupational and Organizational Psychology*, *71*, 29-33. doi:10.1111/j.2044-8325.1998.tb00660.x

Villa, A., & Calvete, E. (2001). Development of the teacher self-concept evaluation scale and its relation to burnout. *Studies in Educational Evaluation*, *27*, 239-255. doi:10.1016/S0191-491X(01)00028-1

Wanous, J. P., Reichers, A. E., & Hudy, M. J. (1997). Overall job satisfaction: How good are single-item measures? *Journal of Applied Psychology*, *82*, 247-252. doi:10.1037/0021-9010.82.2.247

Walters, S. (1996). The determinants of job satisfaction among Canadian and American correctional officers. *Journal of Crime and Justice*, *19*, 145-158. doi:10.1080/0735648X.1996.9721551

West, C. P., Huschka, M. M., Novotny, P. J., Sloan, J. A., Kolars, J. C., Habermann, T. M., & Shanafelt, T. D. (2006). Association of perceived medical errors with resident distress and empathy: A prospective longitudinal study. *JAMA: The Journal of the American Medical Association*, *296*, 1071-1078. doi:10.1001/jama.296.9.1071

Wolpin, J., Burke, R. J., & Greenglass, E. R. (1991). Is job satisfaction an antecedent or a consequence of psychological burnout? *Human Relations*, *44*, 193-209. doi:10.1177/001872679104400205

Ybema, J. F., Smulders, P. G., & Bongers, P. M. (2010). Antecedents and consequences of employee absenteeism: A longitudinal perspective on the role of job satisfaction and burnout. *European Journal of Work and Organizational Psychology*, *19*, 102-124. doi:10.1080/13594320902793691

TPM®

WHY IS PERFECTIONISM A RISK FACTOR FOR WORKAHOLISM? THE MEDIATING ROLE OF IRRATIONAL BELIEFS AT WORK

ALESSANDRA FALCO
LAURA DAL CORSO
DAMIANO GIRARDI
UNIVERSITY OF PADOVA

ALESSANDRO DE CARLO
LUMSA UNIVERSITY OF ROMA

BARBARA BARBIERI
UNIVERSITY OF CAGLIARI

TANIA BOATTO
XYLEM INCORPORATION

WILMAR B. SCHAUFELI
UTRECHT UNIVERSITY
LEUVEN UNIVERSITY

This study investigates the role of irrational beliefs at work in two samples of workers. The first aim was to evaluate the psychometric properties of an Italian adaptation of the Work-related Irrational Beliefs Questionnaire (WIB-Q; Van Wijhe, Peeters, & Schaufeli, 2013). Several confirmatory factor analyses (CFAs), including multiple-group CFAs, supported the four-factor structure (i.e., performance demands, coworkers' approval, failure, and control) of the WIB-Q in both samples. Additionally, the WIB-Q showed satisfactory convergent, discriminant, and criterion-related validity. The second aim of this study was to test a theoretical model in which irrational beliefs at work mediate the association between two dimensions of perfectionism — self-oriented perfectionism (SOP) and socially prescribed perfectionism (SPP) — and workaholism. Failure mediated the association between SOP/SPP and workaholism, whereas the mediating effect of performance demands was marginally significant. Overall, the results of this study suggest that interventions aimed at preventing workaholism should target perfectionists' work-related irrational beliefs related to failure and performance demands.

Key words: Work-related irrational beliefs; Self-oriented perfectionism; Socially prescribed perfectionism; Workaholism; Mediating effect.

Correspondence concerning this article should be addressed to Alessandra Falco, Department FISPPA – Section of Applied Psychology, University of Padova, Via Venezia 14, 35131 Padova (PD), Italy. Email: alessandra.falco@unipd.it

Workaholism may be defined as "the tendency to work excessively hard in a compulsive way" (Schaufeli, Taris, & Bakker, 2008, p. 204). Accordingly, the authors identified two core dimensions of the construct, that is, working excessively (i.e., working beyond what is expected to meet organizational or economic requirements) and working compulsively (i.e., thinking persistently and frequently about work). The former represents the behavioral, whereas the latter represents the cognitive component of workaholism. Workaholism is characterized by the combination of high levels of both working excessively (WE) and working compulsively (WC) (Schaufeli, Bakker, van der Heijden, & Prins, 2009).

Recent studies suggested that workaholism and work engagement (i.e., "a positive, fulfilling, work-related state of mind that is characterized by vigor, dedication, and absorption"; Schaufeli, Salanova, González-Romá, & Bakker, 2002, p.74) should be considered as different types of heavy work investment (Schaufeli, 2016; Shimazu, Schaufeli, Kamiyama, & Kawakami, 2015). Indeed, although workers with high scores on work engagement or workaholism dedicate a lot of time and energy to their work, the former are basically intrinsically motivated, whereas the latter are fueled by extrinsic motivation (Van Beek, Hu, Schaufeli, Taris, & Schreurs, 2012). Furthermore, work engagement is associated with positive outcomes (e.g., life satisfaction, job performance; Barbieri, Dal Corso, Di Sipio, De Carlo, & Benevene, 2016; Shimazu et al., 2015), whereas workaholism is predominantly associated with negative outcomes, such as physical and psychological symptoms, sickness absenteeism and presenteeism, cardiovascular risk, and sleep problems (Falco et al., 2013; Girardi, Falco, Piccirelli, et al., 2015; Kubota et al., 2010; Salanova et al., 2016; for a recent review see also Andreassen, 2014).

Previous studies suggested that several factors, including personal and situational variables, might lead to the onset of workaholism (Liang & Chu, 2009; McMillan & O'Driscoll, 2008; Ng, Sorensen, & Feldman, 2007; Spurk, Hirschi, & Kauffeld, 2016), similarly to what has been described for work-related stress and burnout (Bélanger et al., 2016; Girardi, Falco, De Carlo, et al., 2015). Among these, perfectionism has received considerable attention (Clark, Michel, Zhdanova, Pui, & Baltes, 2016; Spence & Robbins, 1992; see also Stoeber & Damian, 2016, for a review). Perfectionism may be defined as striving for exceedingly high, often unrealistic standards of performance, accompanied by frequent thoughts about the accomplishment of these standards and excessively critical evaluation of one's own behavior (Flett & Hewitt, 2002; Frost, Marten, Lahart, & Rosenblate, 1990; Sirois & Molnar, 2016). Several authors conceptualize perfectionism as a multidimensional construct, although there is no consensus about the central features of the construct (Frost et al., 1990; Hewitt & Flett, 1991; Slaney, Rice, Mobley, Trippi, & Ashby, 2001). According to the influential model proposed by Hewitt and Flett (1991), the one adopted in this study, perfectionism encompasses interpersonal as well as intrapersonal aspects and comprises three dimensions, namely self-oriented perfectionism (SOP; i.e., setting extremely high standards for oneself), socially prescribed perfectionism (SPP; i.e., the attainment of unrealistically high standards imposed by significant others), and other-oriented perfectionism (OOP; i.e., setting excessively high and often unrealistic standards for other people).

Moreover, previous studies have shown that dimensions of perfectionism taken from different theoretical models reflect two underlying factors, namely perfectionistic strivings and perfectionistic concerns (Bieling, Israeli, & Antony, 2004; Frost, Heimberg, Holt, Mattia, & Neubauer, 1993). Perfectionistic strivings (PS) subsume the tendency to set unrealistically high personal standards and to expect nothing less than perfection from oneself. Perfectionistic con-

cerns (PC) capture aspects of perfectionism related to concerns over making mistakes, excessive preoccupation about negative evaluation by others, and an exceptionally critical appraisal of one's own behavior. Indicators of perfectionistic strivings include, among others, SOP, whereas SPP reflects perfectionistic concerns (Sirois & Molnar, 2016; Stoeber & Damian, 2016; Stoeber & Otto, 2006). Interestingly, indicators of PS are typically associated with adaptive characteristics and outcomes (e.g., conscientiousness, problem-focused coping, well-being, and satisfaction with life), whereas facets reflecting PC are often related to maladaptive characteristics and outcomes, such as neuroticism, avoidant coping, and reduced well-being (Bieling et al., 2004; Cox, Enns, & Clara, 2002; Gnilka, Ashby, & Noble, 2012).

Altogether, several empirical studies showed that perfectionism is positively associated with workaholism. Indeed, in a recent meta-analysis Clark et al. (2016) found a positive, strong correlation between perfectionism and workaholism. However, previous studies have at least two limitations. First, past research usually examined the association between overall perfectionism and workaholism, and did not consider possible differences between perfectionistic strivings and concerns, albeit with some exceptions. In this regard, some previous studies showed that both perfectionistic strivings and concerns are positively associated with workaholism, although results for PC were somewhat inconsistent across studies (Clark, Lelchook, & Taylor, 2010; Falco, Piccirelli, Girardi, Di Sipio, & De Carlo, 2014; Stoeber, Davis, & Townley, 2013). Second, and perhaps most importantly, mechanisms that could explain the association between perfectionism and workaholism were not considered (for a recent review see Stoeber & Damian, 2016).

Overview of the Study

In this perspective, cognitive elements of trait perfectionism, such as perfectionistic cognitions and irrational beliefs (Flett, Hewitt, & Cheng, 2008; Flett, Nepon, & Hewitt, 2016) could act as possible mediators. In this study we focused on irrational beliefs, that is, illogical and rigid cognitions that are related to unrealistic demands about the self, other people, and the world in general, and that may lead to maladaptive consequences for the individual (Ellis, David, & Lynn, 2010). Previous studies showed that individuals with high levels of perfectionism have the tendency to endorse several irrational beliefs that reflect awfulizing, catastrophizing, difficulties in tolerating frustration, and the idea that self-worth depends on achievement and the approval by others (Flett & Hewitt, 2008), such as high self-expectations, demand for approval, and anxious overconcern (Flett, Hewitt, Blankstein, & Koledin, 1991; Flett et al., 2008; Watson, Simmons, Weathington, O'Leary, & Culhane, 2009). Moreover, irrational beliefs (e.g., "I must respect the deadline at all costs, or a disaster will happen," "If I delegate my work, it won't get done properly"), may play a central role in the development of workaholism (Burwell & Chen, 2002; Chen, 2006; Van Wijhe, Schaufeli, & Peeters, 2010), whose central element, according to Naughton (1987), is an irrational commitment to excessive work.

Tellingly, Van Wijhe, Peeters, and Schaufeli (2013) developed the first questionnaire that assess irrational beliefs in the work context, namely the Work-related Irrational Beliefs Questionnaire (WIB-Q). The instrument measures four different kinds of irrational beliefs regarding the work context, that is, performance demands, coworkers' approval, failure, and control. The WIB-Q focuses exclusively on the cognitive aspects of these beliefs (and not on the emotional

aspects; Terjesen, Salhany, & Sciutto, 2009), which are considered important for workaholism. Moreover, Van Wijhe et al. (2013) found that workaholism was positively correlated with each of the four work-related irrational beliefs, and that performance demands and failure were positively associated with workaholism in a structural regression model using latent variables. It appeared that performance demands were still positively associated with workaholism, after controlling for the effect of negative affect, whereas failure was not.

Overall, the aim of this study is twofold. Because to the best of the authors' knowledge there is no Italian validation of the WIB-Q (Van Wijhe et al., 2013), the first objective of this study is to evaluate the psychometric properties of the Italian adaptation of the WIB-Q. In this regard, the dimensionality, construct validity (convergent and discriminant validity), and measurement invariance of the WIB-Q were examined through confirmatory factor analyses. Moreover, to assess the criterion-related validity of the WIB-Q, the concurrent correlations between the four irrational beliefs at work and several theoretically related constructs were examined. These constructs, indicated in the literature as possible antecedents and consequences of irrational beliefs, were perfectionism (both SOP and SPP; Flett et al., 1991; Flett et al., 2008), negative affectivity (Davies, 2006; Popov, Majstorović, Matanović, Jelić, & Raković, 2016), anxiety and depressive symptoms (Chang & D'Zurilla, 1996; Ciarrochi, 2004; Nieuwenhuijsen, Verbeek, de Boer, Blonk, & van Dijk, 2010), and burnout (Balevre, Cassells, & Buzaianu, 2012; Ogai & Okayasu, 2010). The second aim consists of testing a theoretical model in which perfectionism (i.e., SOP and SPP) is positively associated with irrational beliefs at work (i.e., performance demands, coworkers' approval, failure, and control), which, in their turn, are positively associated with workaholism. Accordingly, we expect that irrational beliefs at work mediate the association between perfectionism and workaholism. It should be emphasized that, to the best of our knowledge, this is the first study to investigate the mediating role of irrational beliefs at work in the relationship between perfectionism and workaholism, although other studies considered other possible mediators such as work motivation (Stoeber et al., 2013). Finally, in this research we focused solely on SOP and SPP since other-oriented perfectionism neither reflects perfectionistic strivings nor perfectionistic concerns (Stoeber & Otto, 2006).

METHOD

Participants and Procedure

The current study examined two samples: (1) workers from different organizations (S1; N = 506) and (2) workers from a private company in the metal engineering sector (S2; N = 264). Participants from S1 were approached by trained research assistants and invited to complete an anonymous questionnaire (paper-and-pencil) about their work experience. This sample consisted of 289 women (57.1%) and 216 men (42.7%; one gender missing, 0.2%). The majority of the respondents were younger than 40 years (38.7%), 31.8% were older than 50 years, and 27.9% were aged between 40 and 50 years (eight missing data, 1.6%). Most participants worked in the private service sector (47%), followed by industry (15.2%), education (8.3%), healthcare (8.1%), and the public sector (7.3%) whereas 13.3% of the respondents worked in other sectors (four missing data, 0.8%). The majority of the respondents (77.5%) had a permanent contract (17 missing data,

3.4%) and 77.1% were employed full-time (three missing data, 0.6%). With respect to work experience, 37% had been with their current company for five to 19 years, 31.2% for more than 19 years, and 28.4% for less than five years (17 missing data, 3.4%).

With respect to S2, workers were administered a standardized questionnaire (paper-and-pencil) as part of a work-related stress risk assessment. This sample consisted of 168 men (63.6%) and 82 women (31.1%; 14 missing data, 5.3%). The majority of the respondents were aged between 40 and 50 years (45.1%), 27.7% were younger than 40 years, and 22.7% were older than 50 years (12 missing data, 4.5%). Most of the participants were blue-collar workers (50.4%), followed by white-collar workers (34.8%), and managers (11%; 10 missing data, 3.8%). With respect to work experience, 51.6% had been with the company for five to 19 years, 36.7% for more than 19 years, and 6.8% for less than five years (13 missing data, 4.9%). For both S1 and S2, the questionnaire was administered anonymously, and participants took part in the study on a voluntary basis.

Measures

To assess the constructs under investigation, the following self-report measures were used.

Irrational beliefs at work were assessed in both S1 and S2 using the Work-related Irrational Beliefs Questionnaire (WIB-Q; Van Wijhe et al., 2013). The original scale items were translated into Italian by the authors. Subsequently, an English native-speaker translator performed back-translation, to avoid discrepancies between the English and Italian version of the WIB-Q. The scale is composed of 16 items and measures four types of work-related irrational beliefs (four item each), namely performance demands (Cronbach's alpha was .81 in S1, and .74 in S2), coworkers' approval (α was .87 in S1, and .84 in S2), failure (α was .83 in both S1 and S2), and control (α was .86 in S1, and .87 in S2). The response scale ranged from 1 (*completely disagree*) to 5 (*completely agree*).

Perfectionism was assessed in both S1 and S2 using an Italian adaptation (Falco et al., 2014) of a short version of the Multidimensional Perfectionism Scale (HMPS; Hewitt & Flett, 1991). The scale is composed of seven items and measures self-oriented perfectionism (SOP, three items; α was .81 in S1, and .85 in S2) and socially prescribed perfectionism (SPP, four items; α was .82 in S1, and .73 in S2), which reflect perfectionistic strivings and concerns, respectively. The response scale ranged from 1 (*strongly disagree*) to 6 (*strongly agree*).

Workaholism was assessed in S1 using the Dutch Workaholism Scale (DUWAS; Schaufeli et al., 2008) in the Italian adaptation (Falco et al., 2012; Kravina, Falco, Girardi, & De Carlo, 2010; see also Balducci, Avanzi, Consiglio, Fraccaroli, & Schaufeli, 2015; Mazzetti, Schaufeli, & Guglielmi, 2016). The scale is composed of 10 items, designed to detect the two dimensions of WE (six items; α was .80) and WC (four items; α was .87). The 6-point response scale ranged from 1 (*strongly disagree*) to 6 (*strongly agree*). Since workaholism reflects tendency to work excessively hard in a compulsive way (Schaufeli et al., 2009), an overall workaholism score was used. Cronbach's alpha for the overall scale was .86.

Anxiety and *depressive symptoms* were assessed in S2 using two scales taken from the Qu-Bo test, a standardized instrument developed for the Italian context (De Carlo, Falco, & Capozza, 2008). The psychometric properties of the scales taken from the Qu-Bo test are de-

scribed in Trifiletti, Vianello, and Capozza (2013). This scale is composed of seven items, designed to detect anxiety (three items; α was .80) and depressive symptoms (four items; α was .68). The scales assessed how often specific anxiety or depressive symptoms occurred in the past six months, and the response scale ranged from 1 (*never*) to 6 (*every day*).

Burnout was determined in S2 using the scale taken from the Qu-Bo test (De Carlo et al., 2008). The nine-item scale includes three subdimensions, measured by three items each: exhaustion (α was .85), cynicism (α was .88), and reduced sense of personal accomplishment (α was .84). This scale has been developed and tested in the Italian context and support has been found for reliability, validity, and the factor structure. Answers were provided on a 6-point scale ranging from 1 (*strongly disagree*) to 6 (*strongly agree*).

Negative affectivity was assessed in S2 using a scale taken from the Qu-Bo test (De Carlo et al., 2008). The scale is composed of four items with a response scale ranging from 1 (*strongly disagree*) to 6 (*strongly agree*). Cronbach's alpha was .80.

The psychometric properties of the scales used in the present study were evaluated through several confirmatory factor analyses (CFAs), whose results are described in the Results section of this paper.

Data Analysis

The psychometric properties of the WIB-Q were evaluated in terms of factor structure, construct validity, criterion-related validity, and measurement invariance across two different samples of workers (MacKenzie, Podsakoff, & Podsakoff, 2011; Vandenberg & Lance, 2000). Firstly, dimensionality and construct validity, in terms of convergent and discriminant validity, were examined in both S1 and S2 through CFA using LISREL 8.80 (Jöreskog & Sörbom, 2006). Additionally, for each dimension of the WIB-Q the coefficient average variance extracted (AVE) was calculated, which represents the average amount of variation that a latent construct explains in the observed variables, to which it is theoretically related (Fornell & Larcker, 1981). AVE can be used to assess both convergent and discriminant validity. A good convergent validity is verified when all indicators load significantly on their respective latent construct, and AVE scores equal to or higher than .50 for each dimension indicate a good convergent validity. In addition, two dimensions can be considered distinct (i.e., discriminant validity) if the AVE of each of them is higher than the squared correlation between the two dimensions (i.e., shared variance).

Additionally, the measurement invariance across both samples (i.e., S1 and S2) was examined through a multiple-group CFA approach (Vandenberg & Lance, 2000; see also Barbaranelli, 2013; Brown, 2015). More specifically, several increasingly constrained models were tested in a sequential way (i.e., stepwise) to assess different levels of measurement invariance, that is, configural invariance, metric invariance, and invariance of factor variances and covariances (Vandenberg & Lance, 2000).

To evaluate the goodness-of-fit of the CFA models, the χ^2 test was used. A model shows a good fit to data if χ^2 is nonsignificant. However, since the χ^2 is affected by sample size, three additional fit indices were used: the root mean square error of approximation (RMSEA), the comparative fit index (CFI), and the standardized root mean squared residual (SRMR). More specifically, values close to or smaller than .08 for RMSEA and SRMR and values close to or

greater than .90 for CFI indicate an acceptable model fit, whereas values close to .06 and .95 for RMSEA and CFI, respectively, indicate good fit (Bentler, 1990; Brown, 2015; Browne & Cudeck, 1993; Hu & Bentler, 1999). Furthermore, the chi-square difference test ($\Delta\chi^2$) was adopted to assess the tenability of equality constraints in multiple-group CFAs, because a model with constraints is nested in the model without constraints. Accordingly, if the chi-square difference is nonsignificant, the more parsimonious model (i.e., the one with constraints) should be preferred over the less parsimonious one (i.e., the one without constraints). Moreover, as reported above, to assess the criterion-related validity of the WIB-Q, the concurrent correlations between the four irrational beliefs at work and the theoretically related constructs (i.e., perfectionism, negative affectivity, anxiety and depressive symptoms, and burnout) were examined in S2.

Finally, to test the hypothesized relationships between perfectionism, irrational beliefs at work, and workaholism, a structural equation model with observed variables (i.e., path analysis) was estimated in S1 using LISREL 8.80 (Jöreskog & Sörbom, 2006). The structural paths were freely estimated, to test both direct and indirect effects simultaneously (just-identified path models; Kline, 2011). To test the significance of the indirect effect of perfectionism on workaholism through irrational beliefs at work (i.e., mediation), we computed asymmetric confidence intervals for the indirect effect based on the distribution of product method using the RMediation package (Tofighi & MacKinnon, 2011). If a confidence interval does not contain zero, then a statistically significant mediation is supported (MacKinnon, Cheong, & Pirlott, 2012).

Finally, missing values were considered. For CFAs, participants with missing values on any of the items of the WIB-Q were removed from the dataset (i.e., listwise deletion). The final samples for CFAs comprised, therefore, 440 workers for S1 (180 missing values, 2.2%) and 223 workers for S2 (165 missing values, 3.9%). With respect to criterion-related validity of WIB-Q (S2) and path analysis (S1) missing values were estimated using the person-mean substitution approach, a technique designed for handling missing data when composite scores are used (Downey & King, 1998). More specifically, participants with more than 50% of missing items on a given scale were excluded from subsequent analyses (Hawthorne & Elliott, 2005). Next, missing values within a given scale were replaced by the mean of each individual's completed items in that scale (person-mean imputation; Bono, Ried, Kimberlin, & Vogel, 2007; Downey & King, 1998). Overall, 119 missing values were imputed for S1 ($N = 474$, 0.8%), whereas 95 were imputed for S2 ($N = 228$, 1%).

RESULTS

The first aim of this study was to evaluate the psychometric properties of an Italian adaptation of the WIB-Q. Therefore, two CFAs were carried out, to test the original four-factor model (16 items) proposed by Van Wijhe et al. (2013) in both S1 and S2. The fit indices showed an acceptable fit to data for both S1 — $\chi^2(98) = 347.47$, $p < .01$; RMSEA = .076; CFI = .935; SRMR = .068 — and S2 — $\chi^2(98) = 218.31$, $p < .01$; RMSEA = .074; CFI = .929; SRMR = .081. However, Item 16 showed a low standardized factor loading in both samples. Moreover, an inspection of the modification indices revealed substantial cross loadings for Items 1 and 13 in both S1 and S2. Accordingly, these three items were removed, and a new CFA was carried out. The fit indices of the remaining 13 items showed a good fit to data for both S1 — $\chi^2(59) = 162.98$, $p < .01$;

RMSEA = .063; CFI = .966; SRMR = .046 — and S2 — $\chi^2(59) = 123.12$, $p < .01$; RMSEA = .070; CFI = .949; SRMR = .059. The AVE was greater than .50 for each dimension of the WIB-Q in both S1 and S2, and equal to .50 for performance demands in S2. Moreover, the AVE for each subscale was higher than the shared variance between each couple of latent factors. Hence, overall, the WIB-Q scale showed satisfactory convergent and discriminant validity.

Next, the measurement invariance of the WIB-Q across both samples was examined through a multiple-group CFA approach. First, configural invariance was tested, and the model showed a good fit to data, $\chi^2(118) = 286.10$, $p < .01$; RMSEA = .066; CFI = .974. Accordingly, configural invariance was supported. In the second step, factor loadings were constrained to be equal across groups. This model showed a good fit to data, $\chi^2(127) = 296.13$, $p < .01$; RMSEA = .063; CFI = .974. Additionally, the fit of this model was not significantly worse than the fit of the less constrained model (i.e., the configural invariance model), $\Delta\chi^2(9) = 10.03$, $p = .35$, and therefore metric invariance was supported. Finally, factor variances and covariances were constrained to be equal across groups. This model also showed a good fit to data, $\chi^2(137) = 316.03$, $p < .01$; RMSEA = .062; CFI = .973, but the fit of this model was worse than the fit of the previous one, $\Delta\chi^2(10) = 19.90$, $p = .03$. An inspection of the modification indices showed that the covariance between control and failure should be freely estimated (i.e., partial measurement invariance; Byrne, Shavelson, & Muthén, 1989). Accordingly, a new CFA was carried out, and the fit indices showed a good fit to data, $\chi^2(136) = 311.44$, $p < .01$; RMSEA = .062; CFI = .973. Moreover, the fit of this re-specified model was not significantly worse than the fit of the less constrained model (i.e., metric invariance), $\Delta\chi^2(9) = 15.31$, $p = .08$. Therefore, factor loading, variances, and covariances were invariant across S1 and S2, except for the covariance between control and failure, which was larger in S1 ($r = .56$) than in S2 ($r = .42$). The common metric completely standardized solution is summarized in Table 1.

To investigate the criterion-related validity of the WIB-Q, the correlations between the four dimensions of irrational beliefs at work and several theoretically related constructs (i.e., perfectionism, negative affectivity, anxiety and depressive symptoms, and burnout) were examined in S2. Prior to examining these correlations, a CFA was carried out to investigate the psychometric properties of the scales adopted for this purpose (except for the WIB-Q, whose psychometric properties in S2 are described above). The hypothesized model included 27 items and eight latent factors, namely SOP, SPP, negative affectivity, anxiety symptoms, depressive symptoms, and the three dimensions of burnout. Because several scale items (e.g., depressive symptoms, burnout) were not normally distributed, the robust maximum likelihood was adopted as the estimation method. Therefore, to assess model fit, the scaled Satorra-Bentler chi-square test ($SB\chi^2$) was used. The model showed a good fit to data, $SB\chi^2(296) = 437.61$, $p < .01$; RMSEA = .046; CFI = .968; SRMR = .073. Moreover, all items loaded substantially on their respective factors (median standardized factor loading of .80), and correlations between latent factors ranged from .0 to .80 (between emotional exhaustion and depressive symptoms).

Next, the correlations between the WIB-Q and the other constructs in the nomological network (i.e., perfectionism, negative affectivity, anxiety and depressive symptoms, and burnout) were analyzed, and results are reported in Table 2. Overall, the two dimensions of failure and control were positively associated with anxiety and depressive symptoms, burnout (except for cynicism), and negative affectivity. Moreover, failure was positively associated with both the dimensions of

TABLE 1
Factor loading and correlations from the multiple-group confirmatory factor analysis of the WIB-Q: The common metric completely standardized solution

Item	Factor			
	Performance demands	Coworkers' approval	Failure	Control
Item 3	.71			
Item 4	.70			
Item 5	.84			
Item 6		.79		
Item 7		.65		
Item 9		.92		
Item 10		.82		
Item 11			.75	
Item 12			.84	
Item 14			.78	
Item 17				.69
Item 19				.88
Item 20				.89
	1	2	3	4
1. Performance demands	1			
2. Coworkers' approval	.37	1		
3. Failure	.43	.42	1	
4. Control	.09	.33	.56 (S1[a]) .42 (S2[b])	1

Note. [a]Sample 1, $N = 440$. [b]Sample 2, $N = 223$.

perfectionism, whereas control was positively associated with SPP (but not SOP). However, performance demands and coworkers' approval showed a somewhat different pattern of correlations. Indeed, performance demands were positively associated with perfectionism (both SOP and SPP) and negative affectivity, whereas coworkers' approval was positively associated with SPP (but not SOP), negative affectivity, and reduced sense of personal accomplishment. Overall, the WIB-Q showed reasonable criterion-related validity.

Finally, to test the hypothesized relationships between perfectionism, irrational beliefs at work, and workaholism, a path analysis model was estimated in S1. The results of this path analysis are represented in Figure 1. Prior to examining these associations, a CFA was carried out to investigate the psychometric properties of the scales adopted for this purpose, namely perfectionism and workaholism (except for the WIB-Q, whose psychometric properties in S1 are described above). Accordingly, the hypothesized model included 17 items and four latent factors, namely WE, WC, SOP, and SPP. The model showed an acceptable fit to data, $\chi^2(113) = 359.05$, $p < .01$;

TABLE 2
Criterion-related validity of the WIB-Q: Correlations between study variables

	Irrational beliefs at work			
	Performance demands	Coworkers' approval	Failure	Control
Anxiety symptoms	.09	.01	.25***	.18**
Depressive symptoms	.11	–.04	.22**	.25***
Exhaustion	.09	.08	.24***	.19**
Cynicism	.13	.02	.10	.07
Reduced sense of personal accomplishment	.00	.25***	.29***	.32***
Self-oriented perfectionism	.53***	.08	.18**	–.11
Socially prescribed perfectionism	.48***	.21**	.40***	.17**
Negative affectivity	.18**	.22**	.45***	.37***

Note. Sample 2, $N = 228$. ** $p < .01$. *** $p < .001$.

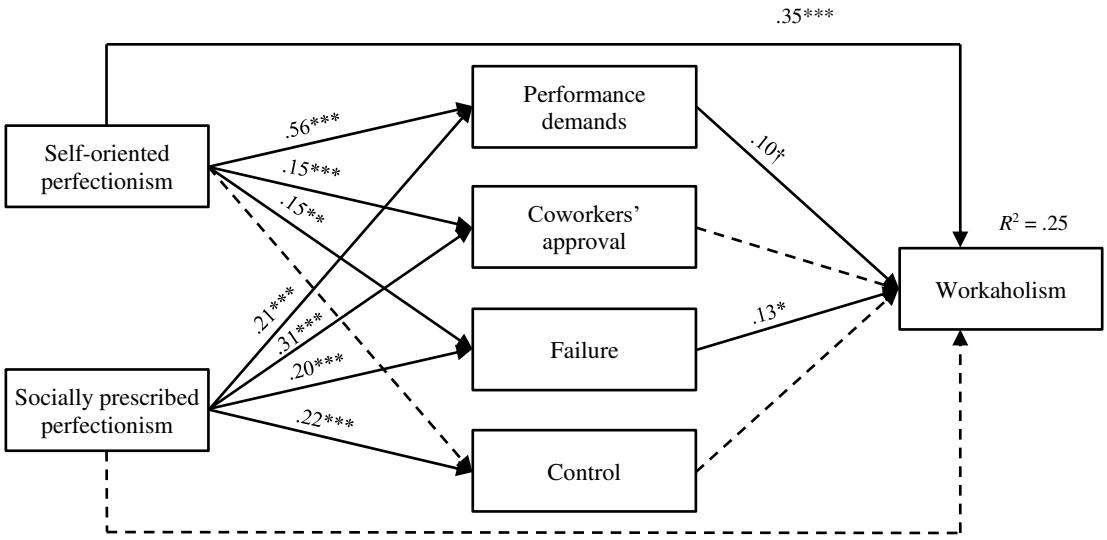

FIGURE 1
The path analysis model in Sample 1 ($N = 474$).
Dashed lines represent nonsignificant paths.
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

RMSEA = .068; CFI = .931; SRMR = .065. Moreover, all items loaded substantially on the respective factor (median standardized factor loading of .74), and correlations between latent factors ranged from .22 to .60. The correlations between workaholism, irrational beliefs at work, and perfectionism are summarized in Table 3.

In the path analysis model, SOP was positively associated with performance demands ($\gamma = .56$, $p < .001$), coworkers' approval ($\gamma = .15$, $p < .001$), and failure ($\gamma = .15$, $p < .01$). Additionally, SPP was positively associated with performance demands ($\gamma = .21$, $p < .001$), coworkers' approval

TABLE 3
Correlations between workaholism, irrational beliefs at work, and perfectionism

	1	2	3	4	5	6	7	8	9
1. Working excessively	1								
2. Working compulsively	.53***	1							
3. Workaholism	.91***	.84***	1						
4. Performance demands	.30***	.39***	.39***	1					
5. Coworkers' approval	.11*	.20***	.17***	.37***	1				
6. Failure	.18***	.24***	.23***	.36***	.42***	1			
7. Control	.06	.03	.05	.14**	.36***	.51***	1		
8. Self-oriented perfectionism	.34***	.49***	.46***	.65***	.28***	.23***	.02	1	
9. Socially prescribed perfectionism	.21***	.28***	.27***	.44***	.37***	.26***	.19***	.40***	1

Note. Sample 1, $N = 474$. * $p < .05$. ** $p < .01$. *** $p < .001$.

($\gamma = .31$, $p < .001$), failure ($\gamma = .20$, $p < .001$), and control ($\gamma = .22$, $p < .001$). However, only failure was positively associated with workaholism, controlling for the effect of both SOP and SPP, ($\beta = .13$, $p < .05$), whereas the association between performance demands and workaholism was positive but marginally significant ($\beta = .10$, $p < .08$). Interestingly, SOP was positively associated with workaholism, controlling for the effect of irrational beliefs at work ($\gamma = .35$, $p < .001$), whereas SPP was not.

The 95% asymmetric confidence intervals for the indirect effect of SOP/SPP on workaholism through failure did not contain zero. The unstandardized point estimate for SOP was .02, 95% CI [.002, .033], the same as unstandardized point estimate for SPP: .02, 95% CI [.004, .042]. Therefore, failure mediated the association between SOP/SPP and workaholism. Moreover, the 95% asymmetric confidence intervals for the indirect effect of SOP/SPP on workaholism through performance demands contained zero, thus suggesting a nonsignificant indirect effect. However, because the association between performance demands and workaholism was marginally significant, 90% asymmetric confidence intervals were also computed. The unstandardized point estimate for SOP was .04, 90% CI [.002, .085], whereas the unstandardized point estimate for SPP was .02, 90% CI [.001, .034]. Accordingly, the indirect effect of SOP/SPP on workaholism through performance demands was marginally significant.

Finally, to obtain a more parsimonious solution, an additional model was estimated, in which the four nonsignificant paths in the previous models were fixed to zero. This model showed a good fit to data: $\chi^2(4) = 5.27$, $p = .26$; RMSEA = .025; CFI = .999; SRMR = .021. In this final model, all structural paths were significant, including the association between performance demands and workaholism ($\gamma = .11$, $p < .05$).

Discussion

This study examined the role of irrational beliefs in the work context, and contributed to the understanding of the relationship between perfectionism and workaholism. First, we investigated the psychometric properties of an Italian adaptation of the Work-related Irrational Beliefs Questionnaire (WIB-Q; Van Wijhe et al., 2013). After removing three items that showed unsatisfactory characteristics (i.e., low factor loadings or cross-loadings), the hypothesized four-factor structure of the WIB-Q (i.e., performance demands, coworkers' approval, failure, and control) was partially invariant across two different samples of workers. More specifically, factor loading and factor variances/covariances were invariant, except for the covariance between control and failure, which was stronger in the first sample (i.e., a multi-occupational sample) than in the second one (i.e., a sample of workers from a private metal engineering company).

Moreover, the WIB-Q showed good construct (i.e., convergent and discriminant) as well as criterion-related validity. It should be noted that failure and control were concurrently and positively correlated with most of the constructs cited in the literature as possible antecedents and consequences of irrational beliefs, namely perfectionism (i.e., SOP and SPP), negative affectivity, anxiety/depressive symptoms, and burnout. Contrarily, performance demands and coworkers' approval were positively associated only with perfectionism (i.e., SOP/SPP) and negative affectivity, whereas the correlations with anxiety/depressive symptoms and burnout were not significant (except for the correlation between coworkers' approval and reduced sense of personal accomplishment). Basically, these results are in line with those of previous research. For example, specific irrational beliefs, namely self-directed shoulds and self-worth taken from the Survey of

Personal Beliefs (Demaria, Kassinove, & Dill, 1989), were not associated with anxiety or depressive symptoms in some studies (Chang & D'Zurilla, 1996; Culhane & Watson, 2003), whereas these associations were significant in other studies that included both clinical and nonclinical samples (Flett et al., 2008; Nottingham, 1992). Interestingly, on the one hand self-directed shoulds refer to inflexible demands directed toward the self, and share some conceptual similarity with performance demands from the WIB-Q. On the other hand, self-worth also reflects reduced self-ratings that result from evaluations by others, and is somewhat comparable to the coworkers' approval scale of the WIB-Q. That being said, we believe this adaptation of the WIB-Q to be a valid instrument to assess irrational beliefs at work in the Italian context. This may have important practical implications in terms of prevention of workaholism, as discussed below.

The second aim of this study was to test a theoretical model in which irrational beliefs at work mediate the association between self-oriented/socially prescribed perfectionism and workaholism. This mediating effect was supported only for failure, whereas the mediating effect for performance demands was marginally significant. We believe these findings to be particularly interesting, because previous research has shown that workaholics perform work activities for their instrumental value (i.e., extrinsic motivation). More specifically, they work hard to preserve and improve feelings of self-worth and self-esteem, and avoid negative emotions (Van Beek et al., 2012; Van Beek, Taris, Schaufeli, & Brenninkmeijer, 2013). In this perspective, individuals with high levels of perfectionism tend to endorse irrational beliefs that reflect the fear of failure and the pursuit of exceedingly high standards of performance. In turn, these irrational beliefs, which identify conditions that have to be met to avoid negative emotions and protect self-worth, could be a risk factor for workaholism.

Overall, we believe that the results of this study make several contributions to the literature. First, our findings showed that both SOP and SPP, which reflect perfectionistic strivings and concerns, are associated, directly or indirectly, with workaholism. Interestingly, a common limitation of several previous studies on workaholism is that they examined overall perfectionism (Clark et al., 2016), and did not consider possible differences between perfectionistic strivings and concerns (with some exceptions; see for example Falco et al., 2014; Stoeber et al., 2013; Taris, Van Beek, & Schaufeli, 2010), which are typically associated with adaptive or maladaptive characteristics and outcomes, respectively.

Furthermore, this study highlights the role of cognitive elements of trait perfectionism, namely work-related irrational beliefs (Flett et al., 2008) that may mediate the association between perfectionism and workaholism. This means that perfectionism is related to workaholism because of the irrational beliefs (particularly as related to failure and performance demands) that it produces, which, in their turn are associated with workaholism. Overall, the findings of this study are in line with previous research, which showed that both SOP and SPP are positively associated with irrational beliefs (Flett et al., 1991, 2008). Moreover, our results are rather consistent with the ones reported by Van Wijhe et al. (2013), who also found that failure and performance demands are positively associated with workaholism. However, to the best of our knowledge, this is the first study to show the mediating role of specific work-related irrational beliefs (i.e., failure and performance demands) in the relationship between dimensions of perfectionism reflecting perfectionistic strivings and concerns (i.e., SOP and SPP) and workaholism.

Moreover, an intriguing finding of this study was that self-oriented perfectionism was positively associated with workaholism, after controlling for the effect of irrational beliefs at

work, whereas socially prescribed perfectionism was not. In other words, work-related irrational beliefs fully mediated the association between SPP and workaholism, whereas the mediation for SOP was partial. Accordingly, although work-related irrational beliefs seem to play a central role in the relationship between SPP and workaholism, other mechanisms (besides work-related irrational beliefs) could be responsible for the association between SOP and workaholism. In this perspective, future studies could investigate possible additional mediators such as work motivation (Stoeber et al., 2013), coping styles (Gnilka et al., 2012), and perfectionistic automatic thoughts (Flett, Hewitt, Nepon, & Besser, 2017; Flett, Newby, Hewitt, & Persaud, 2011). Moreover, according to the definition of workaholism as a syndrome characterized by the tendency to work excessively in a compulsive way, an overall score of workaholism was adopted in this study. Hence, future research could replicate and extend the results of this study by modeling workaholism as a latent variable reflected by WE and WC.

Among the limitations of this study, it should be noted that the cross-sectional design precludes drawing causal inferences. A future longitudinal investigation would be useful to examine the direction of the associations between perfectionism, work-related irrational beliefs, and workaholism. Moreover, the observed relationships could be affected by common method bias (Podsakoff, MacKenzie, & Podsakoff, 2012), because the constructs were determined using the same measurement method (i.e., self-report questionnaires). Accordingly, future studies could use, for example, observer-rating of workaholism (Falco et al., 2012; see also Mazzetti et al., 2016).

Finally, we believe that this study has relevant practical implications for occupational psychologists and psychotherapists. Indeed, our results showed that perfectionism is, directly or indirectly (i.e., through work-related irrational beliefs), related to workaholism. However, perfectionism is a relatively stable trait, and perfectionists are relatively resistant to treatment (Flett & Hewitt, 2008). Therefore, interventions aimed at preventing workaholism should target the cognitive elements of trait perfectionism such as work-related irrational beliefs (as well as perfectionistic automatic thoughts; Flett et al., 2011), especially the ones related to failure and performance demands. More specifically, according to the rationale emotive behavior therapy (REBT) framework, workers should be encouraged to actively restructure their irrational beliefs (e.g., "I absolutely must perform well at work and obtain my supervisor's approval, or else I have little worth as a person") and to assimilate more functional rational beliefs (e.g., "I do not need to perform well at work, but I want it, and I will do my best to do so. However, if I perform badly and sometimes I do not get my supervisor's approval, I'm not worthless, but I'm just a person who acted poorly in that situation"; Ellis et al., 2010). Using this framework could be a relevant implication of this study, because, as we pointed out before, previous research has shown that workaholics perform work activities for their instrumental value (i.e., extrinsic motivation). More specifically, they work hard to preserve and improve feelings of self-worth and self-esteem, and to avoid negative emotions (Van Beek et al., 2012, 2013).

REFERENCES

Andreassen, C. S. (2014). Workaholism: An overview and current status of the research. *Journal of Behavioral Addictions*, *3*, 1-11. doi:10.1556/JBA.2.2013.017

Balducci, C., Avanzi, L., Consiglio, C., Fraccaroli, F., & Schaufeli, W. B. (2015). A cross-national study on the psychometric quality of the Italian version of the Dutch Work Addiction Scale (DUWAS). *European Journal of Psychological Assessment*. Advance online publication. doi:10.1027/1015-5759/a000300

Balevre, P. S., Cassells, J., & Buzaianu, E. (2012). Professional nursing burnout and irrational thinking: A replication study. *Journal for Nurses in Professional Development*, *28*, 2-8. doi:10.1097/NND.0b013 e318240a65a

Barbaranelli, C. (2013). I modelli di equazioni strutturali multi-gruppo e l'esame dell'invarianza fattoriale [Multi-group structural equation modeling and the examination of structural invariance]. In C. Barbaranelli & S. Ingoglia (Eds.), *I modelli di equazioni strutturali: Temi e prospettive* (pp. 153-200). Milano, Italy: LED.

Barbieri, B., Dal Corso, L., Di Sipio, A., De Carlo, A., & Benevene, P. (2016). Small opportunities are often the beginning of great enterprises: The role of work engagement in support of people through the recovery process and in preventing relapse in drug and alcohol abuse. *Work*, *55*, 373-383. doi:10.3233/ WOR-162411

Bélanger, J. J., Pierro, A., Barbieri, B., De Carlo, N. A., Falco, A., & Kruglanski, A. W. (2016). One size doesn't fit all: The influence of supervisors' power tactics and subordinates' need for cognitive closure on burnout and stress. *European Journal of Work and Organizational Psychology*, *25*, 287-300. doi:10.1080/1359432X.2015.1061999

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

Bieling, P. J., Israeli, A. L., & Antony, M. M. (2004). Is perfectionism good, bad, or both? Examining models of the perfectionism construct. *Personality and Individual Differences*, *36*, 1373-1385. doi:10.1016/S0191-8869(03)00235-6

Bono, C., Ried, L. D., Kimberlin, C., & Vogel, B. (2007). Missing data on the Center for Epidemiologic Studies Depression Scale: A comparison of 4 imputation techniques. *Research in Social and Administrative Pharmacy*, *3*, 1-27. doi:10.1016/j.sapharm.2006.04.001

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

Browne, M. W., & Cudek, R. (1993). Alternate ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Burwell, R., & Chen, C. P. (2002). Applying REBT to workaholic clients. *Counselling Psychology Quarterly*, *15*, 219-228. doi:10.1080/09515070210143507

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456-466. doi:10.1037/0033-2909.105.3.456

Chang, E. C., & D'Zurilla, T. J. (1996). Irrational beliefs as predictors of anxiety and depression in a college population. *Personality and Individual Differences*, *20*, 215-219. doi:10.1016/0191-8869(95)00166-2

Chen, C. (2006). Improving work-life balance: REBT for workaholic treatment. In R. J. Burke (Ed.), *Research companion to working time and work addiction* (pp. 310-330). Cheltenham, UK: Edward Elgar.

Ciarrochi, J. (2004). Relationships between dysfunctional beliefs and positive and negative indices of well-being: A critical evaluation of the Common Beliefs Survey-III. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, *22*, 171-188. doi:10.1023/B:JORE.0000047306.55720.4e

Clark, M. A., Lelchook, A. M., & Taylor, M. L. (2010). Beyond the Big Five: How narcissism, perfectionism, and dispositional affect relate to workaholism. *Personality and Individual Differences*, *48*, 786-791. doi:10.1016/j.paid.2010.01.013

Clark, M. A., Michel, J. S., Zhdanova, L., Pui, S. Y., & Baltes, B. B. (2016). All work and no play? A meta-analytic examination of the correlates and outcomes of workaholism. *Journal of Management*, *42*, 1836-1873. doi:10.1177/0149206314522301

Cox, B. J., Enns, M. W., & Clara, I. P. (2002). The multidimensional structure of perfectionism in clinically distressed and college student samples. *Psychological Assessment*, *14*, 365-373. doi:10.1037/1040-3590.14.3.365

Culhane, S. E., & Watson, P. J. (2003). Alexithymia, irrational beliefs, and the rational-emotive explanation of emotional disturbance. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, *21*, 57-72. doi:10.1023/A:1024133218634

Davies, M. F. (2006). Irrational beliefs and unconditional self-acceptance. I. Correlational evidence linking two key features of REBT. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, *24*, 113-124. doi:10.1007/s10942-006-0027-0

Demaria, T. P., Kassinove, H., & Dill, C. A. (1989). Psychometric properties of the Survey of Personal Beliefs: A rational-emotive measure of irrational thinking. *Journal of Personality Assessment*, *53*, 329-341. doi:10.1207/s15327752jpa5302_10

De Carlo, N. A., Falco, A., & Capozza, D. (2008). *Test di valutazione del rischio stress lavoro-correlato nella prospettiva del benessere organizzativo, Qu-Bo* [Test for the assessment of work-related stress risk in the organizational well-being perspective, Qu-Bo]. Milano, Italy: FrancoAngeli.

Downey, R. G., & King, C. V. (1998). Missing data in Likert ratings: A comparison of replacement methods. *The Journal of General Psychology*, *125*, 175-191. doi:10.1080/00221309809595542

Ellis, A., David, D., & Lynn, S. J. (2010). Rational and irrational beliefs: A historical and conceptual perspective. In D. David, S. J. Lynn, & A. Ellis (Eds.), *Rational and irrational beliefs: Research, theory, and clinical practice* (pp. 3-22). New York, NY: Oxford University Press.

Falco, A., Girardi, D., Kravina, L., Trifiletti, E., Bartolucci, G. B., Capozza, D., & De Carlo, N. A. (2013). The mediating role of psychophysic strain in the relationship between workaholism, job performance, and sickness absence: A longitudinal study. *Journal of Occupational and Environmental Medicine*, *55*, 1255-1261. doi:10.1097/JOM.0000000000000007

Falco, A., Kravina, L., Girardi, D., Dal Corso, L., Di Sipio, A., & De Carlo, N. A. (2012). The convergence between self and observer ratings of workaholism: A comparison between couples. *TPM – Testing, Psychometrics, Methodology in Applied Psychology*, *19*, 311-324. doi:10.4473/TPM19.4.54

Falco, A., Piccirelli, A., Girardi, D., Di Sipio, A., & De Carlo, N. A. (2014). "The best or nothing": The mediating role of workaholism in the relationship between perfectionism and burnout. *TPM – Testing, Psychometrics, Methodology in Applied Psychology*, *21*, 213-232. doi:10.4473/TPM21.2.7

Flett, G. L., & Hewitt, P. L. (2002). Perfectionism and maladjustment: An overview of theoretical, definitional, and treatment issues. In G. L. Flett & P. L. Hewitt (Eds.), *Perfectionism: Theory, research, and treatment* (pp. 5-31). Washington, DC: American Psychological Association.

Flett, G. L., & Hewitt, P. L. (2008). Treatment interventions for perfectionism — A cognitive perspective: Introduction to the special issue. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, *26*, 127-133. doi:10.1007/s10942-007-0063-4

Flett, G. L., Hewitt, P. L., Blankstein, K. R., & Koledin, S. (1991). Dimensions of perfectionism and irrational thinking. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, *9*, 185-201. doi:10.1007/BF01061229

Flett, G. L., Hewitt, P. L., & Cheng, W. M. W. (2008). Perfectionism, distress, and irrational beliefs in high school students: Analyses with an abbreviated survey of personal beliefs for adolescents. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, *26*, 194-205. doi:10.1007/s10942-007-0066-1

Flett, G. L., Hewitt, P. L., Nepon, T., & Besser, A. (2017). Perfectionism cognition theory: The cognitive side of perfectionism. In J. Stoeber (Ed.), *The psychology of perfectionism: Theory, research, applications*. London, UK: Taylor & Francis.

Flett, G. L., Nepon, T., & Hewitt, P. L. (2016). Perfectionism, worry, and rumination in health and mental health: A review and a conceptual framework for a cognitive theory of perfectionism. In F. M. Sirois & D. S. Molnar (Eds.), *Perfectionism, health, and wellbeing* (pp. 121-155). Cham, Switzerland: Springer International Publishing.

Flett, G. L., Newby, J., Hewitt, P. L., & Persaud, C. (2011). Perfectionistic automatic thoughts, trait perfectionism, and bulimic automatic thoughts in young women. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, *29*, 192-206. doi:10.1007/s10942-011-0135-3

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*, 39-50. doi:10.2307/3151312

Frost, R. O., Heimberg, R. G., Holt, C. S., Mattia, J. I., & Neubauer, A. L. (1993). A comparison of two measures of perfectionism. *Personality and Individual Differences*, *14*, 119-126. doi:10.1016/01918869 (93)90181-2

Frost, R. O., Marten, P., Lahart, C., & Rosenblate, R. (1990). The dimensions of perfectionism. *Cognitive Therapy and Research*, *14*, 449-468. doi:10.1007/BF01172967

Girardi, D., Falco, A., De Carlo, A., Benevene, P., Comar, M., Tongiorgi, E., & Bartolucci, G. B. (2015). The mediating role of interpersonal conflict at work in the relationship between negative affectivity and biomarkers of stress. *Journal of Behavioral Medicine*, *38*, 922-931. doi:10.1007/s10865-015-9658-x

Girardi, D., Falco, A., Piccirelli, A., Dal Corso, L., Bortolato, S., & De Carlo, A. (2015). Perfectionism and presenteeism among managers of a service organization: The mediating role of workaholism. *TPM – Testing, Psychometrics, Methodology in Applied Psychology*, *22*, 507-521. doi:10.4473/TPM22.4.5

Gnilka, P. B., Ashby, J. S., & Noble, C. M. (2012). Multidimensional perfectionism and anxiety: Differences among individuals with perfectionism and tests of a coping-mediation model. *Journal of Counseling & Development*, *90*, 427-436. doi:10.1002/j.1556-6676.2012.00054.x

Hawthorne, G., & Elliott, P. (2005). Imputing cross-sectional missing data: Comparison of common techniques. *Australian and New Zealand Journal of Psychiatry*, *39*, 583-590. doi:10.1111/j.14401614.2005. 01630.x

Hewitt, P. L., & Flett, G. L. (1991). Perfectionism in the self and social contexts: Conceptualization, assessment, and association with psychopathology. *Journal of Personality and Social Psychology*, *60*, 456-470. doi:10.1037/0022-3514.60.3.456

Hu, L., & 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*, 1-55. doi:10.1080/10705519909540118

Jöreskog, K. G., & Sörbom, D. (2006). *LISREL 8.80 for Windows*. Lincolnwood, IL: Scientific Software International.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: The Guilford Press.

Kravina, L., Falco, A., Girardi, D., & De Carlo, N. A. (2010). Workaholism among management and workers in an Italian cooperative enterprise. *TPM – Testing, Psychometrics, Methodology in Applied Psychology*, *17*, 201-2016. doi:10.4473/TPM.17.4.2

Kubota, K., Shimazu, A., Kawakami, N., Takahashi, M., Nakata, A., & Schaufeli, W. B. (2010). Association between workaholism and sleep problems among hospital nurses. *Industrial Health*, *48*, 864-871. doi:10.2486/indhealth.MS1139

Liang, Y.-W., & Chu, C.-M. (2009). Personality traits and personal and organizational inducements: Antecedents of workaholism. *Social Behavior and Personality: An International Journal*, *37*, 645-660. doi:10.2224/sbp.2009.37.5.645

MacKenzie, S., Podsakoff, P., & Podsakoff, N. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *Management Information Systems Quarterly*, *35*(2), 293-334.

MacKinnon, D. P., Cheong, J., & Pirlott, A. G. (2012). Statistical mediation analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological* (pp. 313-331). Washington, DC: American Psychological Association.

Mazzetti, G., Schaufeli, W. B., & Guglielmi, D. (2016). Are workaholism and work engagement in the eye of the beholder? *European Journal of Psychological Assessment*. Advance online publication. doi:10.1027/1015-5759/a000318

McMillan, L. H. W., & O'Driscoll, M. P. (2008). The wellsprings of workaholism: A comparative analysis of the explanatory theories. In R. J. Burke & C. L. Cooper (Eds.), *The long work hours culture: Causes, consequences and choices* (pp. 85-111). Bingley, UK: Emerald.

Naughton, T. J. (1987). A conceptual view of workaholism and implications for career counseling and research. *Career Development Quarterly*, *35*, 180-187. doi:10.1002/j.2161-0045.1987.tb00912.x

Ng, T. W. H., Sorensen, K. L., & Feldman, D. C. (2007). Dimensions, antecedents, and consequences of workaholism: A conceptual integration and extension. *Journal of Organizational Behavior*, *28*, 111-136. doi:10.1002/job.424

Nieuwenhuijsen, K., Verbeek, J. H. A. M., de Boer, A. G. E. M., Blonk, R. W. B., & van Dijk, F. J. H. (2010). Irrational beliefs in employees with an adjustment, a depressive, or an anxiety disorder: A prospective cohort study. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, *28*, 57-72. doi:10.1007/s10942-007-0075-0

Nottingham, E. J. (1992). Use of the Survey of Personal Beliefs scale: Further validation of a measure of irrational beliefs with psychiatric inpatients. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, *10*, 207-217. doi:10.1007/BF01062375

Ogai, H., & Okayasu, T. (2010). Effects of job stressors and irrational beliefs on burnout in nurses. *The Japanese Journal of Health Psychology*, *23*, 13-20. doi:10.11560/jahp.23.1_13

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, *63*, 539-569. doi:10.1146/annurev-psych-120710-100452

Popov, B., Majstorović, N., Matanović, J., Jelić, D., & Raković, S. (2016). Predictors of employees' psychophysical health and sickness absenteeism: Modelling based on REBT framework. *Psihologija*, *49*(1), 67-86.

Salanova, M., López-González, A. A., Llorens, S., del Líbano, M., Vicente-Herrero, M. T., & Tomás-Salvá, M. (2016). Your work may be killing you! Workaholism, sleep problems and cardiovascular risk. *Work & Stress*, *30*, 228-242. doi:10.1080/02678373.2016.1203373

Schaufeli, W. B. (2016). Heavy work investment, personality and organizational climate. *Journal of Managerial Psychology*, *31*, 1057-1073. doi:10.1108/JMP-07-2015-0259

Schaufeli, W. B., Bakker, A. B., van der Heijden, F. M. M. A., & Prins, J. T. (2009). Workaholism among medical residents: It is the combination of working excessively and compulsively that counts. *International Journal of Stress Management*, *16*, 249-272. doi:10.1037/a0017537

Schaufeli, W. B., Salanova, M., González-Romá, V., & Bakker, A. B. (2002). The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness Studies*, *3*, 71-92. doi:10.1023/A:1015630930326

Schaufeli, W. B., Taris, T. W., & Bakker, A. B. (2008). It takes two to tango: Workaholism is working excessively and working compulsively. In R. J. Burke & C. L. Cooper (Eds.), *The long work hours culture: Causes, consequences and choices* (pp. 203-225). Bingley, UK: Emerald.

Shimazu, A., Schaufeli, W. B., Kamiyama, K., & Kawakami, N. (2015). Workaholism vs. work engagement: The two different predictors of future well-being and performance. *International Journal of Behavioral Medicine*, *22*, 18-23. doi:10.1007/s12529-014-9410-x

Sirois, F. M., & Molnar, D. S. (2016). Conceptualizations of perfectionism, health, and wellbeing: An introductory overview. In F. M. Sirois & D. S. Molnar (Eds.), *Perfectionism, health, and well-being* (pp. 1-21). Cham, Switzerland: Springer International Publishing.

Slaney, R. B., Rice, K. G., Mobley, M., Trippi, J., & Ashby, J. S. (2001). The revised almost perfect scale. *Measurement and Evaluation in Counseling and Development*, *34*(3), 130-145.

Spence, J. T., & Robbins, A. S. (1992). Workaholism: Definition, measurement, and preliminary results. *Journal of Personality Assessment*, *58*, 160-178. doi:10.1207/s15327752jpa5801_15

Spurk, D., Hirschi, A., & Kauffeld, S. (2016). A new perspective on the etiology of workaholism: The role of personal and contextual career-related antecedents. *Journal of Career Assessment*, *24*, 747-764. doi:10.1177/1069072715616127

Stoeber. J., & Damian, L. E. (2016). Perfectionism in employees: Work engagement, workaholism, and burnout. In F. M. Sirois & D. S. Molnar (Eds.), *Perfectionism, health, and well-being* (pp. 265-283). Cham, Switzerland: Springer International Publishing.

Stoeber, J., Davis, C. R., & Townley, J. (2013). Perfectionism and workaholism in employees: The role of work motivation. *Personality and Individual Differences*, *55*, 733-738. doi:10.1016/j.paid.2013.06.001

Stoeber, J., & Otto, K. (2006). Positive conceptions of perfectionism: Approaches, evidence, challenges. *Personality and Social Psychology Review*, *10*, 295-319. doi:10.1207/s15327957pspr1004_2

Taris, T. W., Van Beek, I., & Schaufeli, W. B. (2010). Why do perfectionists have a higher burnout risk than others? The mediational effect of workaholism. *Romanian Journal of Applied Psychology*, *12*(1), 1-7.

Terjesen, M. D., Salhany, J., & Sciutto, M. J. (2009). A psychometric review of measures of irrational beliefs: Implications for psychotherapy. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, *27*, 83-96. doi:10.1007/s10942-009-0093-1

Tofighi, D., & MacKinnon, D. P. (2011). RMediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods*, *43*, 692-700. doi:10.3758/s13428-011-0076-x

Trifiletti, E., Vianello, M., & Capozza, D. (2013). Validazione delle scale del Test Qu-Bo con modelli di analisi fattoriale confermativa [Validation of the scales of the Qu-Bo Test using confirmatory factor analyses]. In N. A. De Carlo, A. Falco, & D. Capozza (Eds.), *Stress, benessere organizzativo e performance: Valutazione & intervento per l'azienda positiva* (pp. 464-482). Milano, Italy: FrancoAngeli.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4-70. doi:10.1177/109442810031002

Van Beek, I., Taris, T. W., Schaufeli, W. B., & Brenninkmeijer, V. (2013). Heavy work investment: Its motivational make-up and outcomes. *Journal of Managerial Psychology*, *29*, 46-62. doi:10.1108/JMP-06-2013-0166

Van Beek, I., Hu, Q., Schaufeli, W. B., Taris, T. W., & Schreurs, B. H. J. (2012). For fun, love, or money: What drives workaholic, engaged, and burned-out employees at work? *Applied Psychology*, *61*, 30-55. doi:10.1111/j.1464-0597.2011.00454.x

Van Wijhe, C. I., Peeters, M. C. W., & Schaufeli, W. B. (2013). Irrational beliefs at work and their implications for workaholism. *Journal of Occupational Rehabilitation*, *23*, 336-346. doi:10.1007/s10926-012-9416-7

Van Wijhe, C. I., Schaufeli, W. B., & Peeters, M. C. W. (2010). Understanding and treating workaholism: Setting the stage for successful interventions. In R. J. Burke & C. L. Cooper (Eds.), *Risky business: Psychological, physical and financial costs of high risk behavior in organizations* (pp. 107-134). Farnham, UK: Gower Publishing.

Watson, P. J., Simmons, N. M., Weathington, B. L., O'Leary, B. J., & Culhane, S. E. (2009). Psychometric analysis and tentative shortening of survey of personal beliefs. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, *27*, 201-212. doi:10.1007/s10942-007-0077-y

TPM®

INSTRUCTIONS TO AUTHORS

The journal Testing, Psychometrics, Methodology in Applied Psychology (TPM) emphasizes the publication of original research which offers new knowledge to the fields of applied psychology, social psychology, work and organizational psychology, personality and individual differences, psychological methods. We are interested in publishing several types of contributions:

1. Empirical investigations that enlarge conceptual understanding in Psychology (original studies or meta-analyses);
2. Literature reviews that advance theory and/or promote novel studies;
3. Proposals of new analytic methods and new measurement tools;
4. Qualitative studies on subjects that are difficult to investigate with quantitative methods;
5. Commentary articles of target contributions are well accepted.

A sample of topics relevant to TPM includes attitudes and stereotypes; work motivation; work stress, health, and well-being; leadership; diversity in work settings; regulatory foci and organizational behavior; prejudice and prejudice reduction; political psychology; intergroup relations; implicit social cognition; acculturation processes; immigration problems; intimate relationships; family processes; individual differences in personality; measurement methods; construction or adaptation of measurement instruments. Contributions that emphasize potential professional applications are well accepted. The contributions containing scales' validation are accepted only if the paper is aimed at broader theoretical and applicative aims and the validation of the scale is a preliminary step to achieve, in the same paper, these aims.

All manuscripts should follow the style guidelines set forth in the sixth edition (2010) of the Publication Manual of the American Psychological Association (APA). Submission of an article implies that the work reported has not been published, in all or in part, previously (also in electronic form or in another language), and it is not under consideration for publication elsewhere.

It is expected that the empirical research submitted was performed in a manner consistent with the APA ethical principles in the conduct of research. It is also expected that the publication of the paper is approved by all authors and tacitly or explicitly by the responsible authorities where the work was performed.

Authors are asked to secure the necessary letters of permission from copyright holders for reproducing copyright material (test and scale items, questionnaires, tables, figures).

Authors must acknowledge, if any, the source of support for the research and any financial involvement that might indicate a conflict of interest should be disclosed at the time of submission.

A statement transferring copyright from the author to the journal will be required after the final acceptance of the manuscript for publication. The managing editor will supply the necessary forms for this transfer.

Manuscripts

Manuscripts, exclusively in English, must be submitted online to the following e-mail address tpmap@tpmap.org. The submission must be completed attaching to the e-mail a specific form in .doc or .odt format filled in following the instructions published on the web page http://www.tpmap.org/InstructionsToAuthors.htm. The direct links to the forms are tpmap.org/manuscript.doc or tpmap.org/manuscript.odt.

Questions regarding journal policy should be addressed to ilaria.ometto@tpmap.org.

Papers should not exceed 10,000 words in length (11-point Times New Roman font), including references and notes. The word count should appear on the title page. Also short research reports of no more of 2,500 words, including references and notes, are invited. Short research reports have a more rapid review time.

Manuscripts should include an abstract of no more than 150 words, and beneath the abstract five keywords.

Authors' names and their affiliations must be indicated on the title page, together with full contact details for one of the authors: affiliation, address, e-mail address, telephone/fax.

Tables

Tables should be numbered consecutively and indicated by numbers in the text. Each table should have a descriptive title at the top. All the abbreviations that are used should be explained. Do not use too many tables, and ensure that the data presented in tables do not duplicate results reported in the article.

Figures

Figures should be numbered consecutively and indicated by numbers in the text. Each figure should have a caption below the figure. A caption should include a brief title and a description of the figure. Keep the text

to a minimum, but explain all symbols and abbreviations used. Figures must be reported at the end of the manuscript.

Appendices
Appendices should be indicated by letters and cited in the text. Each appendix should have a descriptive title at the top. All the abbreviations that are used should be explained. Appendices must be reported at the end of the manuscript.

References
The author is responsible for the accuracy of citations.
Please ensure that every reference cited in the text is also present in the reference list (and vice versa).

Citations in the text
Authors are invited to follow the APA's indications which are reported in the Publication Manual (2010, sixth edition).

Reference list
Authors are invited to follow the APA's indications which are reported in the Publication Manual (2010, sixth edition). They have also to use appropriate DOI for articles.

Review process
The journal operates a policy of anonymous double-blind peer review. Papers will normally be commented on by at least two independent expert referees.
The referees will not be made aware of the identity of the author. Make sure that the manuscript contains no clues as to author's identity. All information about authorship should be confined to the title page, and the text should be free of such clues as identifiable self-citations (e.g., "In our earlier work…").
The author can mail to the managing editor tpmap@tpmap.org in order to verify the status of the submitted article.
The action and managing editor will make revisions aimed at greater clarity and conformity of style in compliance with the author.
All the authors will receive one e-copy of the journal.

FURTHER ADVICE FOR PREPARATION

Title of the article
It should be concise, but informative. Avoid abbreviations.
A running head must be submitted as well.

Authors' affiliation and address
Present the authors' affiliation addresses (where the work was done) below the names.

Corresponding author
Indicate who will follow correspondence at all stages of publication. For the corresponding author provide: the complete postal address; fax and telephone numbers (with country and area code); the e-mail address.

Abstract
The abstract should report the aims of the research, the main results and main conclusions. References should be avoided, but if necessary, cite the author and the year. Avoid abbreviations.

Acknowledgements
Report acknowledgements in a separate section before the references. You can list people who supported you during the research and during the process of writing and editing the manuscript.

Footnotes
Footnotes should be used rarely. Indicate the position of the footnotes in the text using superscript Arabic numbers. Footnotes will be presented separately at the end of the article, before references and acknowledgements.

ISSN 1972-6325