



Luís Miguel Lindinho da Cunha Mendes Grilo

Sumário pormenorizado do Seminário

Modelos de Equações Estruturais com Variáveis Ordinais

Universidade de Évora
Escola de Ciências e Tecnologia
Departamento de Matemática

Novembro de 2025

♦ Évora, Portugal ♦

Sumário pormenorizado

O sumário pormenorizado do seminário com o título “Modelos de Equações Estruturais com Variáveis Ordinais” é apresentado à Universidade de Évora no âmbito das Provas de Agregação em Matemática ao abrigo da alínea c) do Artigo 5.º do Decreto-Lei n.º 239/2007 de 19 de junho, com atualização no Decreto-Lei n.º 64/2023 de 31 de julho.

Évora, 10 de novembro de 2025.

A handwritten signature in blue ink, appearing to read 'Luís Miguel Grilo', written over a horizontal line.

(Luís Miguel Grilo)

Preâmbulo

O sumário pormenorizado do seminário “Modelos de Equações Estruturais com Variáveis Ordinais” é apresentado em provas públicas na Universidade de Évora no âmbito de Provas de Agregação em Matemática. O sumário terá por base investigação científica aplicada na área da modelação estatística e serão apresentados resultados de trabalhos publicados, bem como de alguns outros que se esperam concluir e publicar em breve.

A utilização de variáveis ordinais em modelos de equações estruturais requer considerações especiais, uma vez que pelo facto de estas não serem contínuas (têm, aliás, natureza discreta) violam os pressupostos dos métodos (ou, estimadores) de estimação clássicos, como o método de máxima verosimilhança. Tratar os dados ordinais (por exemplo, escalas de tipo Likert) como contínuos pode levar a estimativas enviesadas dos parâmetros, erros padrão incorretos e a conclusões erradas nas estatísticas relativas à qualidade de ajustamento do modelo, especialmente na presença de variáveis com poucas categorias ou com distribuições assimétricas. Deste modo, considerou-se pertinente um seminário onde fossem apresentados e analisados alguns dos resultados obtidos, com base em dados reais, utilizando estimadores atualmente disponíveis para este tipo de modelos.

Conteúdo

1 Introdução	9
2 Modelos de Equações Estruturais	9
2.1 Abordagem com Base na Covariância	10
2.2 Abordagem com Base na Variância	13
3 Exemplo de Aplicação dos Modelos de Equações Estruturais	14
4 Breve Descrição de Trabalhos Publicados ou em Fase de Publicação	22
5 Considerações Finais.....	25
6 Trabalhos Futuros	25
7 Referências.....	26
APÊNDICE A.....	29

1 Introdução

Com os modelos de equações estruturais (Structural Equation Modeling - SEM) é possível analisar relações complexas entre variáveis observadas (também designadas por variáveis manifestas, variáveis medidas, indicadores ou itens) e variáveis não observadas (que se designam por variáveis latentes, fatores ou construtos latentes). Estes modelos, combinam aspetos de análise fatorial, de regressão múltipla e de análise de caminhos/trajetórias, com o objetivo principal de avaliar em que grau um modelo teórico proposto é suportado pelos dados [1-3].

Os SEM têm sido usados em diferentes áreas do conhecimento, nomeadamente em Psicometria, em Ciências Sociais e Comportamentais, e em Ciências da Saúde, onde é frequente a existência de variáveis qualitativas medidas em escala ordinal (tipo Likert) [4]. Estes modelos permitem: (i) representar construtos latentes a partir de indicadores observáveis (variáveis manifestas); (ii) testar teorias com base em modelos causais complexos especificados *a priori*; (iii) modelar relações diretas e indiretas entre construtos latentes, bem como efeitos de mediação e de moderação [5]; (iv) comparar modelos teóricos concorrentes.

O recurso a diagramas de trajetórias – “desenhados/construídos” considerando a teoria subjacente ao fenómeno em estudo, eventualmente, conjugada com algum conhecimento empírico (*a priori*) – facilita a interpretação e a comunicação dos resultados obtidos com os SEM, especialmente em contextos aplicados.

Relativamente aos modelos multivariados convencionais, os SEM possuem duas vantagens principais (i) capacidade de examinar simultaneamente múltiplas influências e múltiplas respostas, bem como considerar erros de medida de forma explícita [1, 2, 5]; (ii) facilidade em serem aplicados, de forma eficiente, a problemas complexos com grandes volumes de dados e de diferentes tipos [6]. Atendendo a estas características, os SEM são considerados modelos de segunda geração muito eficazes para gerar conhecimento em múltiplos domínios.

2 Modelos de Equações Estruturais

Os SEM estão organizados nos submodelos de medida e de estrutura (ou, estrutural). Depois de centradas, as variáveis podem representar-se formalmente por [1]:

$$x = \Lambda_x \xi + \delta \quad (1)$$

$$y = \Lambda_y \eta + \varepsilon \quad (2)$$

$$\eta = B\eta + \Gamma\xi + \zeta. \quad (3)$$

Os modelos de medida em (1) e (2), definem como construtos latentes independentes (exógenos) ξ e dependentes (endógenos ou de resposta) η são operacionalizadas pelas variáveis observadas ou manifestas x (independentes ou preditoras) e y (dependentes), respetivamente. Em (3), o modelo estrutural define as relações causais ou de associação hipotéticas/supostas entre os construtos latentes ξ e η .

Em (1), Λ_x é a matriz das cargas fatoriais de ξ em x ; δ é o vetor dos erros de medida de x . Em (2), Λ_y é a matriz das cargas fatoriais de η em y ; ε é o vetor dos erros de medida de y . Em (3), B é a matriz dos coeficientes dos construtos latentes endógenos, η , no modelo estrutural; Γ é a matriz dos coeficientes dos construtos latentes exógenos, ξ , no modelo estrutural; ζ é o vetor dos erros do modelo estrutural (*disturbances*).

Esta formalização clássica dos SEM assume que os diferentes erros (ou resíduos) dos submodelos, de medida e estrutural, têm valor esperado nulo e seguem distribuição normal. Os erros de diferentes submodelos não podem estar correlacionados entre diferentes submodelos (embora possam estar correlacionados entre si) e, também, não podem estar correlacionados com os construtos latentes. Tem-se, então, que:

- i) δ e ξ são independentes;
- ii) ε e η são independentes;
- iii) ζ e ξ são independentes;
- iv) ζ , ε e ξ são mutuamente independentes;
- v) os valores esperados dos erros são nulos;
- vi) a matriz B tem zeros na diagonal, pois um construto dependente não é causa e efeito dele mesmo;
- vii) a matriz $(I - B)$, onde I é a matriz identidade, é não singular (i.e., admite inversa).

Relativamente às covariâncias entre os construtos latentes [1], partindo da equação estrutural em (3) obtém-se:

$$\eta = (I - B)^{-1} (\Gamma \zeta + \xi),$$

e considerando

$$\Phi = \text{Cov}(\zeta) \quad \text{e} \quad \Psi = \text{Cov}(\xi),$$

tem-se para a covariância entre os construtos latentes endógenos,

$$\Sigma_{\eta\eta} = (I - B)^{-1} (\Gamma \Phi \Gamma^T + \Psi) [(I - B)^{-1}]^T$$

e a covariância cruzada, entre os construtos exógenos e endógenos, vem dada por,

$$\Sigma_{\xi\eta} = \Phi \Gamma^T [(I - B)^{-1}]^T.$$

Quanto à covariância implícita nos submodelos de medida, a partir das equações de medida em (1) e (2) obtém-se:

$$\Sigma_{xx} = \Lambda_x \Phi (\Lambda_x)^T + \Theta_\delta;$$

$$\Sigma_{yy} = \Lambda_y \Sigma_{\eta\eta} (\Lambda_y)^T + \Theta_\varepsilon;$$

$$\Sigma_{xy} = \Lambda_x \Sigma_{\xi\eta} (\Lambda_y)^T.$$

onde, $\Theta_\delta = \text{Cov}(\delta)$ é a covariância dos erros de medida de x e $\Theta_\varepsilon = \text{Cov}(\varepsilon)$ é a covariância dos erros de medida de y .

A matriz de covariâncias implícita pelo modelo completo obtém-se considerando os resultados anteriores e representa-se por,

$$\Sigma(\theta) = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy}^T & \Sigma_{yy} \end{bmatrix} = \begin{bmatrix} \Lambda_x \Phi \Lambda_x^T + \Theta_\delta & \Lambda_x \Phi \Gamma^T [(I - B)^T]^{-1} \Lambda_y^T \\ \Lambda_y (I - B)^{-1} \Gamma \Phi \Lambda_x^T & \Lambda_y (I - B)^{-1} (\Gamma \Phi \Gamma^T + \Psi) [(I - B)^{-1}]^T \Lambda_y^T + \Theta_\varepsilon \end{bmatrix}.$$

Esta matriz de covariâncias teórica implícita pelo modelo, $\Sigma(\theta)$, é então função de um conjunto de parâmetros,

$$\theta = \{\Lambda_x, \Lambda_y, B, \Gamma, \Phi, \Psi, \Theta_\delta, \Theta_\varepsilon\}.$$

Na aplicação dos SEM podem ser considerados dois tipos de abordagens: (i) com base na covariância; (ii) com base na variância. Apesar da origem comum, ambas as abordagens diferem no modo como tratam os dados e, por isso, cada uma tem vantagens e desvantagens relativamente à outra.

2.1 Abordagem com Base na Covariância

Tendo por base a covariância (Covariance-Based, CB-SEM) os parâmetros do SEM são estimados de modo a minimizar as diferenças entre a matriz de covariâncias observada a partir dos dados (S) e a matriz de covariâncias $\Sigma(\theta)$, prevista pelo SEM, que se espera reproduza a matriz Σ das covariâncias das variáveis manifestas na população.

Formalmente, a estimação consiste em calcular o vetor de estimativas dos parâmetros (cargas fatoriais, coeficientes de caminhos, etc.), θ , que minimiza a função diferença (função de discrepância),

$$f = F [S - \Sigma(\theta)],$$

onde, o ajustamento é considerado perfeito se $f = 0$. O vetor de parâmetros θ pode ser estimado por métodos numéricos exatos ou quase-exatos de equações diferenciais, os quais minimizam a discrepância entre as matrizes mencionadas. Pelo facto da tarefa da estimação dos parâmetros por métodos exatos ser muito morosa os *softwares* utilizam habitualmente algoritmos iterativos que minimizam f . O critério de paragem do algoritmo (por exemplo, gradiente descendente ou Newton-Rapson) ocorre quando o critério de convergência definido previamente for cumprido. Os métodos de ajuste utilizados em SEM e as suas respetivas funções de discrepância variam com o tipo de pressupostos sobre a natureza da medida e em função das distribuições das variáveis [7].

A abordagem CB-SEM tem como objetivo estatístico testar teorias estabelecidas recorrendo a dados, onde o motor da análise é a teoria e os dados servem para a testar. O tradicional estimador de máxima verossimilhança pertence à família CB-SEM, com o forte pressuposto da normalidade multivariada dos dados [7].

Estimador de Máxima Verossimilhança

O estimador de máxima verossimilhança (Maximum Likelihood - ML) tem sido o mais utilizado em CB-SEM, o qual estima os parâmetros que maximizam a verossimilhança de observar a matriz S .

Assume-se que para uma amostra $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, onde cada vetor $\mathbf{x}_i \in \mathbb{R}^p$, se tem que $\mathbf{x}_i \sim N_p(\mu(\theta), \Sigma(\theta))$. A função de verossimilhança conjunta é obtida pelo produto das densidades individuais e pode ser especificada como,

$$L(\theta) = \prod_{i=1}^n (2\pi)^{-p/2} |\Sigma(\theta)|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x}_i - \mu(\theta))^T \Sigma(\theta)^{-1}(\mathbf{x}_i - \mu(\theta))\right).$$

Após o uso de logaritmos (função de log-verossimilhança) e de algumas transformações é possível expressar a função de discrepância ML por,

$$f_{ML} = \log|\Sigma(\theta)| + \text{tr}(S\Sigma(\theta)^{-1}) - \log|S| - p,$$

onde $|\cdot|$ e $\text{tr}(\cdot)$ representam as funções determinante e traço de uma matriz, respetivamente.

Tem-se $f_{ML} \geq 0$, com $f_{ML} = 0$ se e só se $S = \Sigma(\theta)$. O estimador ML produz estimativas centradas e consistentes (ou seja, à medida que a dimensão da amostra, n , aumenta, as estimativas obtidas aproximam-se do verdadeiro valor do parâmetro populacional, com variância mínima e distribuição normal). No entanto, estas propriedades só são válidas quando se verifica o pressuposto das variáveis manifestas terem distribuição normal multivariada ou quando a matriz de covariâncias segue distribuição de Wishart [7]. Tem-se, ainda, que o valor mínimo de F_{ML}^* conduz à estatística de teste,

$$T_{ML} = (n - 1) F_{ML}^*$$

que, sob normalidade multivariada e admitindo especificação correta do modelo, segue aproximadamente uma distribuição qui-quadrado com graus de liberdade dados por,

$$gl = \frac{p(p+1)}{2} - t,$$

onde t é o número de parâmetros livres em θ .

O estimador ML é geralmente considerado robusto à violação deste pressuposto de normalidade, desde que as características da forma da distribuição empírica das variáveis manifestas não se afastem demasiado das características da forma de uma distribuição normal (escalas ordinais devem ter pelo menos 5 categorias e distribuições aproximadamente simétricas, considerando-se com comportamento de variáveis intervalares [7]). É, no entanto, possível considerar versões robustas deste estimador de máxima verossimilhança [2].

Estimador Robusto de Máxima Verossimilhança

Em condições de não normalidade, as estimativas dos parâmetros ML são ainda consistentes, mas os erros-padrão e a estatística do teste do qui-quadrado são enviesados. Assim, a abordagem Robusta de Máxima Verossimilhança (Maximum Likelihood Robust - MLR) modifica a estatística do teste ML para,

$$T_{MLR} = (n - 1) c F_{MLR},$$

onde, c ajusta a não-normalidade e é designado por fator de correção de escala de Satorra-Bentler [8, 9, 10]. Estes ajustes fornecem erros-padrão e estatísticas de teste corrigidas, melhorando a fiabilidade da avaliação do modelo em caso de violações da normalidade [2].

Ambos os estimadores ML e MLR têm por base a matriz de covariâncias completa e têm vindo a ser aplicados em muitos estudos com variáveis manifestas qualitativas, medidas em escalas ordinais. Todavia, o estimador de Mínimos Quadrados Ponderados Diagonalmente é recomendado, especialmente quando se trabalha com dados ordinais (por exemplo, em escalas de Likert) ou quando o pressuposto de normalidade multivariada não é cumprido [11, 12].

Estimador de Mínimos Quadrados Ponderados Diagonalmente

O estimador de Mínimos Quadrados Ponderados Diagonalmente (Diagonally Weighted Least Squares – DWLS), proposto para problemas cujos dados disponíveis são ordinais [11, 12], utiliza a matriz de correlações policóricas. A função de discrepância do estimador DWLS minimiza a diferença entre as correlações observadas e as previstas e é geralmente escrita como,

$$f_{DWLS} = (\mathbf{s} - \sigma(\boldsymbol{\theta}))^T \mathbf{W}^{-1} (\mathbf{s} - \sigma(\boldsymbol{\theta})),$$

onde $\mathbf{S}$ é um vetor de momentos amostrais (correlações ou covariâncias), $\sigma(\boldsymbol{\theta})$ é um vetor de momentos implícitos no modelo e $\mathbf{W}$ é uma matriz de pesos diagonal.

O estimador DWLS centra-se na estimação da matriz de correlações policóricas – que capta as correlações entre variáveis latentes contínuas subjacentes às variáveis ordinais observadas – e, em seguida, utiliza apenas os elementos da diagonal principal dessa matriz para ponderar as diferenças entre as correlações observadas e as implícitas no SEM. Ao evitar a inversão de matrizes de ordem considerável e potencialmente mal condicionadas, o estimador DWLS aumenta a estabilidade e a robustez computacional. No entanto, o DWLS, também, apresenta vulnerabilidades, principalmente por ser sensível a incorretas especificações do modelo e a amostras de pequenas dimensões.

De notar que, o coeficiente de correlação policórica, proposto por Karl Pearson em 1900, é utilizado para estimar a correlação entre características que não são diretamente mensuráveis de forma quantitativa. Parte-se do pressuposto de que variáveis ordinais representam a discretização de variáveis latentes contínuas que seguem uma distribuição

normal bivariada. Nesse contexto, o coeficiente de correlação policórico fornece uma estimativa da associação linear entre duas variáveis latentes subjacentes, as quais são observadas indiretamente através de variáveis ordinais, como é o caso de itens em escalas de tipo Likert [7, 13].

2.2 Abordagem com Base na Variância

A abordagem com base na variância (Variance-Based, VB-SEM) maximiza a variância explicada dos construtos latentes endógenos, estimando relações parciais do modelo numa sequência iterativa de regressões de mínimos quadrados ordinários. Pertence a esta família VB-SEM, o estimador de mínimos quadrados parciais (Partial Least Squares - PLS), com avanços recentes na avaliação dos modelos PLS-SEM [14-23]. O potencial de interação destes modelos com múltiplas áreas de investigação é enorme como se pode constatar pelo elevado número de artigos que ilustram a sua aplicação nas mais diversas áreas do conhecimento: Ciências Económicas [24-26], Ciências Sociais e Humanas [20, 26, 27], Ciências da Saúde [28-30], Marketing [31], mas também em Engenharia de *Software* [32], Sistemas de Informação [14] e Turismo [33]. O desenvolvimento computacional potenciou o aumento considerável da complexidade destes modelos, existindo diversos programas para a análise PLS-SEM, uns comerciais (como por exemplo, o SmartPLS e o ADANCO) e outros de utilização livre (como por exemplo, os *packages* SemPLS, SEMinR e plspm disponíveis no R).

Estimador de Mínimos Quadrados Parciais

O estimador PLS enfatiza a previsão e, simultaneamente, relaxa as condições exigidas aos dados (é robusto à falta de normalidade multivariada e viável para amostras de pequena dimensão). O PLS-SEM pode ser usado para estimar modelos complexos (muitos indicadores, muitos construtos, muitas relações causais entre construtos), mesmo para casos em que o número de variáveis é superior ao número de observações da amostra. Adequa-se muito bem em situações onde a teoria, que sustenta as relações causais, ainda não está muito consolidada, podendo, por isso, ser utilizado de forma mais “exploratória”.

As duas formas possíveis de medir os construtos latentes (reflexivos e formativos), definem se os SEM são reflexivos, formativos ou mistos, o que tem obviamente consequências em termos do método de estimação a aplicar. Em PLS-SEM os construtos latentes são estimados como uma combinação linear (designados compósitos ponderados. *scores* dos construtos latentes, *proxy* ou *scores* compósitos) das suas variáveis manifestas ou blocos de indicadores (itens), independentemente de serem medidos como formativos ou reflexivos [34]. Posteriormente, estima os coeficientes de caminhos entre esses compósitos usando o método dos mínimos quadrados simples, evitando, deste modo, a necessidade da estrutura de covariância completa, $\Sigma(\theta)$, utilizada em CB-SEM.

Estimador Consistente de Mínimos Quadrados Parciais

O estimador consistente PLS (representado, abreviadamente, por PLSc) foi desenvolvido com base em modificações do algoritmo PLS original [15, 16] e é utilizado em modelos reflexivos, possuindo um bom comportamento com escalas ordinais. Esta abordagem fornece estimativas corrigidas do modelo (cargas fatoriais e coeficientes de caminho), mantendo todas as vantagens do método PLS tradicional (como, robustez a amostras pequenas e a dados não normais), sendo que foi desenvolvido para imitar o modelo de fatores comuns CB-SEM, que é utilizado para

estudos confirmatórios [14, 19]. O estimador PLSc modifica o estimador PLS para corrigir o enviesamento, em resultado de os compósitos serem representações imperfeitas (*proxies*) dos construtos latentes, fornecendo, deste modo, estimadores consistentes. A matriz de covariância implícita pelo modelo representa-se, neste caso, por Σ_{PLSc} e é semelhante à matriz $\Sigma(\theta)$, considerada em CB-SEM, mas foi construída com as estimativas baseadas no PLS e corrigida para a consistência.

Alguns dos novos recursos sobre os estimadores PLS e PLSc que têm surgido nos últimos anos incluem [14-19, 23]:

(i) detalhes sobre a distinção entre estimadores VB-SEM e CB-SEM, bem como os objetivos de investigação e os modelos que são favoráveis ao uso de PLS-SEM; (ii) aplicação destes estimadores com dados secundários; (iii) diretrizes para determinar dimensões mínimas de amostras, usando o método da raiz quadrada inversa; (iv) cobertura detalhada da confiabilidade da consistência interna usando métricas recentes e testes de inferência na avaliação da validade discriminante; (v) investigações sobre avaliação pelo método de reamostragem bootstrapping; (vi) análise do poder preditivo, fora da amostra, dos modelos recorrendo ao procedimento PLSpredict; (vii) métricas para comparações de modelos (por exemplo, o critério de informação Bayesiano).

A Raiz Quadrática Média Residual Padronizada (Standardized Root Mean Square Residual – SRMR) em PLSc-SEM, embora não seja uma função de discrepância *per se*, é comumente utilizada para avaliar a qualidade de ajustamento do modelo [35],

$$\text{SRMR} = \sqrt{\frac{2}{p(p+1)} \sum_{i < j} (r_{ij} - \hat{r}_{ij})^2},$$

onde, r_{ij} é a correlação observada (empírica) entre os indicadores i e j ; $\hat{r}_{ij}$ é a correlação reproduzida/estimada (pelo modelo ajustado) entre os indicadores i e j , e p é o número de indicadores (variáveis observadas).

O ajustamento é considerado perfeito se e somente se $\text{SRMR} = 0$, mas valores positivos de SRMR iguais ou inferiores ao valor de referência de 0,08 são indicadores de um bom ajustamento.

3 Exemplo de Aplicação dos Modelos de Equações Estruturais

Com o intuito de ilustrar a aplicação dos SEM considerou-se um estudo empírico ainda em desenvolvimento [36], na área da saúde mental, sobre o efeito do “stress” no “burnout”, nos “sleeping troubles” e nos “depressive symptoms” de trabalhadores de uma empresa industrial portuguesa.

Neste estudo foi considerada a versão portuguesa [37] do Copenhagen Psychosocial Questionnaire (COPSOQ), desenvolvida por Kristensen [36]. Foi utilizado o COPSOQ II, versão média com 76 questões relativas aos fatores de riscos psicossociais, à saúde e ao bem-estar dos trabalhadores. Na Tabela 1, apresentam-se apenas os construtos latentes, os itens (indicadores) e as questões relativas à saúde mental dos trabalhadores.

Tabela 1. Construtos latentes, itens e questões disponíveis no COPSOQ II.

Construto latente	Item	Questão no COPSOQ II
Stress	ST1	69. Com raiva?
	ST2	70. Ansioso?
Burnout	BO1	67. Fisicamente exausto?
	BO2	68. Emocionalmente exausto?
Sleeping troubles	SL1	65. Dificuldade em dormir?
	SL2	66. Acordou várias vezes durante a noite e depois não conseguiu dormir mais?
Depressive symptoms	DS1	71. Triste?
	DS2	72. Falta de interesse pelas coisas do dia-a-dia?

As variáveis observadas são qualitativas e medidas numa escala ordinal (itens do tipo Likert), com 5 categorias, codificadas de acordo com a Tabela 2. As respostas dos colaboradores da empresa refletem implicitamente a intensidade das suas percepções sobre questões relacionadas com a sua saúde e bem-estar.

Tabela 2. Categorias das questões disponíveis no COPSOQ II.

Categoria	Codificada
Nunca/quase nunca (de todo)	1
Raramente (uma pequena parte do tempo)	2
Por vezes (parte do tempo)	3
Frequentemente (uma grande parte do tempo)	4
Sempre (o tempo todo)	5

Amostra

O questionário COPSOQ II foi aplicado com o principal objetivo de avaliar os riscos psicossociais dos trabalhadores da empresa, com um número total de trabalhadores permanentes superior a 500 (população alvo), dos quais aproximadamente 65% são mulheres. Os responsáveis dos Departamentos de Recursos Humanos e da Segurança e Saúde no Trabalho (SST) da empresa apresentaram os objetivos do estudo aos trabalhadores e convidaram-nos a participar voluntariamente. Os trabalhadores foram informados de que não seria solicitada qualquer informação pessoal que os pudesse identificar e que poderiam interromper o preenchimento do questionário a qualquer momento. O questionário foi disponibilizado individualmente aos trabalhadores em papel, por todos os setores da empresa. Os trabalhadores podiam levar os questionários para os preencher fora da empresa, com mais tempo, e depois teriam de depositar em caixas disponibilizadas na empresa. A dimensão da amostra (não probabilística) disponível foi de 256 questionários validados e os resultados para algumas variáveis sociodemográficas dos trabalhadores foram: 74,6% são do sexo feminino (3 dados omissos); 55,1% têm idades entre 20 e 39 anos (4 dados omissos) e 47,5% possuem um nível de escolaridade superior ao 9.º ano (12 dados omissos).

Características dos Dados

Na Tabela 3 são apresentados os valores de assimetria (Skewness - Sk), curtose (Kurtosis - Ku) e os valores das razões críticas, obtidos com base nos respetivos erros-padrão (SE), Sk/SE_{Sk} e Ku/SE_{Ku} . Os resultados apresentados indicam que as distribuições empíricas da maioria dos itens são aproximadamente simétricas, dado que, para um nível de significância de 5%, os valores das razões críticas Sk/SE_{Sk} pertencem ao intervalo (-1,96; 1,96), pelo que a hipótese nula de que as distribuições são simétricas não é rejeitada. Apenas os itens DS1 e DS2 têm distribuições assimétricas positivas (i.e., os valores correspondentes das razões críticas são superiores a 1,96). Em termos de curtose, todas as distribuições dos itens, exceto uma com achatamento aproximadamente mesocúrtico – item BO1, com um valor de

Ku/SE_{Ku} a pertencer ao intervalo $(-1,96; 1,96)$ –, são aproximadamente platicúrticas, uma vez que os valores das razões críticas são inferiores a $-1,96$. Na última coluna da Tabela 3 constam os valores-p relativos ao teste de normalidade univariada de Shapiro Wilk, os quais permitem rejeitar a hipótese de normalidade (tem-se para todos os itens dos construtos latentes valores- $p < 0.001$).

Tabela 3. Construtos latentes, itens, dimensão da amostra (n), medidas de assimetria (Sk), curtose (Ku) e razões críticas; valores-p do teste de normalidade univariado de Shapiro-Wilk.

Construto latente	item	n	Skewness			Kurtosis			Shapiro-Wilk p
			Sk	SE_{Sk}	Sk/SE_{Sk}	Ku	SE_{Ku}	Ku/SE_{Ku}	
Stress	ST1	256	0,026	0,152	0,171	-0,983	0,303	-3,244	< 0,001
	ST2	256	0,065	0,152	0,428	-0,908	0,303	-2,997	< 0,001
Burnout	BO1	255	-0,196	0,153	-1,281	-0,533	0,304	-1,753	< 0,001
	BO2	256	-0,062	0,152	-0,408	-0,867	0,303	-2,861	< 0,001
Sleeping troubles	SL1	256	0,211	0,152	1,388	-0,986	0,303	-3,254	< 0,001
	SL2	256	0,007	0,152	0,046	-1,090	0,303	-3,597	< 0,001
Depressive symptoms	DS1	256	0,308	0,152	2,026	-0,756	0,303	-2,495	< 0,001
	DS2	253	0,396	0,153	2,588	-0,785	0,305	-2,574	< 0,001

Na Figura 1 os box-plots (diagramas de extremos-e-quartis) de alguns itens confirmam a assimetria das suas distribuições. Apenas são identificados 3 outliers moderados em dois itens (BO1 e DS1), que correspondem a uma percentagem relativamente inferior a 5% do total de observações da amostra (o que, segundo o estudo de simulação em [21], não compromete a qualidade dos resultados estatísticos obtidos).

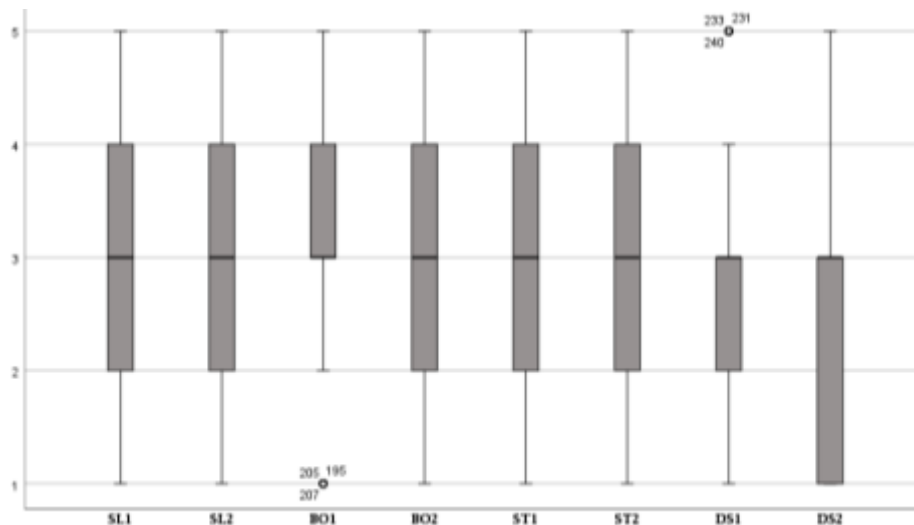


Figura 1. Box-plot de cada um dos itens, considerando o conjunto completo de dados.

O Modelo de Equações Estruturais Reflexivo Proposto

A importante fase de especificação do modelo, que consiste no “desenho” formal do modelo proposto para testar as hipóteses de investigação, tem por base o referencial teórico sobre o tema em estudo (perturbações associadas ao stress nos trabalhadores [39, 40]), bem como algum conhecimento empírico (*a priori*) de especialistas da empresa em Medicina do Trabalho, que avaliam a segurança e o bem-estar dos trabalhadores. Algumas das hipóteses de investigação a considerar são:

- H1: Existe um efeito positivo direto significativo do “stress” no “burnout” dos trabalhadores
- H2: Existe um efeito positivo direto significativo do “stress” nos “sleeping troubles” dos trabalhadores
- H3: Existe um efeito positivo direto significativo do “stress” nos “depressive symptoms” dos trabalhadores
- H4: Existe um efeito positivo direto significativo do “burnout” nos “sleeping troubles” dos trabalhadores
- H5: Existe um efeito positivo direto significativo do “burnout” nos “depressive symptoms” dos trabalhadores
- H6: Existe um efeito positivo direto significativo dos “sleeping troubles” nos “depressive symptoms” dos trabalhadores
- H7: Existe um efeito positivo indireto significativo do “stress”, através do mediador “burnout”, nos “depressive symptoms” dos trabalhadores
- H8: Existe um efeito positivo indireto significativo do “stress”, através do mediador “sleeping troubles”, nos “depressive symptoms” dos trabalhadores
- H9: Existe um efeito positivo indireto significativo do “burnout”, através do mediador dos “sleeping troubles”, nos “depressive symptoms” dos trabalhadores.

No modelo de trajetórias proposto da Figura 2, o construto latente exógeno é o “stress”, que se manifesta nos dois itens ST1 e ST2, e o construto endógeno “depressive symptoms” é considerado o alvo do estudo, manifestando-se, também, em dois itens, DS1 e DS2. A natureza dos itens é considerada reflexiva (i.e., a relação de causa e efeito ocorre no sentido dos construtos latentes para os itens) e é um modelo estrutural linear recursivo (i.e., não há relações de *feedback*), constituído apenas por construtos latentes de primeira ordem.

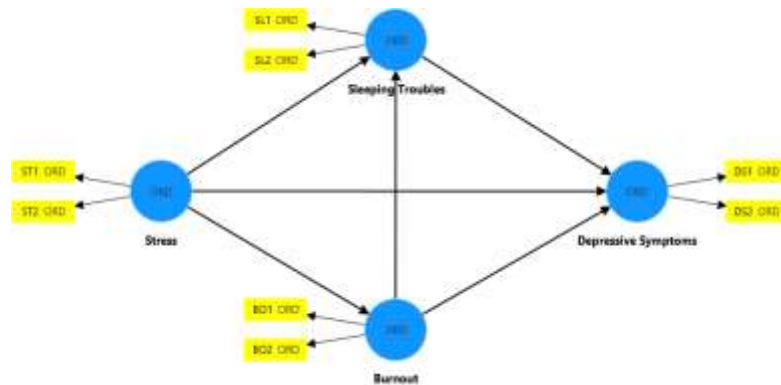


Figura 2. Modelo teórico causal proposto para os “depressive symptoms” de trabalhadores.

Para estimar o modelo teórico causal da Figura 2, com o estimador PLSc-SEM, calculou-se o número mínimo de observações (n_{min}) recomendado para representar a população subjacente utilizando o método da raiz quadrada inversa [18],

$$n_{min} > \left(\frac{2,486}{|p_{min}|} \right)^2. \quad (4)$$

Para um nível de significância de 5%, uma potência de teste de 80% e uma magnitude dos coeficientes de caminho mínimo ($|p_{min}|$) de 0,2, obteve-se o resultado usando (4), $n_{min} = 155$, aproximado por excesso às unidades. Com base na amostra disponível de $n = 256$ (bem acima do n_{min}) e recorrendo ao algoritmo PLSc, estimou-se o modelo da Figura 3, onde foi considerado: um número máximo de 3000 iterações, com o valor 10E-07 como critério de paragem e, ainda, o método de exclusão por pares (*pairwise deletion*) no tratamento de dados omissos. O algoritmo convergiu em apenas 4 iterações, o que constitui, obviamente, uma solução rápida e estável [17]. O modelo foi estimado com o *software* SmartPLS 4 [41], com cargas fatoriais (entre os construtos latentes e os itens no modelo de medida) e valores

dos coeficientes de caminho (entre os construtos latentes no modelo estrutural) padronizados e valores-p (entre parênteses).

O modelo estimado (Figura 3) disponibiliza alguma informação sobre a saúde e bem-estar dos trabalhadores da empresa. No modelo de medida, os itens ST1, BO2, SL2 e DS1 (Tabela 1) apresentam cargas fatoriais elevadas de aproximadamente 0,9 e, relativamente ao modelo estrutural, o “stress” tem fortes efeitos diretos positivos tanto nos “depressive symptoms” (0,903) como no “burnout” (0,879), bem como um efeito indireto nos “sleeping troubles” através da variável mediadora “burnout” ($0,688 \approx 0,879 \times 0,782$). Para os valores do coeficiente de determinação (R^2), que estão apenas dentro dos construtos latentes endógenos, tem-se $R^2 = 81,7\%$ para a quantidade de variância do construto “depressive symptoms” que é explicada pelos construtos preditores, neste caso apenas o “stress”. De notar, ainda, que algumas das hipóteses de investigação consideradas no modelo teórico proposto não estão presentes no modelo estimado, cujas equações estruturais são,

$$\begin{cases} \text{Depressive Symptoms} = 0.903 \text{ Stress} \\ \text{Burnout} = 0.879 \text{ Stress} \\ \text{Sleeping Troubles} = 0.782 \text{ Burnout} \end{cases}$$

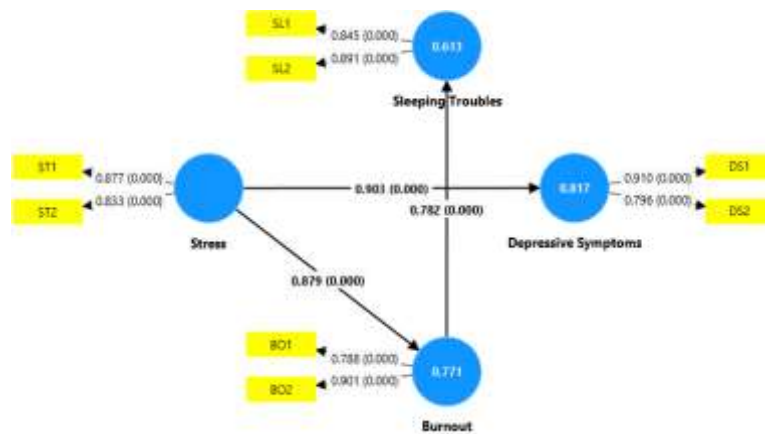


Figura 3. Modelo reflexivo estimado com o PLSc do SmartPLS 4.

Avaliação do Modelo Estimado

Na avaliação do SEM estimado (Figura 3) é necessário considerar a avaliação dos submodelos de medida e estrutural, seguindo as regras práticas disponíveis na literatura da especialidade que servem como linhas orientadoras gerais ([17-20], entre outros). Como o PLSc-SEM é um método de estimação não paramétrico para fazer "inferência estatística" sobre os diferentes parâmetros do modelo, recorre-se ao procedimento de reamostragem *bootstrap* (com o número recomendado na literatura de 10000 subamostras com reposição e de dimensão igual à amostra original, para garantir a estabilidade dos resultados obtidos [19, 20]). Assim, em algumas tabelas do Apêndice A estão disponíveis os valores obtidos com a "amostra original (O)" e os obtidos para a "média amostral (M)", recorrendo ao *bootstrap*. Este procedimento, também, permitiu obter para alguns parâmetros, intervalos de confiança a 95% (considerando o método Bias-Corrected and accelerated – BCa), estatísticas *t* e valores-p.

Avaliação do Submodelo de Medida (Outer Model)

Cargas fatoriais (outer loadings) – magnitude e significância: A avaliação do modelo de medida reflexivo inicia-se com um exame aos outer loadings da Figura 3, onde se pode observar que todos estes valores estão acima do valor de referência de 0,708 e são, ainda, estatisticamente significativos (valores- $p < 0,001$).

Fiabilidade usando o coeficiente ρ_A – consistência interna: As estimativas relativas à fiabilidade da consistência interna (Tabela A1, Apêndice A) foram obtidas com o coeficiente de fiabilidade exato (ou consistente) ρ_A , que se situa geralmente entre os valores do α de Cronbach (mais conservador) e os do ρ_C (mais liberal). Todas as estimativas estão acima de 0,8, indicando níveis elevados de fiabilidade, pois estão contidas no intervalo de referência (0,7 a 0,9), considerado de "satisfatório a bom". Além disso, na última coluna da Tabela A1, têm-se todos os valores- $p < 0,001$, relativos aos testes de significância deste parâmetro.

Average Variance Extracted (AVE) – validade convergente: Os valores da variância média extraída (AVE), que avalia a validade convergente de um construto latente, são todos superiores ao valor de referência de 0,50 (Tabela A2, Apêndice A), pelo que todos os construtos explicam pelo menos 50% da variância dos itens que os operacionalizam. Os valores- $p (< 0,001)$ da Tabela A2 confirmam a significância estatística desta medida.

Heterotrait-Monotrait ratio (HTMT) – validade discriminante: Para avaliar a extensão em que um construto latente é empiricamente distinto de outros foi utilizada a medida HTMT. Na Tabela A3 (Apêndice A), pode observar-se que apenas um valor de HTMT (Stress $\leftrightarrow$ Sintomas Depressivos) está ligeiramente acima de 0,90 (caso em que as variáveis latentes podem ser consideradas conceptualmente semelhantes). Todavia, o intervalo de confiança a 95% não contém o valor 1, o que corrobora a hipótese de que a validade discriminante está presente e que os construtos se podem considerar distintos.

Avaliação do Submodelo Estrutural (Inner Model)

Variance Inflation Factor (VIF) – colinearidade: O fator de inflação da variância (VIF) foi utilizado para medir o grau de multicolinearidade no modelo estrutural. Na Tabela A4 (Apêndice A), podemos observar que os valores de VIF são todos iguais a 1 e, por isso, inferiores ao valor de referência 3, o que é considerado ideal para não enviesar os resultados das regressões. Não existem, portanto, preocupações quanto à colinearidade entre os construtos analisados.

Standardized Root Mean Square Residual (SRMR) – ajustamento global do modelo: A raiz quadrada média residual padronizada (SRMR) permite avaliar a magnitude média das discrepâncias entre as correlações observadas e esperadas e tem sido usada como medida absoluta do critério de ajuste em PLSc-SEM. Uma vez $SRMR = 0,032$ (Tabela A5, Apêndice A) é inferior ao valor de referência de 0,08 (e, até, inferior a 0,05), pode considerar-se que o modelo estimado é adequado para explicar os “depressive symptoms” dos trabalhadores, com base no construto exógeno “stress”. Acresce que, para um nível de significância de 5%, o modelo estimado não é rejeitado, uma vez que os valores obtidos para esta medida de discrepância (Tabela A5) são inferiores ao valor do quantil 95%, da sua distribuição de referência.

Embora não existam todos os índices de ajuste global disponíveis em CB-SEM para os estimadores PLS e PLSc, é possível avaliar o modelo localmente, bem como a sua capacidade preditiva dentro e fora da amostra.

Coefficiente de determinação (R^2) – avaliação local do modelo: Os valores de R^2 (Figura 3 e Tabela A6, Apêndice A) permitem avaliar a proporção da variação prevista nos construtos endógenos “depressive symptoms” (81,2%), “burnout” (76,9%) e “sleeping troubles” (61,1%) a partir dos construtos preditores, considerando apenas os dados disponíveis na amostra. Dado o contexto específico da área de investigação deste estudo empírico (psicologia e saúde mental), pode considerar-se que os valores percentuais obtidos para o R^2 representam variâncias explicadas medianas (> 50%) e grandes (> 75%). Estes coeficientes de determinação são, ainda, estatisticamente significativos na população (valores- p < 0,001, na última coluna da Tabela A6).

Métrica Q^2 – relevância preditiva: O indicador Stone-Geisser Q^2 permite avaliar a precisão preditiva do modelo estimado, indicando o quão bem consegue prever resultados futuros, e não apenas explicar os dados existentes. Este procedimento consiste na remoção de pontos únicos na matriz de dados, que são posteriormente substituídos pela média, permitindo analisar também o poder preditivo fora da amostra. Na Tabela A7 (Apêndice A), os valores de Q^2 , dos itens associados aos três construtos endógenos, são todos superiores a zero, o que indica relevância preditiva. Como os valores Q^2 são superiores a 0,14 ou a 0,35 (valores de referência) a relevância preditiva é considerada média ou grande, respetivamente.

PLSpredict - poder preditivo do modelo fora da amostra: Com o algoritmo PLSpredict é possível determinar a capacidade preditiva do modelo relativamente a novos dados, considerando uma amostra de validação. Foi usada a estatística Raiz do Erro Quadrático Médio (Root Mean Squared Error – RMSE), que é particularmente útil quando os erros grandes são indesejáveis. Quando se comparam os erros de previsão (em termos de RMSE), na Tabela A7 do Apêndice A, verifica-se que o mesmo número de itens, do modelo estimado, tem valores de RMSE mais elevados do que o “naive benchmark” LM (Linear Model). Na coluna LM-PLSc o número de valores negativos e positivos são iguais, o que é indicativo do poder preditivo médio.

Bayesian Information Criterion (BIC) - comparação de modelos: O critério de informação Bayesiano (BIC) é uma métrica para comparar modelos, que apresenta uma elevada precisão na seleção de configurações alternativas de modelos, avaliando a previsão fora da amostra sem considerar uma amostra de validação. Comparando os valores apresentados na Tabela A8 (Apêndice A), é possível verificar que a variável endógena “depressive symptoms”, alvo do estudo, minimiza o valor do BIC (-424,378), indicando o modelo estimado como o melhor equilíbrio entre a qualidade do ajustamento e simplicidade.

Estimadores CB-SEM e VB-SEM

O SEM proposto na Figura 2 foi, também, estimado utilizando os estimadores ML, MLR e DWLS (da família CB-SEM), sendo que os respetivos algoritmos necessitaram entre 35 a 62 iterações para convergir. Importa salientar que, quando são utilizados estimadores desta família, os dados não são as observações relativas a cada indivíduo, mas sim as variâncias e covariâncias (ou, correlações) entre as variáveis manifestas, que se designam por elementos não redundantes da matriz de variâncias-covariâncias (ou, matriz de correlações).

Na Tabela 4, os valores obtidos com os estimadores CB-SEM, para os coeficientes de caminho, estão relativamente próximos dos obtidos com o estimador PLSc. O estimador DWLS fornece estimativas sempre ligeiramente superiores aos valores obtidos com os outros estimadores.

Tabela 4. Valores dos coeficientes de caminho obtidos com os estimadores ML, MLR e DWLS (CB-SEM) e com o estimador PLSc (VB-SEM).

Relação	Estimador CB-SEM			Estimador VB-SEM
	ML	MLR	DWLS	PLSc
Stress → Burnout	0,897	0,894	0,903	0,879
Burnout → Sleeping Troubles	0,767	0,763	0,803	0,782
Stress → Depressive Symptoms	0,908	0,909	0,918	0,903

Na Tabela 5, os valores obtidos com os estimadores CB-SEM, para os coeficientes de determinação (R^2), estão, também, relativamente próximos dos obtidos com o PLSc. Uma vez mais, os valores de R^2 obtidos com o DWLS são, também, ligeiramente superiores aos obtidos com os outros estimadores.

Tabela 5. Valores dos coeficientes de determinação (R^2) nos construtos latentes endógenos.

Variable latente	Estimador CB-SEM			Estimador VB-SEM
	ML	MLR	DWLS	PLSc
Burnout	0,805	0,799	0,815	0,771
Sleeping Troubles	0,588	0,582	0,644	0,613
Depressive Symptoms	0,824	0,827	0,843	0,817

Relativamente à avaliação da qualidade do ajustamento realizado, em CB-SEM têm vindo a ser propostos muitos índices. Alguns dos mais utilizados e disponíveis em diferentes *softwares* constam da Tabela 6. Todos estes índices são, no entanto, alvo de críticas severas ([7, 42], entre outros): o teste de qui-quadrado, porque testa a hipótese irrealista de que o ajustamento do modelo é perfeito e, também, porque é muito sensível à dimensão da amostra e à não normalidade dos dados; os outros índices, porque avaliam a qualidade do ajustamento global e podem não refletir eventuais problemas de ajustamento local dos parâmetros e, também, porque não possuem distribuição amostral conhecida (ou seja, não são verdadeiros testes estatísticos). Por este motivo, as decisões tomadas sobre a qualidade do ajustamento são realizadas com base em “regras-de-polegar”, que são obviamente muito discutíveis [43].

Tabela 6. Índices de qualidade do ajustamento global e valores de referência.

Estatística	Valor de referência	Tipo de índice
χ^2 e valor-p	Quanto menor, melhor; valor-p > 0,05	Absoluto
CFI - Comparative Fit Index TLI - Tucker-Lewis Index	> 0,95, ajustamento muito bom	Relativo
NFI - Normed Fit Index	> 0,90, ajustamento muito bom	
SRMR - Standardized Root Mean Square Residual	< 0,08, ajustamento muito bom;	Discrepância
Mean Square Residual	< 0,05, ajustamento excelente	
RMSEA - Root Mean Square Error Approximation	≤ 0,05, ajustamento muito bom	

Os três estimadores (ML, MLR e DWLS) conduziram, neste caso, a modelos com excelente qualidade de ajustamento. Em particular, para os índices da Tabela 6 os valores obtidos quando foi usado estimador DWLS foram: $\chi^2 = 18,1$ (valor-p = 0,320 > 0,05); CFI = 1,000 (> 0,95); TLI = 1,000 (> 0,95); NFI = 0,999 (> 0,90); SRMR = 0,028 (< 0,05) e RMSEA = 0,023 (< 0,05).

Os estimadores que foram utilizados neste estudo, que envolveu um modelo relativamente simples, conduziram a resultados relativamente semelhantes para os coeficientes de caminho e para os coeficientes de determinação (e, até, para os valores das cargas fatoriais, não apresentados neste relatório). Em particular, as interpretações dos coeficientes de caminho obtidos pelos estimadores CB-SEM e PLSc (utilizado com construtos reflexivos), podem ser considerados

semelhantes, especialmente em termos de sentido, magnitude e significado das relações entre os construtos (assumindo uma especificação apropriada do modelo e uma dimensão de amostra adequada). No entanto, estes estimadores resultam de diferentes “filosofias” de estimação e servem diferentes propósitos de modelação, especialmente em relação ao ajustamento do modelo, aos pressupostos e ao tratamento dos erros de medida. Estas diferenças e semelhanças, são mencionadas na literatura, sendo que os estimadores CB-SEM têm por objetivo estatístico testar teorias estabelecidas, i.e., análise confirmatória, e os estimadores PLS e PLSc são mais indicados em análises exploratórias além de serem, também, mais orientados para a previsão de resultados. Todavia, os avanços recentes na avaliação destes modelos estimados pelos versáteis estimadores PLS e PLSc (que podem ser utilizados numa variedade de modelos de medida, incluindo construtos reflexivos, formativos e mistos) têm levado alguns autores a considerá-los adequados tanto para investigação exploratória como confirmatória [14, 18].

4 Breve Descrição de Trabalhos Publicados ou em Fase de Publicação

Num trabalho, já aceite e a aguardar publicação [44], sobre os riscos psicossociais a que estão sujeitos os trabalhadores de uma empresa industrial portuguesa, o construto latente exógeno foi “quality of leadership” e o construto endógeno alvo do estudo foi o “stress” (Figura 4). No estudo de simulação desenvolvido, em que foram consideradas diferentes dimensões de amostras e diferentes números de categorias das variáveis ordinais, o estimador PLSc-SEM apresenta um desempenho, em termos de robustez das estimativas dos parâmetros, semelhante ou até ligeiramente superior ao dos estimadores CB-SEM (MLR e DWLS), em casos onde a dimensão da amostra é relativamente pequena ($n = 125$) e para um número de 4 ou 5 categorias. Apenas o estimador Bayesiano, também incluído no estudo, superou o desempenho de todos os estimadores (em termos de convergência e robustez das estimativas), em casos de amostras de dimensão muito pequena ($n = 50$) e, também, para casos com um número muito pequeno de categorias (apenas uma ou duas). Para amostras de maior dimensão ($n = 250$) e um número de pelos menos 3 categorias, nas variáveis ordinais, o estimador DWLS teve um desempenho ligeiramente superior aos outros estimadores em termos de robustez das estimativas dos parâmetros.

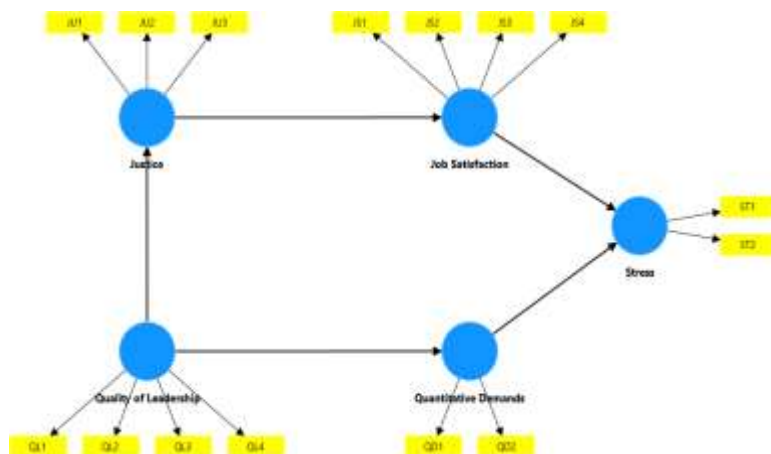


Figura 4. Modelo teórico causal proposto para o “stress” de trabalhadores.

Num estudo em desenvolvimento [45], estão a ser utilizados SEM, considerando o “student burnout” como construto de segunda ordem (Figura 5). Devido à complexidade do modelo reflexivo envolvido, à não normalidade dos dados e ao facto da dimensão da amostra não ser grande, está a ser utilizado o estimador PLSc com um processo de estimação específico para construtos de ordem superior. Neste caso, os estimadores CB-SEM não convergem ou produzem

soluções com mensagens de aviso sobre a possibilidade da matriz de covariâncias dos parâmetros estimados, não ser definida positiva (havendo valores próprios que, embora sejam aproximadamente zero, têm sinal negativo). A mensagem informa, ainda, que poderá ser um sintoma de modelo não-identificado (i.e., o número de parâmetros a estimar é superior à informação fornecida pelas variáveis manifestas – variâncias e covariâncias – sendo os graus de liberdade inferiores a zero).

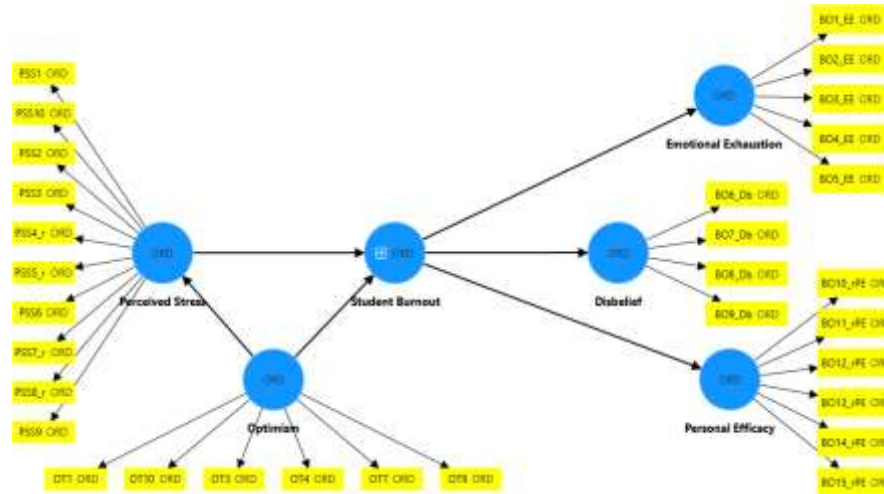


Figura 5. Modelo teórico causal proposto para o “student burnout”.

Num outro artigo, publicado recentemente [46], investigou-se de que forma o Sistema de Incentivos à Inovação (SII), do Quadro de Referência Estratégico Nacional (QREN 2014-2020), influenciou a internacionalização das PME portuguesas. Foram considerados quatro construtos latentes: "product innovation", "marketing innovation", "organizational innovation" e "working conditions", como sendo possíveis preditores do construto alvo: "internationalization of companies". Após a realização de uma Análise Tétrade Confirmatória (Confirmatory Tetrad Analysis - CTA) foi proposto um modelo misto (com construtos reflexivos e formativos), sendo depois estimado com o PLSc-SEM (Figura 6), pelo facto do modelo ser relativamente complexo, a amostra não ser grande (120 empresas) e os dados disponíveis (recolhidos através de questionário, com escala tipo Likert) não terem distribuição normal multivariada. Foram, ainda, discutidas as condições de regularidade para o PLSc na perspetiva da covariância semicontínua e os aspetos estocásticos, especialmente os relacionados com o uso de PLSc-SEM, onde podem ocorrer saltos nas covariâncias. O modelo estimado apresenta, entre todos os construtos, as “working conditions” com maior efeito direto positivo na “internationalization of companies”.

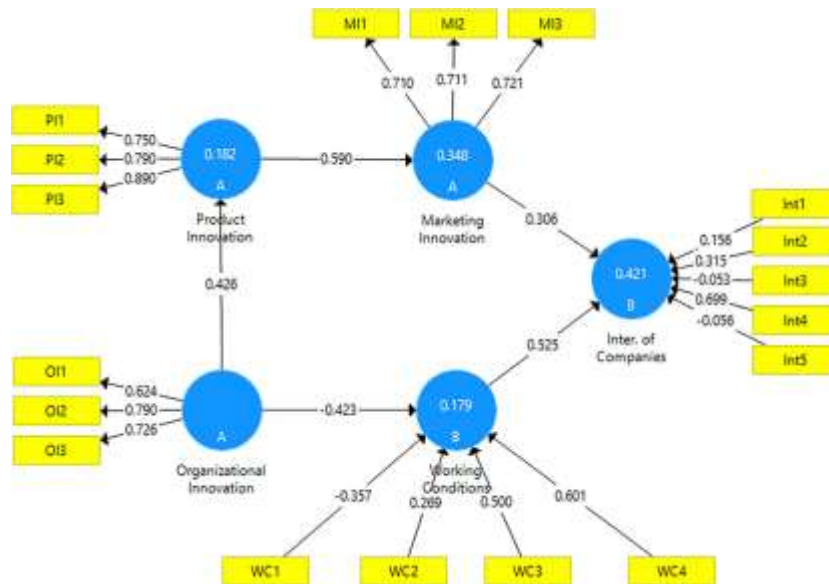


Figura 6. Modelo misto estimado com o PLSc-SEM do SmartPLS.

No trabalho [47] analisou-se como se relacionavam as três componentes do “burnout”: “emotional exhaustion”, “cognitive fatigue” e a “physical fatigue” nos inspetores do SEF do aeroporto de Lisboa (Figura 7). Os dados foram recolhidos utilizando a versão Portuguesa da Shirom Melamed Burnout Measure (SMBM), que conceptualiza o “burnout” nos três construtos latentes reflexivos mencionados, os quais são operacionalizados com base em indicadores medidos numa escala ordinal de 7 pontos. O PLSc-SEM foi utilizado para estimar um modelo proposto que considerava relações causais entre os 3 construtos. Foram avaliadas as violações da continuidade das covariâncias e foi apresentado um exemplo, de uma combinação de indicadores, onde a matriz de correlação dos construtos é definida positiva. O modelo estimado mostra que o maior efeito direto positivo se verifica entre a “cognitive fatigue” e a “physical fatigue”. Após estabelecida a invariância parcial no modelo de medida, foi desenvolvida uma análise multigrupos (por meio do teste de permutação não paramétrico), para testar se os coeficientes de trajetória diferem relativamente à variável categórica “anos de serviço”. Como esperado, o efeito direto positivo entre “emotional exhaustion” e “physical fatigue” é um pouco maior para trabalhadores com mais tempo de serviço, não sendo, no entanto, esta diferença estatisticamente significativa.

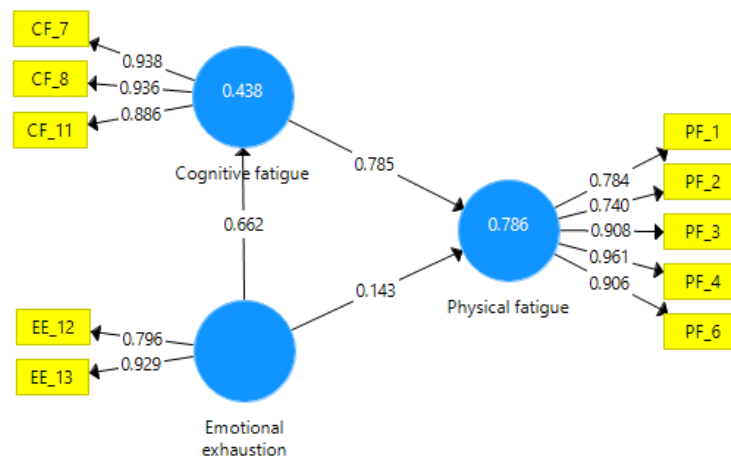


Figure 7. Modelo reflexivo estimado com o algoritmo PLSc-SEM do SmartPLS.

5 Considerações Finais

Em algumas das áreas de investigação em que os SEM são utilizados, os dados disponíveis foram recolhidos com base em inquéritos através de questionários, cujas questões correspondem a variáveis qualitativas medidas em escala ordinal (tipo Likert). No sumário deste seminário apresentaram-se e analisaram-se os resultados da aplicação dos SEM em alguns estudos empíricos com variáveis ordinais. Foram utilizados diferentes métodos de estimação com o objetivo de mostrar e discutir a sua adequação em cada caso. No primeiro estudo, na área da saúde mental e bem-estar de trabalhadores de uma empresa industrial portuguesa, procurou ilustrar-se o procedimento de modelação estatística recorrendo ao estimador PLSc, pertencente à família VB-SEM. Estimadores da família CB-SEM (estimador ML tradicional, uma versão robusta do ML e, ainda, o estimador DWLS) foram, também, utilizados neste estudo e os principais resultados (coeficientes de caminho e de determinação) encontram-se muito próximos dos obtidos com o estimador PLSc. Embora estes estimadores pertençam a famílias diferentes (pois recorrem a métodos de estimação diferentes) e possam servir diferentes propósitos de modelação, as interpretações dos coeficientes de caminho são consideradas semelhantes (em termos de sentido, magnitude e significado das relações entre os construtos). Posteriormente, foram apresentados resultados obtidos em estudos em desenvolvimento ou em fase de publicação, onde se verificaram as interessantes e conhecidas propriedades dos estimadores PLSc e Bayesiano (como a capacidade de lidar com: modelos complexos; dados não normais, amostras pequenas e, também, variáveis manifestas em escala ordinal com poucas categorias). Por fim, analisaram-se resultados de trabalhos publicados que procuraram mostrar a versatilidade do PLSc em trabalhar com modelos reflexivos e, também, mistos (construtos reflexivos e formativos), com variáveis ordinais, e em situações onde não existe continuidade das covariâncias. As principais conclusões resultantes destes estudos confirmam os estimadores PLSc e Bayesiano como uma opção viável na estimação de SEM com variáveis ordinais e mostrando-se como uma alternativa em casos em que o estimador DWLS, considerado mais adequado na presença deste tipo de variáveis, se revela incapaz de produzir uma solução (i.e., não converge ou não parece produzir uma solução válida).

6 Trabalhos Futuros

Na área de investigação em SEM existem várias questões em aberto, pelo que se apresenta uma que pode ser alvo de estudo num futuro próximo:

- . avaliar as condições de regularidade estocástica, na perspetiva da covariância semicontínua com ocorrência de saltos nas covariâncias, para o estimador DWLS e, eventualmente, estabelecer comparações com o estimador PLSc;

Encontram-se, ainda, em desenvolvimento os trabalhos:

Grilo, L. M., Grilo, M. L., Ghosh, S. e Mubayi, A. (2025). Assessment of workers' health and well-being through structural equation modeling with ordinary variables: a multi-group analysis by educational level;

Grilo, L. M. e Costa, C. (2025). Modeling the reflective higher-order construct 'student burnout' using the disjoint two-stage approach with PLSc-SEM;

Mendez, P., Paz, A., Paillacar, C., Caiceo, J., Grilo, L. M. e Stehlík, M. (2025). Direct Government Subsidy (AFD): 40 Years in Chilean Higher Education.

7 Referências

1. Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.
2. Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling* (4th ed.). New York: Guilford Press.
3. Byrne, B. M. (2016). *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming* (3rd ed.). New York: Routledge.
4. Hair, J. F., Black, W. C., Babin, B. J. e Anderson, R. E. (2019). *Multivariate Data Analysis* (8th ed.). Cengage.
5. Schumacker, R. E. e Lomax, R. G. (2016). *A Beginner's Guide to Structural Equation Modeling* (4th ed.). New York: Routledge.
6. Hair, J. F., Hult, G. T. M., Ringle, C. M. e Sarstedt, M. (2022). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (3rd ed.). Thousand Oaks, CA: Sage.
7. Marôco, J. (2014). *Análise de Equações Estruturais. Fundamentos teóricos, Software & Aplicações*, 2.^a Edição, ReportNumber, Pêro Pinheiro.
8. Satorra, A. e Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage
9. Bentler, P. M. e Dudgeon, P. (1996). Covariance structure analysis: Statistical practice, theory, and directions. *Annual Review of Psychology*, 47(1), 563–592. <https://doi.org/10.1146/annurev.psych.47.1.563>
10. Yuan, K.-H. e Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, 30(1), 165–200. <https://doi.org/10.1111/0081-1750.00078>
11. Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48(3), 936–949. <https://doi.org/10.3758/s13428-015-0619-7>
12. Rhemtulla, M., Brosseau-Liard, P. É. e Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17(3), 354–373. <https://doi.org/10.1037/a0029315>
13. Marôco, J. (2024). Factor Analysis of Ordinal Items Old Questions, Modern Solutions? *Stats*, 7(3) 984-1001; <https://doi.org/10.3390/stats7030060>
14. Benitez, J., Henseler, J., Castillo, A. e Schuberth, F. (2020). How to perform and report an impactful analysis using partial least squares: Guidelines for confirmatory and explanatory IS research, *Information & Management*, Volume 57, Issue 2, 103168, ISSN 0378-7206. <https://doi.org/10.1016/j.im.2019.05.003>
15. Dijkstra, T. K. e Henseler, J. (2015). Consistent partial least squares path modeling. *MIS Quarterly*, 39(2), 297-316. <https://doi.org/10.25300/MISQ/2015/39.2.02>
16. Dijkstra, T. K. e Henseler, J. (2015). Consistent and asymptotically normal PLS estimators for linear structural equations. *Computational Statistics & Data Analysis*, 81, 10–23. <https://doi.org/10.1016/j.csda.2014.07.008>
17. Hair, J. F., Hult, G. T. M., Ringle, C. M. e Sarstedt, M. (2017). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (2nd ed.). Thousand Oakes, CA: Sage.
18. Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P. e Ray, S. (2021). *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R. A Workbook*, Springer.
19. Hair, J. F., Risher, J. J., Sarstedt, M. e Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM, *European Business Review*, 31, 1: pp. 2-24. <https://doi.org/10.1108/EBR-11-2018-0203>

20. Ringle, C. M., Sarstedt, M., Mitchell, R. e Gudergan, S. P. (2020). Partial least squares structural equation modeling in HRM research, *The International Journal of Human Resource Management*, 31:12, 1617-1643. <https://doi.org/10.1080/09585192.2017.1416655>
21. Schamberger, T., Schuberth, F., Henseler, J. e Dijkstra, T. K. (2020). Robust partial least squares path modeling, *Behaviormetrika* 47, 307–334. <https://doi.org/10.1007/s41237-019-00088-2>.
22. Schuberth, F., Henseler, J. e Dijkstra, T. K. (2018). Partial least squares path modeling using ordinal categorical indicators, *Qual Quant*, 52: 9-35. <https://doi.org/10.1007/s11135-016-0401-7>
23. Shmueli, G., Sarstedt, M., Hair, J. F., Cheah, J., Ting, H., Vaithilingam, S. e Ringle, C. M. (2019). Predictive model assessment in PLS-SEM: guidelines for using PLSpredict, *European Journal of Marketing*, 53(11), 2322-2347. <https://doi.org/10.1108/EJM-02-2019-0189>
24. Peng, D. e Lai, F. (2012). Using partial least squares in Operations Management research: a practical guideline and summary of past research, *J. Oper. Manage.* 30 (6) 467–480. <https://doi.org/10.1016/j.jom.2012.06.002>
25. Avkiran, N. e Ringle, C. (2018). *Partial Least Squares Structural Equation Modeling*, Springer, Cham, Switzerland.
26. Ribau, C. P., Moreira, A. C. e Raposo, M. (2017). SMEs Innovation Capabilities and Export Performance: an Entrepreneurial Orientation View, *Journal of Business Economics and Management*, 920-934. <https://doi.org/10.3846/16111699.2017.1352534>
27. Grilo, L. M., Honrado, G. e Vitorino, R. (2022). Assessing global satisfaction of Portuguese expatriate workers using PLS-SEM. International Conference of Computational Methods in Sciences and Engineering (ICCMSE 2021), *AIP Conf. Proc.* 2611. <https://doi.org/10.1063/5.0119303>
28. Grilo, L. M., Grilo, H. L. e Martire, E. (2018). SEM using PLS Approach to Assess Workers Burnout State, *AIP Conf. Proc.* 2040, 110008-1–110008-5. <https://doi.org/10.1063/1.5079172>
29. Grilo, L. M., Mubayi, A., Dinkel, K., Amdouni, B., Ren, J. e Bhakta, M. (2019). Evaluation of Academic Burnout in College Students. Application of SEM with PLS approach, *AIP Conf. Proc.* 2186, 090008. <https://doi.org/10.1063/1.5138004>
30. Mutonyi, B. R., Slatten, T. e Lien, G. (2021). Fostering innovative behavior in health organizations: a PLS-SEM analysis of Norwegian hospital employees, *BMC Health Services Research*, 21:470. <https://doi.org/10.1186/s12913-021-06505-1>
31. Hair, J., Sarstedt, M., Ringle, C. e Mena, J. (2012). An assessment of the use of partial least squares structural equation modeling in Marketing research, *J. Acad. Mark. Sci.* 40 (3) 414–433. <https://doi.org/10.1007/s11747-011-0261-6>
32. Russo, D. e Stol, K.-J. (2021). PLS-SEM for Software Engineering Research: An Introduction and Survey, *ACM Comput. Surv.* 54, 4, article 78. <https://doi.org/10.1145/3447580>
33. Müller, T., Schuberth, F. e Henseler, J. (2018). PLS path modeling: a confirmatory approach to study tourism technology and tourist behavior, *J. Hosp. Tour. Technol.* 9 (3) 249–266. <https://doi.org/10.1108/JHTT-09-2017-0106>
34. Sanchez, G. (2013). *PLS Path Modeling with R*, Trowchez Editions. Berkeley.
35. Jöreskog, K. G. e Sörbom, D. (1981). *LISREL V: Analysis of linear structural relationships by maximum-likelihood and least-squares methods* (User's guide, version V). Chicago: International Educational Services.
36. Grilo, L. M., Grilo, M. L., Ghosh, S. e Mubayi, A. (2025). Assessment of workers' health and well-being through structural equation modeling with ordinary variables: a multi-group analysis by educational level (em desenvolvimento).
37. Silva, C. F. (2006). *Copenhagen Psychosocial Questionnaire (COPSOQ). Portuguese version* (FCT-MEC, Universidade de Aveiro, Análise Exacta). <https://www.copsoq-network.org/assets/Uploads/COPSOQ-Manual-Portugal2013.pdf>

38. Kristensen, T. S. (2021) A new tool for assessing psychosocial factors at work: The Copenhagen Psychosocial Questionnaire. *Scandinavian Journal of Public Health*, 19(1), 45-47. <https://doi.org/10.1177/1403494809354434>
39. Hsieh, H.-F.; Liu, Y., Hsu, H.-T., Ma, S.-C., Wang, H.-H. e Ko, C.-H. (2021). Relations between Stress and Depressive Symptoms in Psychiatric Nurses: The Mediating Effects of Sleep Quality and Occupational Burnout. *Int. J. Environ. Res. Public Health*, 18, 7327. <https://doi.org/10.3390/ijerph18147327>
40. Beer, L. T, Pienaar, J. e Rothmann Jr, S. (2014). Job burnout's relationship with sleep difficulties in the presence of control variables: a self-report study, *South African Journal of Psychology*, Vol. 44(4) 454–466. <https://doi.org/10.1177/0081246314538249>
41. Ringle, C. M., Wende, S., e Becker, J.-M. 2024. "SmartPLS 4." Bönningstedt: SmartPLS,
42. Barrett., P. (2007). Structural equation modelling: Adjudging model fit, *Personality and Individual Differences*, Volume 42, Issue 5, 815-824, ISSN 0191-8869. <https://doi.org/10.1016/j.paid.2006.09.018>
43. Nguyen, T. C. (2025). A commentary on “Magic number .95? Or was it .08? A refresher on SEM approximate fit indices thresholds for applied psychologists and management scholars.” *Research in Statistics*, 3(1). <https://doi.org/10.1080/27684520.2025.2544728>
44. Rojas, G. e Grilo, L. M. (2025). *Four Estimation Approaches for Structural Equation Modeling: an Empirical Application in Psychosocial Risk*. Springer Nature Switzerland AG. Trends in Mathematics. Advances in Mathematical Modeling in Science, Engineering, and Social Sciences: The Second International Workshop on Mathematics and Physical Sciences, Évora, Portugal, July 11-12, 2024. Editado por Mourad Bezzeghoud, Fernando Carapau, Feliz Minhós and Ashwin Vaidya (aceite e a aguardar publicação).
45. Grilo, L. M. e Costa, C. (2025). Modeling the reflective higher-order construct 'student burnout' using the disjoint two-stage approach with PLSc-SEM (em desenvolvimento).
46. Grilo, L. M., Pereira, E. J., Maidana, J. P. e Stehlík, M. (2024a). On stochastic aspects of impact modeling of the innovation incentive system and business internationalization: evidence from Portuguese SMEs. *Stochastic Analysis and Applications*, 42(1), 20–44. <https://doi.org/10.1080/07362994.2023.2166532>
47. Grilo, L. M., Braz, T. F., Grilo, H. L., Maidana, J. P. e Stehlík, M. (2024b). Modeling the facets of burnout in Lisbon airport border officers using PLSc-SEM estimator. *Stochastic Analysis and Applications*, 42(2), 414–434. <https://doi.org/10.1080/07362994.2023.2266492>

APÊNDICE A

Os *outputs* relativos à avaliação dos submodelos de medida e de estrutura apresentam-se neste apêndice, como foram produzidos pelo *software* SmartPLS 4 [39].

Resultados da avaliação ao modelo de medida (outer model)

Tabela A1: Composite reliability (ρ_A), mean, Stdev, t statistics, p-values.

Construct latent	ρ_A		Standard deviation (Stdev)	T statistics (O/Stdev)	p-values
	Original sample (O)	Sample mean (M)			
Burnout	0,841	0,843	0,026	32,976	0,000
Depressive Symptoms	0,851	0,854	0,026	32,129	0,000
Sleeping Troubles	0,861	0,862	0,022	38,746	0,000
Stress	0,846	0,847	0,026	32,554	0,000

Tabela A2: Average variance extracted (AVE), mean, StDev, t statistics, p-values.

Construct latent	AVE		Standard deviation (Stdev)	T statistics (O/Stdev)	p-values
	Original sample (O)	Sample mean (M)			
Burnout	0,717	0,718	0,037	19,272	0,000
Depressive Symptoms	0,731	0,733	0,038	19,256	0,000
Sleeping Troubles	0,754	0,755	0,034	21,983	0,000
Stress	0,731	0,733	0,040	18,310	0,000

Tabela A3: Heterotrait-monotrait ratio (HTMT), 95% bootstrap BCa confidence interval.

Relationship	HTMT		95% bootstrap BCa CI	
	Original sample (O)	Sample mean (M)	2.5%	97.5%
Depressive Symptoms ↔ Burnout	0,801	0,801	0,715	0,876
Sleeping Troubles ↔ Burnout	0,787	0,787	0,699	0,858
Sleeping Troubles ↔ Depressive Symptoms	0,659	0,659	0,540	0,755
Stress ↔ Burnout	0,876	0,877	0,803	0,943
Stress ↔ Depressive Symptoms	0,904	0,904	0,835	0,961
Stress ↔ Sleeping Troubles	0,705	0,705	0,603	0,788

Resultados da avaliação ao modelo estrutural (inner model)**Tabela A4.** Variance Inflation Factor (VIF) - collinearity analysis.

Relationship	VIF
Burnout → Sleeping Troubles	1,000
Stress → Burnout	1,000
Stress → Depressive Symptoms	1,000

Tabela A5: Standardized Root Mean Square Residual (SRMR), percentils.

	SRMR		Percentil	
	Original sample (O)	Sample mean (M)	95%	99%
Estimated model	0,032	0,026	0,040	0,048

Tabela A6: R-square, mean, Stdev, t statistics, p-values.

Construct latent	R ²		Standard deviation (Stdev)	T statistics (O/Stdev)	p-values
	Original sample (O)	Sample mean (M)			
Burnout	0,771	0,772	0,063	12,320	0,000
Depressive Symptoms	0,817	0,817	0,060	13,679	0,000
Sleeping Troubles	0,613	0,614	0,063	9,673	0,000

Tabela A7: PLSpredict MV summary.

Item	Q ² predict	PLSc-SEM RMSE	LM RMSE	LM-PLSc
BO1	0,347	0,908	0,911	0,003
BO2	0,565	0,833	0,829	- 0,004
DS1	0,562	0,830	0,836	0,005
DS2	0,429	0,952	0,954	0,002
SL1	0,274	1,086	1,080	- 0,007
SL2	0,328	1,060	1,045	- 0,015

Tabela A8. Model selection criteria, Bayesian information criterion (BIC).

Construct latent	BIC
Burnout	-367,480
Depressive Symptoms	-424,378
Sleeping Troubles	-232,619