Computers in Human Behavior 27 (2011) 1875–1883

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Reliability and validity of self-reported burnout in college students: A cross randomized comparison of paper-and-pencil vs. online administration

Juliana Alvares Duarte Bonini Campos [a], Miriane Lucindo Zucoloto [a], Fernanda Salloume Sampaio Bonafé [a], Paula Cristina Jordani [a], João Maroco [b,*]

[a] Faculdade de Odontologia de Araraquara-SP, Universidade Estadual Paulista – UNESP, Rua Humaitá, 1680, 14801-903 Araraquara, São Paulo, Brazil
[b] Departamento de Estatística e Unidade de Investigação em Psicologia e Saúde, ISPA – Instituto Universitário, Rua Jardim do Tabaco, 34, 1149-041 Lisboa, Portugal

ARTICLE INFO

Article history:
Available online 17 May 2011

Keywords:
Internet questionnaires
Paper-and-pencil questionnaires
Reliability
Validity
Student Burnout

ABSTRACT

Internet data collection is becoming increasingly popular in all research fields dealing with human perceptions, behaviors and opinions. Advantages of internet data collection, when compared to the traditional paper-and-pencil format, include reduced costs, automatic database creation, and the absence of researcher-related bias effects, such as availability and complete anonymity. However, the validity and reliability of internet gathered data must be established, in comparison to the usual paper-and-pencil accepted formats, before an inferential analysis can be done. In this study, we compared questionnaire data gathered from the internet with that from the traditional paper-and-pencil in a sample of college students. The questionnaires used were the Maslach Burnout Inventory – Student Survey (MBI-SS), the Oldenburg Burnout Inventory (OBI-SS) and the Copenhagen Burnout Inventory (CBI-SS). Data was gathered through a within-subject cross randomized and counterbalanced design, on both internet and paper-and-pencil formats. The results showed no interference in the application order, and a good reliability for both formats. However, concordance between answers was generally higher in the paper-and-pencil format than on the internet. The factorial structure was invariant in the three burnout inventories. Data gathered in this study supports the Internet as a convenient, user-friendly, comfortable and secure data gathering method which does not affect the accepted factorial structures existent in the paper format of the three burnout inventories used.

1. Introduction

In the last few decades, researchers have started using the internet as a vehicle to gather data. Several questionnaires have been adapted to online forms and data has been flowing steadily from these tools. The psychometric quality of data gathered on the internet, by means of online questionnaires, has been extensively scrutinized and several studies have tried to understand how the internet may affect the validity and the reliability of different questionnaires and psychometric scales (Bates & Cox, 2008; Bressani & Downs, 2002; Buchanan et al., 2005; Carlbring et al., 2007; Fish, McGuire, Hogan, Morrison, & Stewart, 2010; Hedman et al., 2010; Herrero & Meneses, 2006; Hewson & Charlton, 2005; Im et al., 2005; Luce et al., 2007; McCabe, Boyd, Young, Crawford, & Pope, 2005; Meyerson & Tryon, 2003; Miller et al., 2002; Naus, Philipp, & Samsi, 2009; Riva, Teruzzi, & Anolli, 2003; Suris, Borman, Lind, & Kashner, 2007). Some of the advantages of the internet data collection, when compared to the traditional paper-and-pencil self-report, mail and telephone surveys, include a lower cost, a larger sampling frame which may include geographically distant areas, respondent commodity, absence of interviewer biased responses, easy database creation, reduced data typos and commodity of data analysis. (Luce et al., 2007; Miller et al., 2002; Naus et al., 2009; Reips, 2001; Riva et al., 2003). However, as pointed out by Riva et al. (2003), Buchanan et al. (2005), Herrero and Meneses (2006), Luce et al. (2007), Naus et al. (2009) and Hedman et al. (2010), the psychometric properties of the data from measurement scales deployed online, are not necessarily equivalent to the psychometric properties evaluated in paper-and-pencil application of the same instruments, which have been previously shown to produce valid and reliable data. Thus, the psychometric evaluation of data gathered from online measurement instruments must be performed before one can use the data in further inferential analysis. Additionally, as stated by Bowling (2005), different response methods can produce different bias in the data and these should be investigated whenever a measurement instrument is deployed in a form different from the one validated originally.

Studies of the equivalence between internet and paper-and-pencil forms have been conducted with several measurement instruments for diverse areas, such as alcohol and drug abuse (McCabe et al., 2005; Miller et al., 2002), sexual desire clues

* Corresponding author. Tel.: +351 218811700; fax: +351 218860954.
E-mail address: joao.maroco@ispa.pt (J. Maroco).

0747-5632/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chb.2011.04.011

(Meyerson & Tryon, 2003), stress and depression (Herrero & Meneses, 2006; Naus et al., 2009), panic/agoraphobia (Carlbring et al., 2007), aggression, impulsivity, health functioning (Suris et al., 2007), psychological assessment (Naus et al., 2009; Suris et al., 2007), prospective memory (Buchanan et al., 2005), social anxiety disorder (Hedman et al., 2010), personality testing (Buchanan & Smith, 1999; Naus et al., 2009) and Chronic Pain Acceptance (Fish et al., 2010).

The influence of the internet on possible bias in responses, lack of reliability, validity and the factorial structure of measurement instruments is controversial. Several studies have shown that some instruments maintain their psychometric properties when adapted to the internet, when compared to the traditional paper-and-pencil format (Bates & Cox, 2008; Bressani & Downs, 2002; Fish et al., 2010; Herrero & Meneses, 2006; Hewson & Charlton, 2005; McCabe et al., 2005; Meyerson & Tryon, 2003; Thorndike et al., 2009). However, several other reports have shown that the reliability, validity and factor structure of instruments adapted to the internet were not equivalent to the observed in paper-and-pencil applications (Buchanan et al., 2005; Finegan & Allen, 1994; Hedman et al., 2010; Im et al., 2005; Luce et al., 2007; Naus et al., 2009; Suris et al., 2007; Whitener & Klein, 1995). It needs to be pointed out that, although invariance of factor structures of internet vs. paper-and-pencil applications are a common concern amongst researchers, only a very few studies have explored the formats' equivalence using appropriate cross-sampling within-subject designs (Bressani & Downs, 2002; Carlbring et al., 2007; Naus et al., 2009), and/or multi-group structural equation modeling analysis to demonstrate invariance of factorial structures (Buchanan et al., 1999; Fish et al., 2010; Herrero & Meneses, 2006; Hewson & Charlton, 2005).

As far as we know, there is no published research on Burnout inventories application through the internet, nor any studies on the psychometric properties of the data gathered on another format different from paper-and-pencil.

Burnout is a multifactorial syndrome characterized by three key dimensions: emotional exhaustion, cynicism and reduced efficiency (Maslach, Jackson, & Leiter, 1996). The first studies about burnout in the work place showed that this syndrome can impair productivity, damage human relations, cause depression and be the precursor to more serious mental and psychological conditions. Initially, burnout was considered a psychological syndrome specific to professionals performing aid or support tasks to other people (e.g., doctors, lawyers, psychologists, teachers, etc.). Research on the burnout syndrome has shown, however, that it is not exclusive to aid-related professionals. On the contrary, it is extensible to all professional activities (Leiter & Schaufeli, 1996; Maslach, Schaufeli, & Leiter, 2001). The concept of burnout has also been applied to people involved in such activities as full-time motherhood or undergraduate/graduate full-time studies (Koeske & Koeske, 1991; Maroco & Tecedeiro, 2009; Maroco, Tecedeiro, Martins, & Meireles, 2008), that are not generally labeled as professions, but share some of the characteristics of the so-called 'classical' jobs. According to Maroco et al. (2008), college students constitute a population susceptible to burnout, since they experience multiple socio-economic constrains, academic work requirements (term papers, tests and examinations), social and personal pressures related to teachers and colleagues. On the other hand, they frequently experience a lack of quality time spent with family and friends, and may experience stress related with future professional expectations and usefulness of their studies. Student burnout has been mainly evaluated with the Maslach Burnout Inventory (MBI), adapted to college students by Schaufeli, Martinez, Pinto, Salanova, and Bakker (2002) from the MBI-General Survey proposed by Maslach et al. (1996). MBI's use to measure burnout has been criticized by some researchers, and there are two other burnout measuring instruments in the public domain, namely the Oldenburg Burnout Inventory (OLBI) (Demerouti, Bakker, Nachreiner, & Schaufeli, 2001) and the Copenhagen Burnout Inventory (CBI) (Kristensen, Borritz, Villadsen, & Christensen, 2005). As far as we know, these inventories have been applied only in the paper-and-pencil format. More recently, the MBI has been adapted to the internet (Maroco & Tecedeiro, 2009; Maroco et al., 2008). However, no studies of the psychometric properties as well as of the factorial invariance of the internet format, when compared to paper-and-pencil format, have been reported. In this study, we investigated the reliability, inter-format agreement, factorial validity and the invariance of the MBI-SS, OLBI and CBI applied through the internet, as an online format vs. paper-and-pencil format, using a within-subjects counterbalanced design.

2. Methods

2.1. Design

A crossover experiment was used in two sequential steps. In the first evaluation step, half the participants were randomly assigned the paper-and-pencil questionnaire (group 1) and the other half the internet questionnaire (group 2). After a 1-week washout period, the two groups were switched: group 1 was therefore assigned the internet questionnaire in the 2nd moment, while group 2 was assigned to the paper-and-pencil questionnaire. The pairing of the two questionnaire forms (internet and paper-and-pencil) was assured by an anonymous alphanumeric code given to each participant who filled it in on both questionnaire formats. Paper and pencil questionnaires were filled in a classroom while the internet questionnaires were filled in the computer lab. In both situations, after giving the filling instructions, the researcher left the rooms unattended. Items' order in the three questionnaires was maintained as in the original versions, but the three questionnaires order was randomly assigned.

2.2. Participants

Participants were recruited from a pool of 170 graduating students from the Faculdade de Odontologia de Araraquara, UNESP, Brazil. After a brief presentation of the project and its objectives, 151 participants, from both genders, signed up voluntary (88.8% sampling rate). From these 151 students, one did not answer the internet questionnaire and was therefore removed from the study. Only students that completely answered both paper-and-pencil and internet questionnaires were included in the final sample. Thus, the final, non-probabilistic convenience, sample was composed by 150 participants. Participants average age was 21.2 (SD = 2.32) years old. In terms of gender, 24.7% were male and 75.3% were female.

2.3. Instruments

The burnout status was evaluated by a set of three burnout inventories, that were available in the researched literature: the Maslach Burnout Inventory – Student Survey (Schaufeli et al., 2002); the Oldenburg Burnout Inventory (Demerouti et al., 2001) and the Copenhagen Burnout Inventory (Kristensen et al., 2005).

2.3.1. Maslach Burnout Inventory – Student Survey (MBI-SS)

The Maslach Burnout Inventory (MBI), originally proposed by Maslach and Jackson (1981), is the most used Burnout inventory both in research and clinical practice to diagnose Burnout. Schaufeli et al. (2002) adapted the MBI to college students, generating the Maslach Burnout Inventory – Student Survey (MBI-SS).

The MBI-SS is composed by 15 items scored on a 7-point frequency rating scale that ranges from 0 (Never) to 6 (Always). The MBI-SS defines three scales: Emotional Exhaustion, Cynicism and Efficacy. In this research, the MBI-SS used was the Portuguese adaptation, in consonance with the 2009 linguistic agreement adopted by the Portuguese speaking countries, from the Maroco et al. (2008) (European Portuguese) and Carlotto and Câmara (2006) (Brazilian Portuguese) previous adaptations of the MBI-SS. In these studies, the three factor structure of the original MBI-SS inventory was confirmed in both the Portuguese and Brazilian samples.

2.3.2. Oldenburg Burnout Inventory (OLBI)

Although MBI is the most used burnout inventory, its use is not free from criticism, namely on the item formulation and core dimensions assessed. Building on some of those criticisms, Demerouti et al. (2001) developed the Oldenburg Burnout Inventory (OLBI) for the general population with empirical construct validity (Demerouti, Bakker, Vardakou, & Kantas, 2003; Halbesleben & Demerouti, 2005). The OLBI is composed of 16 Likert type items, scored on a 4-point rating scale, ranging from 1 (Strongly agree) to 4 (Strongly disagree). The OLBI defines two scales: Exhaustion and Disengagement. Scores for 4 items of each dimension need to be reversed before the sub-scales' scores can be added. The Halbesleben and Demerouti (2005) English version of the OLBI was translated to Portuguese and back-translated to English, with some items adapted to the college context.

2.3.3. Copenhagen Burnout Inventory (CBI)

Kristensen et al. (2005) proposed the Copenhagen Burnout Inventory (CBI) for use with the general population. The CBI is composed of 19 items on a 5-point rating scale from 1 (Never) to 5 (Always). It defines three scales: personal burnout, work related Burnout and Client related burnout. Kristensen et al. (2005) provided empirical support of the reliability and validity of the CBI, and showed it was adapted to different populations, from 15 different professional occupations. As far as we know, no previous adaptation to college students had been attempted. In this research we translated and adapted the CBI to the Portuguese speaking college student population. In this process, the original client-related Burnout scale was increased to include two new scales: fellow students' related burnout and teacher related burnout. The work-related burnout was also adapted to course-work related burnout.

2.4. Procedure

The invitation to participate in this research project was done verbally by the project researchers. After the project presentation and information regarding voluntary participation, students' consent was obtained in accordance with the Local Ethical Committee on research with Humans (Comitê de Ética em Pesquisa em Seres Humanos da Faculdade de Odontologia de Araraquara – UNESP). Each participant was anonymously assigned an identification code, used with the single propose to pair both forms of the questionnaire. A specific website was built to host the internet questionnaires. These questionnaires were available online for a 2-month period (November/December 2009). The participants were shown all the questionnaire items and all answers to the items could be checked and changed before submission. Paper-and-pencil questionnaires were distributed amongst the participants, after regular class hours, by a research team member who left the room unattended during the questionnaire filling. To evaluate the reproducibility of the answers to the paper-and-pencil and internet questionnaires, the participants were required to fill both forms within a 1-week interval. This interval intended, on one hand, to prevent memory effects in the answers and, on the other, to prevent changes in burnout determinants (e.g., course work, mid-terms, etc...) that could motivate different answers from the previous ones. The order in which the questionnaire was filled by each participant was randomly assigned.

2.5. Psychometric and statistical analysis

2.5.1. Face related validity

Face related validity was assessed by a multidisciplinary team composed of eight psychometricians, psychologists and linguistic experts. The semantic, idiomatic, cultural and conceptual equivalence of the original inventories and the translated/adapted inventories was assessed. The consensual versions were pre-tested with a group of 20 voluntary students, to evaluate the comprehension index (CI) and, if necessary, item reformulation was performed until the overall CI was larger than 80%.

2.5.2. Construct related validity

Construct related validity for the burnout inventories was assessed by means of a Confirmatory factor analysis. Factorial validity was assumed when the goodness of fit statistics χ^2/df ranged from 2 to 5, CFI (*confirmatory fit index*) and GFI (*goodness of fit index*) were larger than 0.9 and RMSEA (*root mean square error of approximation*) was smaller than 0.05 with $P[\text{rmsea} \leqslant 0.05] \geqslant 0.05$ (see, e.g. Schumacker & Lomax, 1996). As appropriate, items with low individual reliability were removed previously to further analysis so that lack of factorial validity would not confound the structural invariance analysis. Factorial invariance of Burnout inventories in the paper-and-pencil vs. internet formats was assessed by a multigroup confirmatory factor analysis. Factorial invariance was accepted when the measurement weights and factor covariances did not differ significantly ($p > 0.05$) accordingly to the multi-group Qui-square difference test (see, e.g., Loehlin, 2003 or Maroco, 2010). Convergent validity was evaluated with an average variance extracted (AVE) and composite reliability (CR) per factor. Convergent validity was accepted when AVE $\geqslant 0.5$ and CR $\geqslant 0.7$ (Fornell & Larcker, 1981; Hair, Black, Babin, Anderson, & Tatham, 2005). Scale discriminant validity was accepted when the AVE for every two scales was larger the squared Pearson correlation between the two scales (Fornell & Larcker, 1981).

2.5.3. Reliability

Reliability was estimated through internal consistency using Cronbach's alpha, and test–retest reliability. The formats' stability was evaluated with Pearson and intraclass correlations coefficients, for each scale and application format. Stratified Cronbach's alpha was estimated for the inventories totals. Reliability was assumed for Cronbach's alpha and correlations' coefficients greater than 0.7 (Maroco & Garcia-Marques, 2006). The overall equivalence for burnout, reporting from both questionnaires forms (Paper-and-pencil and Internet), was evaluated by point estimates and by a 95% interval estimates of Pearson correlation coefficients between the two forms with Z tests for statistical significance. Statistical significant results were assumed for $p < 0.05$.

3. Results

3.1. Psychometric sensitivity

Psychometric sensitivity was judged from the distribution of the items' for the different burnout inventories. The Mean, Median, Skewness and Kurtosis of all the items, in the three burnout inventories, are presented in Table 1.

Most items in all sub-scales of MBI-SS, OLBI-SS and CBI-SS show skewness and kurtosis values that do not differ markedly from the Normal Distribution, both in the paper-and-pencil and internet

Table 1
Descriptive Statistics for all items in MBI-SS, OLBI-SS and CBI-SS $n = 150$) for both paper-and-pencil and internet formats. Values outside parenthesis are for the paper-and-pencil format; values inside parenthesis are for the internet format.

Item	MBI-SS					OLBI-SS					CBI-SS				
	Mean	Median	SD	Ku	Sk	Mean	Median	SD	Ku	Sk	Mean	Median	SD	Ku	Sk
it1	2.86(2.88)	3(3)	1.51(1.50)	−0.68(−0.79)	0.16(0.07)	2.85(2.80)	3(3)	0.77(0.79)	−0.24(−0.18)	−0.28(−0.36)	3.24(3.27)	3(3)	0.87(0.82)	0.68(0.66)	−0.12(−.47)
it2	3.25(3.29)	3(3)	1.50(1.47)	−0.94(−0.87)	0.13(0.16)	2.57(2.64)	3(3)	0.80(0.73)	0.02(−0.32)	−0.12(0.03)	2.93(3.00)	3(3)	0.95(0.96)	0.06(−0.31)	0.33(0.00)
it3	2.88(2.75)	3(3)	1.67(1.45)	−1.02(−0.74)	0.11(0.18)	2.44(2.37)	2(2)	0.80(0.83)	−0.37(−0.55)	0.24(0.06)	2.96(2.98)	3(3)	1.05(0.99)	−0.23(−0.30)	0.12(0.08)
it4	2.03(2.08)	2(2)	1.55(1.56)	−0.29(−0.35)	0.62(0.62)	2.56(2.43)	3(2)	0.69(0.72)	−0.15(0.76)	−0.16(−0.62)	2.45(2.57)	2(2)	1.08(1.05)	−0.36(−0.56)	0.43(0.27)
it5	2.57(2.67)	3(2)	1.74(1.60)	−0.88(−0.71)	0.31(0.43)	1.75(1.91)	2(2)	0.70(0.73)	0.71(0.60)	0.75(0.66)	2.66(2.72)	3(3)	0.93(0.95)	0.19(0.21)	0.28(0.16)
it6	1.13(1.23)	1(1)	1.49(1.54)	1.29(1.10)	1.39(1.36)	2.86(2.93)	3(3)	0.70(0.62)	0.38(1.09)	−0.41(−0.47)	2.17(2.26)	2(2)	0.97(0.98)	−0.05(0.80)	0.38(0.72)
it7	1.15(1.27)	1(1)	1.38(1.53)	1.52(1.39)	1.39(1.41)	2.75(2.60)	3(3)	0.93(0.90)	−0.75(−0.65)	−0.29(−0.33)	3.11(3.17)	3(3)	0.98(0.99)	0.04(0.51)	0.20(−0.40)
it8	1.43(1.43)	1(1)	1.57(1.65)	0.32(0.48)	1.08(1.17)	2.63(2.47)	3(2)	1.04(0.97)	−0.98(−0.97)	−0.18(0.05)	2.43(2.45)	2(2)	1.00(0.96)	−0.12(0.11)	0.39(0.27)
it9	1.65(1.71)	1(1)	1.66(1.62)	0.35(0.01)	1.01(0.91)	2.87(2.89)	3(3)	0.78(0.75)	0.21(0.44)	−0.55(−0.59)	2.51(2.53)	2(2)	0.96(0.89)	0.43(0.00)	0.58(0.33)
it10	4.58(4.47)	5(5)	1.47(1.27)	0.91(0.52)	−1.17(−0.87)	2.60(2.47)	3(2)	0.67(0.72)	−0.10(0.40)	−0.17(0.09)	3.09(3.07)	3(3)	1.27(1.23)	−1.02(−0.89)	−0.09(−0.13)
it11	3.94(3.98)	4(4)	1.53(1.53)	−0.25(−0.69)	−0.59(−0.49)	2.34(2.27)	2(2)	0.66(0.66)	−0.02(0.29)	0.19(−0.23)	2.70(2.59)	3(3)	1.00(1.04)	−0.22(−0.18)	0.02(−0.07)
it12	4.23(4.33)	5(5)	1.39(1.27)	0.47(0.39)	−0.86(−0.85)	2.76(2.66)	3(3)	0.79(0.66)	1.69(−0.30)	−0.78(0.09)	2.00(2.07)	2(2)	1.01(1.00)	0.30(0.20)	0.75(0.64)
it13	5.15(4.99)	6(5)	1.14(1.26)	1.74(1.84)	−1.44(−1.39)	2.64(2.59)	3(3)	0.79(0.75)	−0.21(0.38)	−0.35(−0.40)	2.26(2.40)	2(2)	1.11(1.06)	−0.54(−0.41)	0.49(0.39)
it14	5.27(4.93)	6(5)	0.99(1.16)	6.13(1.69)	−1.97(−1.24)	2.09(2.15)	2(2)	0.73(0.73)	−0.11(0.34)	0.28(0.49)	2.38(2.43)	2(2)	1.05(1.05)	−0.22(−0.14)	0.36(0.27)
it15	4.26(4.14)	4(4)	1.24(1.38)	0.02(−0.35)	−0.61(−0.55)	2.65(2.58)	3(3)	0.74(0.65)	−0.39(−0.40)	0.06(0.37)	2.35(2.29)	2(2)	1.09(1.07)	−0.24(−0.19)	0.49(0.50)
It16	–	–	–	–	–	2.21(2.20)	2(2)	0.81(0.66)	−0.30(0.11)	0.31(0.20)	1.97(2.05)	2(2)	1.02(1.06)	−0.10(−0.41)	0.68(0.58)
It17	–	–	–	–	–	–	–	–	–	–	2.61(2.55)	3(3)	1.16(1.15)	−0.72(−0.53)	0.19(0.29)
It18	–	–	–	–	–	–	–	–	–	–	1.95(1.98)	2(2)	1.05(1.03)	0.08(0.49)	0.87(0.93)
It19	–	–	–	–	–	–	–	–	–	–	1.75(1.92)	1(2)	0.95(1.07)	0.64(0.19)	1.12(0.90)
It20	–	–	–	–	–	–	–	–	–	–	2.17(2.17)	2(2)	0.89(0.87)	−0.20(0.09)	0.25(0.35)
It21	–	–	–	–	–	–	–	–	–	–	1.97(1.99)	2(2)	0.95(0.88)	0.90(1.24)	0.92(0.92)
It22	–	–	–	–	–	–	–	–	–	–	1.91(2.97)	2(2)	0.91(1.87)	1.42(−1.43)	1.01(0.41)
It23	–	–	–	–	–	–	–	–	–	–	1.82(1.91)	2(2)	0.94(0.90)	0.95(1.42)	1.00(0.96)
It24	–	–	–	–	–	–	–	–	–	–	1.83(1.83)	2(2)	0.92(0.91)	0.41(0.51)	0.83(0.83)
It25	–	–	–	–	–	–	–	–	–	–	1.66(1.65)	1(1)	0.89(0.84)	1.00(2.50)	1.19(1.37)

Table 2
Factor weights (λ) and overall fit indices for the factor structures for the three burnout inventories evaluated in paper-and-pencil, internet and both (merged) samples.

Inventory	λ	χ^2/df	CFI	GFI	RMSEA
$MBI_{(paper\text{-}and\text{-}pencil)}$	0.44–0.90	2.105	0.933	0.893	0.086
$MBI_{(internet)}$	0.75–0.94	2.737	0.925	0.871	0.108
$MBI_{(Both\ samples)}$	0.60–0.92	2.360	0.913	0.856	0.068
$OLBI_{(paper\text{-}and\text{-}pencil)}$	0.38–0.82	1.498	0.965	0.958	0.058
$OLBI_{(internet)}$	0.16–0.90	1.669	0.954	0.951	0.067
$OLBI_{(Both\ samples)}$	0.28–0.85	1.758	0.924	0.925	0.050
$CBI_{(paper\text{-}and\text{-}pencil)}$	0.51–0.90	1.910	0.903	0.784	0.078
$CBI_{(internet)}$	0.58–0.91	1.830	0.929	0.811	0.075
$CBI_{(Both\ samples)}$	0.55–0.87	1.813	0.917	0.798	0.052

formats. However, item 14 of the MBI-SS and item 25 of CBI-SS showed a leptokurtic distribution for the paper-and-pencil and internet forms, respectively (see Table 1). According to Kline (2004) items with absolute values of skewness and kurtosis smaller than 3 and 7, respectively, do not present non-normality problems that limit further use. This is the case for the large majority of items in our study, for both paper-and-pencil and internet formats.

3.2. Construct related validity

Construct related validity was assessed by both factorial, convergent and discriminant validities, as described in the Methods section. A Confirmatory factor analysis showed that 3 of the items from the MBI-SS' Efficacy dimension had factor loadings smaller than 0.5 in the paper-and-pencil format but not in the internet format. Thus, items 9, 10 and 14 were removed to improve factorial validity and to prevent possible lack of fit in the invariance analysis due to poor factorial validity (see Table 2). Loadings smaller than 0.5 were also observed in both paper-and-pencil and internet formats for the OLBI-SS. The global fit assessment for the OLBI's 2 factor proposed structure was quite poor in this sample, suggesting a lack of factorial validity of this inventory for both paper-and-pencil and internet formats. Thus, to improve the factorial validity of OLBI-SS, items 3, 5, 6, 8, 9, 12, 14 and 16 were removed to improve fit (see Table 2) and only then factor invariance was evaluated. Regarding the CBI-SS inventory, only item 10 showed a factor loading lower than 0.5 in both paper-and-pencil and internet formats, while item 22 only showed a loading smaller than 0.5 in the internet format. Thus, items 10 and 22 were removed to assure good factorial validity on the overall sample (see Table 2).

Regarding the overall fit of the 2-factor (OLBI-SS), 3-factor (MBI-SS and CBI-SS) and 4-factor (CBI-SS) of the data gathered from the combined sample of internet and paper-and-pencil, the fit, after the removal of items with low individual reliability, was quite good judging from the overall fit indices results (see "both samples" in Table 2).

Convergent validity, as evaluated by the average variance extracted, is low (<0.5) for Efficacy in the MBI-SS paper-and-pencil format, but not in the internet format (see Table 3). Low convergent validity was observed for the two dimensions of the OLBI-SS in both paper-and-pencil and internet formats although composite reliability was within acceptable ranges. Within acceptable range AVE values for convergent validity were observed in the four dimensions of CBI-SS for both paper-and-pencil and internet formats (see Table 3). Composite reliability for all sub-scales was also appropriate in both formats (see Table 3). Discriminant validity, as evaluated by the AVE of the scales being larger than the squared Pearson correlation between two consecutive scales, was present in all scales and application formats. The exceptions are Personal and Study related burnout in the CBI-SS, for the internet formats (see Table 3).

Factorial invariance between paper-and-pencil and internet formats of the three inventories was evaluated by chi-square difference tests for models with free factor loadings vs. models with equal factor loadings; equal factor loadings and equal covariances and equal factor loadings, equal covariances and equal residual variances in the two deployment formats. For MBI-SS, the models with equal constrained factor loadings did not have a significantly poorer fit than the model with free loadings ($X^2(9) = 13.793$; $p = .130$) and free covariances ($X^2(15) = 21.871$; $p = .111$ (see Table 4). Thus, factorial structure was invariant between paper-and-pencil and internet formats. Factor invariance between paper-and-pencil and internet formats was also observed for

Table 4
Factor invariance between paper-and-pencil and internet formats for MBI-SS, OLBI-SS and CBI-SS. Values are, respectively, Chi-square statistics (χ^2_{diff}), degrees of freedom (df) and p values (p) for models with free parameters vs. models with equally constrained factor weights; factor weights plus covariances and factor weights plus covariances plus error residuals.

Inventory	Model	χ^2_{diff}	df	p
MBI-SS	Factor weights	13.793	9	0.130
	Covariances	21.871	15	0.111
	Error residuals	57.495	27	0.001
OLBI-SS	Factor weights	6.473	6	0.372
	Covariances	9.941	9	0.355
	Error residuals	36.505	17	0.004
CBI-SS	Factor weights	15.918	19	0.663
	Covariances	23.143	29	0.770
	Error residuals	79.751	52	0.008

Table 3
Average variance extracted (AVE), composite reliability (CR) and squared correlations between sub-scales (r^2) for MBI-SS, OLBI-SS and CBI-SS for both paper-and-pencil and internet formats.

Inventories	Paper-and-pencil			Internet		
	AVE	CR	r^2	AVE	CR	r^2
MBI-SS						
Exhaustion	0.577	0.872	0.127–0.332	0.658	0.905	0.061–0.398
Cynicism	0.725	0.887	0.227–0.332	0.786	0.916	0.188–0.398
Efficacy	0.416	0.733	0.127–0.227	0.551	0.830	0.061–0.189
OLBI-SS						
Exhaustion	0.361	0.689	0.210	0.445	0.760	0.122
Disengagement	0.442	0.749	0.210	0.397	0.681	0.122
CBI-SS						
Personal burnout	0.577	0.889	0.059–0.677	0.609	0.902	0.106–0.671
Study related burnout	0.520	0.866	0.059–0.677	0.600	0.899	0.090–0.671
Colleagues related burnout	0.652	0.918	0.059–0.131	0.696	0.932	0.090–0.190
Teacher related burnout	0.642	0.898	0.131–0.280	0.710	0.924	0.119–0.190

Table 5
Point estimates and 95% confidence intervals for Pearson correlation coefficients (r) for paper-and-pencil and Internet formats between group1 (paper application first) and group 2 (internet application first). p-values were obtained from a Z test for the equality of correlation coefficients for group 1 and group 2 for both deployment formats.

Burnout inventory	Paper and pencil (n = 75)		Internet (n = 75)		p
	r	$IC_{95\%}$	r	$IC_{95\%}$	
MBI-SS					
Exhaustion	0.606	0.44–0.73	0.604	0.39–0.76	0.989
Cynicism	0.423	0.22–0.59	0.446	0.24–0.61	0.860
Efficacy	0.549	0.37–0.69	0.613	0.45–0.74	0.564
Scale total	0.574	0.40–0.71	0.583	0.41–0.72	0.936
OLBI-SS					
Exhaustion	0.507	0.32–0.66	0.685	0.54–0.79	0.094
Disengagement	0.660	0.51–0.77	0.678	0.53–0.78	0.841
Scale total	0.653	0.50–0.77	0.687	0.54–0.79	0.710
CBI-SS					
Personal burnout	0.664	0.51–0.77	0.558	0.38–0.70	0.308
Study related burnout	0.547	0.37–0.69	0.712	0.58–0.81	0.096
Colleagues related burnout	0.651	0.45–0.79	0.622	0.46–0.74	0.774
Teacher related burnout	0.274	0.05–0.47	0.397	0.19–0.57	0.405
Scale total	0.527	0.34–0.67	0.665	0.52–0.76	0.195

Table 6
Inter-item correlation (r) and Cronbach's alpha (α) for MBI-SS, CBI-SS and OLBI-SS for both paper-and-pencil as well as internet formats (n = 150). Alpha for the scales' total is the Stratified Cronbach's Alpha.

Burnout inventory	Papel		Internet	
	$r_{inter\text{-}item}$	α	$r_{inter\text{-}item}$	α
MBI-SS				
Exhaustion	0.568	0.868	0.654	0.904
Cynicism	0.715	0.883	0.777	0.912
Efficacy	0.386	0.716	0.548	0.829
Scale total		0.983		0.986
OLBI-SS				
Exhaustion	0.356	0.690	0.437	0.756
Disengagement	0.407	0.733	0.304	0.636
Scale total		0.976		0.972
CBI-SS				
Personal burnout	0.564	0.886	0.597	0.899
Study related burnout	0.512	0.863	0.593	0.897
Colleagues related burnout	0.648	0.917	0.685	0.929
Teacher related burnout	0.639	0.898	0.708	0.924
Scale total		0.998		0.999

OLBI-SS ($X^2(6) = 6.473$; $p = .372$) and CBI-SS ($X^2(19) = 15.918$; $p = .663$) (see Table 4). Pattern and scalar invariance was demonstrated for all the burnout inventories assessed in this study. Models with constrained residuals showed poorer fit than models with free residuals ($p < .05$). However, this condition (equal residuals) is only required to assume strict factorial invariance (Meredith & Teresi, 2006).

3.3. Reliability

Pearson correlations between burnout scales from the three inventories did not show significant differences for both formats (Paper-and-pencil vs. internet) (see Table 5).

Reliability, as estimated by internal consistency, was appropriate for both methods and inventories. With the exception of the OLBI-SS, the other scales estimates of Cronbach's alfa from the internet application were marginally larger than the ones obtained with the paper-and-pencil application (see Table 6).

Format concordance for both total scores and burnout scales, as estimated by the test–retest, the intraclass correlation point estimates and the 95% confidence intervals, was observed in all of the burnout scales (see Table 7). However, format concordance were generally higher ($p < .05$) in the paper-and-pencil than internet format for the MBI-SS, OLBI-SS and CBI-SS.

4. Discussion

In this paper we evaluated the psychometric equivalence of three Burnout inventories deployed via paper-and-pencil and internet formats. As shown in Table 5, no significant effects were observed in burnout scores due the order of application of the inventories (whether paper-and-pencil was followed by internet or internet followed by paper-and-pencil). These results are consistent with those of Naus et al. (2009) on depression, quality of life and personality of college students, evaluated both on paper-and-pencil and internet formats. There are, however, reports of higher scores on the second application of several psychometric scales, either in paper-and-pencil or internet formats, which has been associated with learning effects (see e.g. Bressani & Downs, 2002; Carlbring et al., 2007). Nevertheless, it must be pointed out that in the previous studies no cross-over designs were used. That fact may be more important in explaining the order and learning effects observed, than the format of the scales applied.

Reliability for the three Burnout inventories was quite good for both paper-and-pencil as well as internet (Table 6). Similar high

Table 7
Test–retest (r) and intraclass correlation coefficient (ICC) with respective 95% confidence intervals for MBI-SS, OLBI-SS and CBI-SS in both paper-and-pencil and internet formats (n = 150).

Inventory	Test–retest					ICC				
	Paper × Paper		Internet × Internet			Papel × Papel		Internet × Internet		
	r	$IC_{95\%}$	r	$IC_{95\%}$	p	ICC	$IC_{95\%}$	ICC	$IC_{95\%}$	p
MBI-SS										
Exhaustion	0.842	0.79–0.88	0.697	0.60–0.77	0.002	0.839	0.78–0.88	0.696	0.60–0.77	0.002
Cynicism	0.711	0.62–0.78	0.592	0.48–0.69	0.074	0.708	0.62–0.78	0.589	0.47–0.68	0.076
Efficacy	0.699	0.61–0.77	0.483	0.35–0.60	0.004	0.699	0.61–0.77	0.482	0.35–0.60	0.004
Scale total	0.732	0.65–0.80	0.645	0.54–0.73	0.154	0.732	0.65–0.80	0.644	0.54–0.73	0.150
OLBI-SS										
Exhaustion	0.719	0.63–0.79	0.581	0.46–0.68	0.038	0.714	0.626–0.784	0.577	0.46–0.67	0.042
Disengagement	0.567	0.45–0.67	0.530	0.40–0.64	0.650	0.567	0.448–0.666	0.525	0.40–0.63	0.608
Scale total	0.751	0.67–0.81	0.563	0.44–0.66	0.004	0.748	0.668–0.811	0.562	0.44–0.66	0.004
CBI-SS										
Personal burnout	0.787	0.72–0.84	0.665	0.56–0.74	0.025	0.787	0.72–0.84	0.665	0.56–0.74	0.025
Study related burnout	0.766	0.69–0.82	0.642	0.54–0.73	0.033	0.766	0.69–0.82	0.642	0.54–0.73	0.033
Colleagues related burnout	0.717	0.63–0.79	0.678	0.58–0.76	0.514	0.710	0.62–0.78	0.673	0.58–0.75	0.543
Teacher related burnout	0.698	0.61–0.77	0.498	0.37–0.61	0.007	0.697	0.60–0.77	0.488	0.36–0.60	0.005
Scale total	0.785	0.72–0.84	0.704	0.61–0.78	0.116	0.784	0.71–0.84	0.703	0.61–0.78	0.118

reliabilities were also observed in internet applications of several scales and it was consensual that the reliability of several psychometric scales was not compromised by its internet deployment (Buchanan et al., 1999; Carlbring et al., 2007; Fish et al., 2010; Hedman et al., 2010; Herrero & Meneses, 2006; Meyerson & Tryon, 2003; Naus et al., 2009).

Paper-and-pencil deployment of some sub-scales showed slightly higher reproducibility and concordance than the internet deployment (see Table 7). This may be explained by a researcher–subject interaction during the distribution of the paper questionnaires in addition to a researcher presence effect vs. total anonymity in the internet application. It is consensual that the researcher's presence alone during questionnaire filling by participants may condition their responses accordingly to their perceptions of social desirability (see e.g. Bowling, 2005; Buchanan et al., 2005; Herrero & Meneses, 2006; Richman, Kiesler, Weisband, & Drasgow, 1999; Whitener & Klein, 1995).

Construct related validity, as evaluated by factorial, convergent and discriminant validities, does not seem to suffer alterations caused by the format of application. Only slight differences between the paper-and-pencil and internet formats were observed in the overall fit of the original factorial structures proposed for the three inventories. Some of the items with lower factor weights were generally associated with the paper-and-pencil format. The evaluation of convergent and discriminant validity shows no apparent effect of the inventories format (see Table 3). For both MBI-SS, CBI-SS and OLBI-SS, the internet format showed slight higher convergent and discriminant validity than the paper-and-pencil format. It is apparent that the structural composition of the inventories is more important for construct related validity than the format of application.

Factorial structure invariance between paper-and-pencil and internet formats was also analyzed after overall factorial fit was improved by removing items with low individual reliability. It did not escape our attention that the lower factor weights were generally associated with reversed items. Habituation effects to previous, non-reversed, items may cause primacy effects resulting in a larger probability of answering the first points in the scale when items are visually similar (2005). After the elimination of problematic items, the invariance analysis showed no differences in factor weight structures of MBI-SS, CBI-SS and OLBI-SS. Factorial invariance is a key requirement when deploying psychometric scales in different formats (paper-and-pencil, telephone, internet...). Unfortunately, factorial invariance between different deployment formats has not been evaluated in only but a few studies comparing equivalence of scales (see e.g. Fish et al., 2010; Herrero & Meneses, 2006; Hewson & Charlton, 2005; Thorndike et al., 2009). The fact that factor structure invariance may not be present in different formats of psychometric scales has been acknowledged in several similar studies comparing paper-and-pencil, internet and telephone interview filling methods (Carlbring et al., 2007; Hedman et al., 2010; Naus et al., 2009; Suris et al., 2007). In an earlier study on the equivalence of computerized vs. traditional research methods, Whitener and Klein (1995) identified several factors that may explain the lack of factor equivalence and threaten validity. These include socio-demographic differences between samples as has been acknowledged in other investigations (Bowling, 2005; Herrero & Meneses, 2006; Hewson & Charlton, 2005; Im et al., 2005; Klovning, Sandvik, & Hunskaar, 2009); single vs. group effects; mode of administration and instruments presentation and social desirability (Bowling, 2005; Fan & Yan, 2010; Hewson & Charlton, 2005; Thorndike et al., 2009; Whitener & Klein, 1995). To minimize these effects, namely single vs. group effects and mode of administration, we adopted a cross-over design with all the items being presented simultaneously, in the same order, and with the possibility to review and change answers if desired for both paper-and-pencil and internet formats. The webpage, used for the internet deployment of the inventories, reproduced exactly the item order and presentation format of the paper inventories. However, as Buchanan et al. (2005) and Bates and Cox (2008) pointed out, they may have been some affecting effects more pronounced in internet than in paper-and-pencil formats. These include higher perception of anonymity, absence of the interviewer supervision, lower social interactions with interviewers and other respondents, lower social desirability pressure and larger environment variability during the questionnaires filling. These effects may explain the slight larger convergent and discriminant validity for the internet format in some of the scales. However, in face of the scarcity of results, these effects may still be speculative and empirical evidence is required before reliable conclusions can be drawn. Furthermore, the cognitive capacities of the respondents must also be probed in a way to assure that the internet deployment of scale maintains construct validity, internal validity and reliability (Riva et al., 2003). In this study, we used a cross-over design to assure internal validity. This type of methodological concern is required for reliable results in the comparison of different deployment methods since it provides additional strength to the conclusions related with the control of subject

characteristics. Unfortunately, this type of design has been used in only but a few studies comparing the equivalence of formats (e.g. Bressani & Downs, 2002; Carlbring et al., 2007; Naus et al., 2009). Similarly, factor invariance, demonstrated by means of a multi-group confirmatory factor analysis, is also required before equivalence of formats can be accepted.

It must be pointed out that the conclusions from this paper-and-pencil vs. internet equivalence study may be hampered by the relatively modest sample used, as well as confounding effects of language translation, and conversion of the OBLI-SS and CBI-SS from working professionals to college students. Results from other college student samples (both Brazilian and Portuguese) have suggested factorial invariance of the translated inventories when compared to the original ones (our own data not shown). Despite these limitations, our results do show that the format of the Burnout inventories deployment may not influence the inventories psychometric properties, namely the constructs' related validity and reliability. Empirical studies, as this one, showing the equivalence of formats as well as demonstrating the psychometric qualities of the measuring scales, must be pursued before a different format of the scale can be used in collecting data (see also Hewson & Charlton, 2005; Naus et al., 2009; Whitener & Klein, 1995). It also needs to be acknowledged that the psychometric properties of a measuring instrument are specifically related to a given instrument in a given sample and not to the instrument *per se* (see also Honaker, 1988; Suris et al., 2007). This implies that the study of the psychometric properties of a given measuring instrument is required for every sample for which a generalization of results is necessary, and internet samples do not escape these principles. However, data gathered in this study shows that it is feasible to use parallel versions of student burnout inventories in both medial representations.

Acknowledgements

We deeply acknowledge the financial support provided by the Pró-reitoria de Pós-graduação da Universidade Estadual Paulista–UNESP and Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP (process: 2010/09295-5). We also thank Raquel Lara Velez Oliveira for proofreading the manuscript.

References

Bates, S. C., & Cox, J. M. (2008). The impact of computer versus paper-pencil survey, and individual versus group administration, on self-reports of sensitive behaviors. *Computers in Human Behavior, 24*(3), 903–916.

Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health (Oxford), 27*(3), 281–291.

Bressani, R. V., & Downs, A. C. (2002). Youth independent living assessment: Testing the equivalence of web and paper/pencil versions of the Ansell–Casey life skills assessment. *Computers in Human Behavior, 18*(4), 453–464.

Buchanan, T., Ali, T., Heffernan, T. M., Ling, J., Parrott, A. C., Rodgers, J., et al. (2005). Nonequivalence of on-line and paper-and-pencil psychological tests: The case of the prospective memory questionnaire. *Behavior Research Methods, 37*(1), 148–154.

Buchanan, T., & Smith, J. L. (1999). Using the Internet for psychological research: Personality testing on the world wide web. *British Journal of Psychology, 90*, 125–144.

Carlbring, P., Brunt, S., Bohman, S., Austin, D., Richards, J., Ost, L. G., et al. (2007). Internet vs. paper and pencil administration of questionnaires commonly used in panic/agoraphobia research. *Computers in Human Behavior, 23*(3), 1421–1434.

Carlotto, M. S., & Câmara, S. G. (2006). Características psicométricas do Malasch Burnout Inventory – Student Survey (MBI-SS) em estudantes universitários brasileiros. *Psico-USF, 11*(2), 167–173.

Demerouti, E., Bakker, A. B., Nachreiner, F., & Schaufeli, W. B. (2001). The job demands-resources model of burnout. *Journal of Applied Psychology, 86*(3), 499–512.

Demerouti, E., Bakker, A. B., Vardakou, I., & Kantas, A. (2003). The convergent validity of two burnout instruments – A multitrait-multimethod analysis. *European Journal of Psychological Assessment, 19*(1), 12–23.

Fan, W. M., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior, 26*(2), 132–139.

Finegan, J. E., & Allen, N. J. (1994). Computerized and written questionnaires – Are they equivalent. *Computers in Human Behavior, 10*(4), 483–496.

Fish, R. A., McGuire, B., Hogan, M., Morrison, T. G., & Stewart, I. (2010). Validation of the Chronic Pain Acceptance Questionnaire (CPAQ) in an Internet Sample and Development and Preliminary Validation of the CPAQ-8. Pain.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50.

Hair, J. F., Black, W. C., Babin, B., Anderson, R. E., & Tatham, R. L. (2005). *Multivariate data analysis* (6th ed.). Prentice Hall.

Halbesleben, J. R. B., & Demerouti, E. (2005). The construct validity of an alternative measure of burnout: Investigating the English translation of the Oldenburg Burnout Inventory. *Work and Stress, 19*(3), 208–220.

Hedman, E., Ljotsson, B., Rück, C., Furmark, T., Carlbring, P., Lindefors, N., et al. (2010). Internet administration of self-report measures commonly used in research on social anxiety disorder: A psychometric evaluation. *Computers in Human Behavior, 26*(4), 736–740.

Herrero, J., & Meneses, J. (2006). Short web-based versions of the Perceived Stress (PSS) and Center for Epidemiological Studies-Depression (CESD) Scales: a comparison to pencil and paper responses among Internet users. *Computers in Human Behavior, 22*(5), 830–846.

Hewson, C., & Charlton, J. P. (2005). Measuring health beliefs on the internet: A comparison of paper and Internet administrations of the multidimensional health locus of control scale. *Behavior Research Methods, 37*(4), 691–702.

Honaker, L. M. (1988). The equivalency of computerized and conventional MMPI administration – a critical-review. *Clinical Psychology Review, 8*(6), 561–577.

Im, E. O., Chee, W., Bender, M., Cheng, C. Y., Tsai, H. M., Mi Kang, N., et al. (2005). The psychometric properties of pen-and-pencil and internet versions of the Midlife Women's Symptom Index (MSI). *International Journal of Nursing Studies, 42*(2), 167–177.

Kline, R. B. (2004). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press.

Klovning, A., Sandvik, H., & Hunskaar, S. (2009). Web-based survey attracted age-biased sample with more severe illness than paper-based survey. *Journal of Clinical Epidemiology, 62*(10), 1068–1074.

Koeske, G. F., & Koeske, R. D. (1991). Student "burnout" as a mediator of the stress-outcome relationship. *Research in Higher Education, 32*, 415–431.

Kristensen, T. S., Borritz, M., Villadsen, E., & Christensen, K. B. (2005). The Copenhagen burnout inventory: A new tool for the assessment of burnout. *Work and Stress, 19*(3), 192–207.

Leiter, M. P., & Schaufeli, W. B. (1996). Consistency of the burnout construct across occupations. *Anxiety Stress and Coping, 9*(3), 229–243.

Loehlin, J. C. (2003). *Latent variable models: a introduction to factor, path and structural equation analysis* (4th ed.). New Jersey: Lawrence Erlbaum.

Luce, K. H., Winzelberg, A. J., Das, S., Osborne, M. I., Bryson, S. W., & Taylor, C. B. (2007). Reliability of self-report: Paper versus online administration. *Computers in Human Behavior, 23*(3), 1384–1389.

Maroco, J. (2010). *Análise de Equações Estruturais: Fundamentos teóricos, Software & Aplicações*. Pêro Pinheiro: ReportNumber.

Maroco, J., & Garcia-Marques, T. (2006). Qual a fiabilidade do alfa de Cronbach? Questões antigas e soluções modernas? *Laboratório de Psicologia, 4*(1), 65–90.

Maroco, J., & Tecedeiro, M. (2009). Inventário de Burnout para estudantes portugueses. *Psicologia, Saúde & Doenças, 10*(2), 227–235.

Maroco, J., Tecedeiro, M., Martins, P., & Meireles, A. (2008). Estrutura fatorial de segunda ordem da Escala de Burnout de Malasch para estudantes numa amostra portuguesa. *Análise Psicológica, 4*(XXVI), 639–649.

Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. *Journal of Occupational Behaviour, 2*(2), 99–113.

Maslach, C., Jackson, S. E., & Leiter, M. P. (1996). *Maslach Burnout Inventory manual* (3rd ed.). Consulting Psychologists Press.

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

McCabe, S. E., Boyd, C. J., Young, A., Crawford, S., & Pope, D. (2005). Mode effects for collecting alcohol and tobacco data among 3rd and 4th grade students: A randomized pilot study of web-form versus paper-form surveys. *Addictive Behaviors, 30*(4), 663–671.

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care, 44*(11), 68–77.

Meyerson, P., & Tryon, W. W. (2003). Validating internet research: A test of the psychometric equivalence of internet and in-person samples. *Behavior Research Methods Instruments & Computers, 35*(4), 614–620.

Miller, E. T., Neal, D. J., Roberts, L. J., Baer, J. S., Cressler, S. O., Metrik, J., et al. (2002). Test-retest reliability of alcohol measures: Is there a difference between internet-based assessment and traditional methods? *Psychology of Addictive Behaviors, 16*(1), 56–63.

Naus, M. J., Philipp, L. M., & Samsi, M. (2009). From paper to pixels: A comparison of paper and computer formats in psychological assessment. *Computers in Human Behavior, 25*(1), 1–7.

Reips, U. D. (2001). The web experimental psychology lab: Five years of data collection on the Internet. *Behavior Research Methods Instruments & Computers, 33*(2), 201–211.

Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology, 84*(5), 754–775.

Riva, G., Teruzzi, T., & Anolli, L. (2003). The use of the internet in psychological research: Comparison of online and offline questionnaires. *Cyberpsychology & Behavior, 6*(1), 73–80.

Schaufeli, W. B., Martinez, I. M., Pinto, A. M., Salanova, M., & Bakker, A. B. (2002). Burnout and engagement in university students – A cross-national study. *Journal of Cross-Cultural Psychology, 33*(5), 464–481.

Schumacker, R. E., & Lomax, R. G. (1996). A begginer's guide to structural equation modeling. New Jersey.

Suris, A., Borman, P. D., Lind, L., & Kashner, T. M. (2007). Aggression, impulsivity, and health functioning in a veteran population: Equivalency and test-retest reliability of computerized and paper-and-pencil administrations. *Computers in Human Behavior, 23*(1), 97–110.

Thorndike, F. P., Carlbring, P., Smyth, F. L., Magee, J. C., Gonder-Frederick, L., Ost, L. G., et al. (2009). Web-based measurement: Effect of completing single or multiple items per webpage. *Computers in Human Behavior, 25*(2), 393–401.

Whitener, E. M., & Klein, H. J. (1995). Equivalence of computerized and traditional research methods – The roles of scanning, social-environment, and social desirability. *Computers in Human Behavior, 11*(1), 65–75.