Supplementary material to the paper "Gender and organizational commitment: evidence from a nationwide survey in Poland" by Vera A. Adamchik and Piotr Sedlak

CONTENTS:

Online Supplement # 1. Development of the survey questionnaire, validity and reliability of its instruments, data collection process, and representativeness of the survey data

Sedlak&Sedlak (S&S) is the oldest Polish HR consulting company operating in Poland since 1990 (https://sedlak.pl/eng/about_us). The company offers a wide array of services in the area of HRM and HRD, including a number of opinion surveys, such as surveys of salary satisfaction, job satisfaction, work-life balance, occupational stress, employee engagement, workplace bullying (mobbing), workplace discrimination, manager's competencies (https://badaniaHR.pl/english). S&S has its own methodology, ready-to-use questionnaires, and standardized and reliable research tools. In 2013, S&S created the badaniaHR.pl platform (in English - surveysHR.pl) for online research.

An S&S survey used in this study is the web-based Polish Job Satisfaction Survey (in Polish - Ogólnopolskie Badanie Satysfakcji z Pracy, OBSZP). The survey is conducted annually since 2011, provides nationally representative data on different facets of job satisfaction. As explained on the S&S website and during an interview with the S&S employees, the development and validation of the survey instruments included several stages (see Box S1 at the end of this online Supplement or https://badaniahr.pl/biblioteka/kwestionariusz-do-pomiaru-satysfakcji-z-pracy).

Face and *content validity* were assessed through expert panel reviews. Although various validated instruments for this type of survey can be found in the literature, S&S relies only on its own survey questions in order to avoid intellectual property rights violations and possible licensing problems. S&S is a for-profit consulting firm that charges clients for its products and services, and hence should have copyright on all content the company creates. So the survey questions have to be, to some extent, "home-made."

In 2011, a group of specialists from S&S drafted a pool of questions concerning theoretically important dimensions of the survey. Qualitative face and content validity were carried out by a panel of experts who evaluated relevance, appropriateness, accuracy, clarity, simplicity, grammar and vocabulary of the proposed contents. Quantitative face and content validity were assessed by calculating impact scores, content validity ratios, and content validity indices. Response process validity was achieved through cognitive pretesting, that is, a series of face-to-face interviews during which a handful of participants read each question and then explained their thought process in selecting an answer. A pilot survey using a full set of questions was conducted with more than 5,000 respondents to check readability and average time spent by each participant, to remove spelling/grammar errors, and to do some minor modifications if needed. The data obtained from the pilot survey were analyzed to determine *construct validity*, i.e., the extent to which survey measures accurately quantify underlying latent constructs. EFA and CFA were used to confirm the survey scales, thus enabling reliable measurement of employee attitudes and behaviors; the items that did not load onto a factor or heavily cross-loaded were dropped from the survey.

Since 2011, the main goal has been to optimize the length and content of the questionnaire, while maintaining its accuracy and reliability. The survey questionnaire undergoes periodic revisions and refinements every few years in order to include new questions and/or emerging topics, but the core of the survey remains the same to allow for comparison with the previous years' results. In 2022, new thematic areas (e.g., workplace stress) were added to the survey, and now the survey questionnaire consists of 80 questions, forming 13 main dimensions (or scales).

The aim of the OBSZP survey is to give a broad overview of a variety of employees' attitudes; hence, the survey includes only a limited number of questions per each survey dimension. It has been shown that longer questionnaires suffer from their length, which renders them cumbersome in studies that include measures of multiple phenomena as well as time-consuming to complete. Nowadays, shorter questionnaires are becoming more and more popular in order to increase their usability – in some cases, reducing the number of questions per each survey dimension to just one. Shorter surveys have higher completion rates and overall better data quality. For instance, Kost *et al.* (2018) tested ultrashort, short, and long surveys of 13, 25, and 72 questions, respectively. The authors revealed that "shorter surveys were reliable and produced higher response and completion rates then long surveys" (p. 31). In

addition to fatigue, boredom and loss of interest concerns, in the case of longer surveys, some participants pause a long survey and complete it later in different moods. Such temporal separation may allow different factors to intervene and thus bias the responses. Weisberg (2005) and Peytchev and Peytcheva (2017) documented the direct association between survey length and measurement error.

In order to ensure valid and reliable survey instruments, S&S rechecks validity and reliability for each survey. A high value (0.976) of the Kaiser-Meyer-Olkin measure of sampling adequacy indicates that the survey data are suited for factor analysis. The factor structure of the questionnaire is very well explained by the question correlation matrix. All the scales are characterized by high *reliability*, achieving an average of Cronbach's alpha of 0.85. *Criterion validity* involves comparing the survey instruments with other criteria which are believed to be representative of the underlying constructs. To assess *retrospective criterion validity*, S&S regularly compares current and past survey results, followed by an analysis of the correlation between them. *Concurrent criterion validity* is evaluated through a comparison of the survey results to the average in other nationwide surveys. The most recent comparison shown on the company's website reveals good agreement between the 2022 S&S survey and similar nationwide surveys. [Cited from https://badaniahr.pl/biblioteka/kwestionariusz-do-pomiaru-satysfakcji-z-pracy; also see Box S1 at the end of this online Supplement.]

The survey questionnaire is located on the S&S website. The invitations to the survey are distributed through email campaigns, text links connected to various articles published by S&S employees on the Internet, and through cooperation with partner companies, web pages, and paper magazines. After more than 30 years of presence in the Polish HR market, S&S has amassed a huge database of more than 120,000 working people who consented to data collection. The company also uses targeted Google ads for groups that are typically underrepresented in online surveys (e.g., people with only primary education).

Since the survey data are self-reported, it may raise the concern of common method bias. However, several studies of self-reports conclude that the assumption that self-reports result in upwardly biased correlations between self-reported variables is unfounded and overstated (Conway and Lance, 2010; Goffin and Gelatly, 2001; Spector, 2006). Furthermore, self-report measures appear to be the only way to measure attitudinal phenomena covered in the survey. To ensure data reliability and quality, S&S employs a number of quantitative and qualitative checks of the survey responses along with a sophisticated data cleaning procedure. It involves plausibility analysis (that is, checking for inconsistent and/or conflicting answers), examining questionnaire completion time, analysis of outliers, *etc*. On average, every year about 5% of all survey responses are excluded from the database.

It is worth noting that the survey does not constitute a random sample from a target population of all workers in Poland. The survey data may be defined as a 'voluntary response sample' because it only includes those people who voluntarily chose to participate in the survey. Compared to random samples, voluntary response samples are typically prone to different biases and, hence, are not generally suitable for making statistical inferences. We tested the 2020 survey (used in this study) against the official data of the Polish Central Statistical Office (in Polish -- Główny Urząd Statystyczny, GUS) and found quite a few strong similarities between the composition of workers in the survey and those in the official governmental statistics. The survey data are representative across a number of socio-demographic characteristics. For instance, in the 2020 survey, about 60% of respondents are male and about 40% are female. This corresponds to the official statistics reported by GUS: according to GUS, the share of women among the total number of employees in the working-age category (18-64 for men and 18-59 for women) fluctuates about 42% (GUS, 2014, p. 5). Further, in the 2020 survey, the share of private sector workers is 77% [79% by GUS (2021)]; the share of workers employed in manufacturing is 15%[17%], in construction 6% [8%], in banking 5% [2%]; the share of workers residing in the Mazowieckie voivoshhip is 25% [21%], in the Śląskie voivodship 12% [10%], in the Małopolskie voivodship 10% [9%], in the Dolnośląskie voivodship 10% [7%]. Hence, the numbers from the 2020 survey are not dissimilar from the official governmental statistics, and we can rule out the presence of a *strong* self-selection bias, which increases our confidence in the quality of the survey data, and hence in the quality of our data set in this study.

References:

Conway, J., & Lance, C. (2010). What reviewers should expect from authors regarding common method bias in organizational research. *Journal of Business and Psychology*, 25, 325-334.

Goffin, R., & Gellatly, I. (2001). A multi-rater assessment of organizational commitment: Are self-report measures biased? *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, 22(4), 437-451.

GUS. (2014). *Kobiety i mężczyźni na rynku pracy*. [*Women and men in the labor market*.]. GUS.

GUS. (2021). *Employment in national economy in 2020* [*Pracujący w gospodarce narodowej w 2020 r*]. GUS.

Kost, R., & da Rosa, J. (2018). Impact of survey length and compensation on validity, reliability, and sample characteristics for Ultrashort-, Short-, and Long-Research Participant Perception Surveys. *Journal of Clinical and Translational Science*, 2(1), 31-37.

Peytchev, A., & Peytcheva, E. (2017). Reduction of measurement error due to survey length: evaluation of the split questionnaire design approach. *Survey Research Methods*, 11(4), 361-368.

Spector, P. (2006). Method variance in organizational research: truth or urban legend? *Organizational Research Methods*, 9(2), 221-232.

Weisberg, H. (2009). *The total survey error approach: a guide to the new science of survey research*. University of Chicago Press.

BOX S1. An excerpt from the badaniaHR.pl website by Sedlak&Sedlak about the development of the survey questionnaire, validity and reliability of its instruments. Source: https://badaniahr.pl/biblioteka/kwestionariusz-do-pomiaru-satysfakcji-z-pracy Retrieved on August 12, 2023.

PROCEDURA KONSTRUKCJI NARZĘDZIA

Opisywany w artykule Kwestionariusz badania satysfakcji z pracy stanowi efekt kilkunastoletnich prac rozwojowych. Narzędzie wraz z kolejnymi badanami normalizacyjnymi oraz rozwojem światowej wiedzy w zakresie satysfakcji zawodowej było udoskonalane. Na jego aktualny kształt wpłynęły także doświadczenia wynikające z wieloletniej praktyki w realizacji badań komercyjnych.

Proces tworzenia kwestionariusza pomiaru opinii i postaw wobec miejsca pracy przebiegał etapowo. I tak na początku grupa specjalistów firmy Sedlak & Sedlak stworzyła pulę pytań dotyczących istotnych z perspektywy teoretycznej wymiarów, opisujących satysfakcję z pracy. Następnie z wykorzystaniem pełnego zestawu pytań przebadano ponad 5,000 osób. W trzecim etapie przeprowadzono analizy statystyczne, w których wyróżniono najlepiej sprawdzające się w praktyce pytania. Następnie spośród wyselekcjonowanych pytań utworzono skale badawcze, umożliwiające rzetelny pomiar ocen pracowniczych. W latach 2011-2020 prowadzano badania mające na celu optymalizację długości oraz składu treściowego kwestionariusza, przy jednoczesnym zachowaniu jego trafności i rzetelności. W 2022 r. przeprowadzono najnowszą edycję badań normalizacyjnych na próbie N=3,541 pracowników. Uzyskane wyniki pozwoliły na opracowanie najnowszych norm porównawczych dla poszczególnych skal kwestionariusza oraz wyłonienie dodatkowych obszarów tematycznych, które włączono do najnowszej wersji narzędzia.

Przed rozpoczęciem analizy przeprowadzono m.in. jakościową ocenę rzetelności wypełnienia ankiet. Usunięto m.in. rekordy, które budziły wątpliwości co do staranności wypełnienia (np. niepełne i niespójne) oraz ankiety wypełnione przez osoby niespełniające warunków zakwalifikowania do badania (np. bezrobotne). Po selekcji baza danych na której oparto dalsze analizy składała się z wyników dla 3,541 osób.

Skale pomiaru satysfakcji z pracy cechują się bardzo dobrymi własnościami pomiarowymi. Miara adekwatności doboru prób pytań do skal KMO=0.976 osiągnęła doskonały poziom dopasowania (w praktyce przyjmuje się skale, w przypadku których KMO jest większe od 0.5). Struktura czynnikowa kwestionariusza bardzo dobrze tłumaczy macierz korelacji pytań. Skale charakteryzują się wysoką rzetelnością (tabela 1).

Tabela 1. Współczynniki rzetelności poszczególnych skal kwestionariusza

skala	liczba pozycji	współczynnik Alfa Cronbacha
wynagrodzenia	8	0.88
współpraca z przełożonym	7	0.94
relacje ze współpracownikami	5	0.85
rozwój zawodowy	6	0.88
autonomia i empowerment	4	0.86
stres zawodowy	4	0.68
dopasowanie do pracy	4	0.87
wizerunek firmy jako pracodawcy	5	0.87
produkty i usługi firmy	3	0.80
komunikacja w firmie	8	0.88
warunki i narzędzia pracy	3	0.80
organizacja pracy	5	0.78
więź z firmą	4	0.89
praca zdalna	7	0.81
praca na stanowisku kierowniczym	7	0.83

Dodatkowo, kwestionariusz umożliwia ocenę 3 szerszych wymiarów funkcjonowania firmy. Są to indeksy atmosfery, efektywności oraz rozwoju organizacyjnego. Składają się one głównie z pytań podstawowych skal kwestionariusza. Oba indeksy również wykazują bardzo wysoką rzetelność (tabela 2.)

Tabela 2. Współczynniki rzetelności dodatkowych indeksów kwestionariusza

skala	liczba pozycji	współczynnik Alfa Cronbacha
atmosfera organizacyjna	9	0.87
efektywność organizacyjna	12	0.88
rozwój organizacyjny	7	0.87

PROCEDURA BADANIA I INTERPRETACJI WYNIKÓW

W tej części artykułu przedstawiono opis procedury badania i wskazówki do interpretacji wyników uzyskanych za pomocą kwestionariusza do pomiaru satysfakcji zawodowej.

OPIS KWESTIONARIUSZA I PROCEDURA BADANIA

Kwestionariusz badania satysfakcji z pracy składa się z 80 pytań, tworzących 13 wymiarów głównych:

- wynagrodzenia,
- współpraca z przełożonym,
- relacje ze współpracownikami,
- rozwój zawodowy,
- autonomia i empowerment,
- stres zawodowy,
- dopasowanie do pracy,
- wizerunek firmy jako pracodawcy,
- produkty i usługi firmy,
- komunikacja w firmie,
- warunki i narzędzia pracy,
- organizacja pracy,
- więź z firmą

oraz dwa dodatkowe:

- praca zdalna,
- praca na stanowisku kierowniczym.

Ponadto umożliwia on wyliczenie trzech indeksów:

- efektywność organizacyjna,
- atmosfera organizacyjna,
- rozwój organizacyjny.

Narzędzie może być stosowane w postaci ankiety online oraz tradycyjnej, papierowej. Osoba badana na skali pięciopunktowej określa, w jakim stopniu zgadza się z prezentowanymi jej twierdzeniami. Istnieje ponadto możliwość zaznaczenia odpowiedzi „nie wiem/nie dotyczy mnie to pytanie".
W wersji elektronicznej, w każdym momencie badany może cofnąć się do poprzedniej strony kwestionariusza, naciskając przycisk „wstecz". W górnej części ekranu wyświetlana jest skala, informująca o postępie w wypełnianiu kwestionariusza. Całość badania zajmuje około 20 minut.

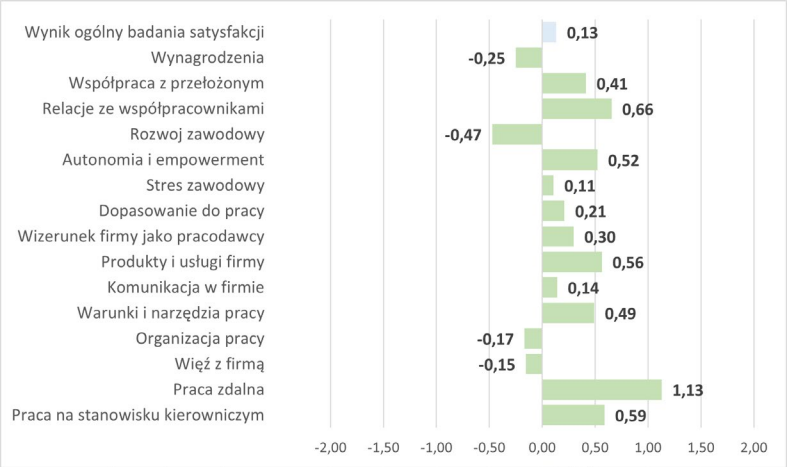
OBLICZANIE WYNIKÓW

Wynik w kwestionariuszu i jego poszczególnych skalach oblicza się poprzez zsumowanie odpowiedzi diagnostycznych (od -2 do 2 punktów za każdą odpowiedź diagnostyczną). Wyższy wynik liczbowy w danej skali wskazuje na większe nasilenie danego czynnika lub bardziej pozytywną ocenę omawianego zagadnienia.

INTERPRETACJA WYNIKÓW

Surowy wynik uzyskany w badaniu nie jest w pełni wystarczający do jego interpretacji. Nie dostarcza on bowiem informacji jak wysoki jest rezultat uzyskany przez pracowników firmy w porównaniu do wyników przeciętnych. Stąd wraz z rozwojem kwestionariusza opracowywaliśmy i aktualizowaliśmy kolejne normy, pozwalające odnieść poziom satysfakcji pracownika z poszczególnych obszarów pracy do wyników ogólnopolskich (wykres 1.).

Wykres 1. Ogólnopolskie normy porównawcze dla poszczególnych wymiarów kwestionariusza.



źródło: Ogólnopolskie Badanie Satysfakcji z Pracy 2022, Sedlak & Sedlak

Online Supplement # 2. Reliability and validity of the composite measures in the study

To evaluate reliability and validity of the composite measures (or survey scales, or survey dimensions), we use *correlation-based techniques*.

Reliability

Reliability refers to the consistency or reproducibility of scores. It could be measures across individual items on the construct (internal consistency), across different raters (inter-rater reliability), and over time (test-retest reliability).

Internal consistency measures how well the scores for individual items on the composite measure correlate with each other. In this study, internal consistency is assessed using Cronbach's alpha. It is worth noting that Cronbach's alpha relies upon the assumption of tau-equivalence, which is typically not justifiable. However, when simple average scores are used (as in our case), "one should report reliability using Cronbach's alpha" because simple average scores are "consistent with the tau-equivalent assumption of Cronbach's alpha" (Cheung, 2023, p. 16). As Table S2.1 shows, 3 alphas are in the 0.7-0.8 range, 7 alphas are in the 0.8-0.9 range, and 2 alphas are above 0.9. Although researchers widely use 0.7 as the threshold for adequate internal consistency, Nunnally (1978), Carmines and Zeller (1979), and Lance *et al*. (2006) point out that Cronbach's alpha of 0.7 indicates only modest internal consistency of the scale, and the value of 0.8 should be used as a standard. Overall, Cronbach's alphas reported in Table S2.1 indicate acceptable internal consistency of our composite measures.

Inter-rater reliability reflects the extent to which different raters are consistent in their judgments. Inter-rater reliability was assessed using two forms of the intra-class correlation coefficient (ICC): "two-way random effects, consistency, multiple raters" and "two-way random effects, absolute agreement, multiple raters" (Koo and Li, 2016, Table 3 on p. 157 and Figure 1 on p. 159). Unlike some other inter-rater reliability estimates (e.g., Cohen's kappa), which quantify reliability based on all-or-nothing agreement, ICC incorporates the magnitude of the disagreement among raters, with larger magnitude disagreements resulting in lower ICCs and vice versa. The values of ICC reported in Table S2.1 are all greater than 0.9 and indicate excellent inter-rater reliability (Koo and Li, 2016, p. 158).

Table S2.1. Cronbach's alpha and intra-class correlation coefficients

	Cronbach's alpha	ICC1	ICC2
COMMIT	0.873	0.998	0.996
PAY	0.888	0.994	0.987
FIT	0.871	0.999	0.997
AUTON	0.816	0.997	0.993
DEVEL	0.879	0.998	0.996
SUPER	0.923	0.999	0.996
COWOR	0.842	0.997	0.994
COMMU	0.790	0.995	0.990
ORGAN	0.762	0.999	0.998
MGT	0.910	0.996	0.991
CONSU	0.808	0.998	0.996
EMPL	0.721	0.997	0.994

Notes: COMMIT = Organizational commitment; PAY = Pay satisfaction; FIT = Job fit; AUTON = Job autonomy; DEVEL = Training and professional development; SUPER = Relationships with direct supervisors; COWOR = Relationships with coworkers; COMMU = Information and communication; ORGAN = Working conditions and work organization; MGT = Effectiveness of management in running a firm; CONSU = A firm's reputation in the consumer market; EMPL = A firm's reputation as employer in the labor market.
Source: Authors' own calculations.

Discriminant validity

The intent of testing for *discriminant validity* is to ensure that the underlying latent theoretical constructs are truly distinct from each other, that is, they are different and not measuring the same phenomena.

Discriminant validity is indicated by "low correlations between the measure of interest and other measures that are supposedly not measuring the same variable or concept" (Churchill, 1979, p. 70). A scale score correlation is the most commonly reported measure of validity (Rönkkö and Cho, 2022, Table 1, p. 8). As Cheung *et al*. (2023, p. 17) point out, "While discriminant validity is commonly defined as "two distinct constructs" and measured by the correlation between the two constructs, there is no generally accepted level of "distinctiveness" regarding the level of cross-construct correlation that establishes discriminant validity." Researchers have proposed different threshold values: 0.9 (John and Benet-Martínez, 2000), 0.85 (Garson, 2002, p. 195; Kenny, 2016), 0.8 (Rönkkö and Cho, 2022, Table 9), 0.75 (Voorhees *et al*., 2016, Table 4), 0.7 (Cheung *et al*, 2023, p. 17).

Out of the 66 pair-wise correlations in our sample, 2 are in the 0.3-0.4 range, 9 in the 0.4-0.5 range, 29 in the 0.5-0.6 range, 22 in the 0.6-0.7 range, and 4 in the 0.7-0.8 range (see Table S2.2). Hence, the predominant majority of the pairwise correlation coefficients is below the most conservative threshold of 0.7. For the four coefficients in the 0.7-0.8 range, we calculated the 90% confidence intervals using the bootstrap method with 10,000 iterations with replacement. The lower limits of these confidence intervals all fall within the 0.7-0.85 range, implying that there are no major discriminant validity concerns (Cheung *et al*., 2023, Table 2, p. 15).

Table S2.2. Pairwise correlations

	COMMIT	PAY	FIT	AUTON	DEVEL	SUPER	COWOR	COMMU	ORGAN	MGT	CONSU	EMPL
COMMIT	1											
PAY	0.637	1										
FIT	0.798	0.682	1									
AUTON	0.572	0.465	0.650	1								
DEVEL	0.698	0.603	0.775	0.611	1							
SUPER	0.613	0.548	0.683	0.569	0.660	1						
COWOR	0.456	0.370	0.517	0.474	0.446	0.556	1					
COMMU	0.564	0.473	0.593	0.511	0.570	0.563	0.504	1				
ORGAN	0.586	0.555	0.612	0.535	0.550	0.610	0.510	0.633	1			
MGT	0.676	0.608	0.697	0.547	0.661	0.681	0.547	0.756	0.719	1		
CONSU	0.534	0.447	0.512	0.438	0.528	0.473	0.392	0.513	0.522	0.637	1	
EMPL	0.651	0.596	0.629	0.511	0.603	0.588	0.438	0.567	0.595	0.682	0.599	1

Notes: COMMIT = Organizational commitment; PAY = Pay satisfaction; FIT = Job fit; AUTON = Job autonomy; DEVEL = Training and professional development; SUPER = Relationships with direct supervisors; COWOR = Relationships with coworkers; COMMU = Information and communication; ORGAN = Working conditions and work organization; MGT = Effectiveness of management in running a firm; CONSU = A firm's reputation in the consumer market; EMPL = A firm's reputation as employer in the labor market.
Source: Authors' own calculations.

In the early studies, the researchers proposed to test for discriminant validity by counting the number of times that the item (e.g., a survey question) correlates higher with items of other constructs than with items of its own theoretical construct. It was suggested that the count should be less than 50 percent of the potential comparisons in order to claim discriminant validity (Campbell and Fiske, 1959). An examination of our 65-question correlation matrix revealed that the number of violations was 29 percent of the 3,814 comparisons, that is, clearly below the recommended threshold.

One recently popularized approach of examining discriminant validity on the construct level is the assessment of the heterotrait-monotrait (HTMT) ratio of correlations. The original version of the HTMT relies on the arithmetic mean (Henseler *et al*., 2015), and the modified version of HTMT relies on the geometric mean (Roemer *et al*., 2021). Both versions have been heavily criticized (Rönkkö and Cho, 2022; Cheung *et al*., 2023). Cheung *et al*. (2023, pp. 11-12) proposed a correct formula of HTMT that uses inter-item covariances instead of inter-item correlations. The authors explain that "The corrected HTMT (…) is equivalent to the correlation between two composite scores formed by simple item averages, disattenuated with Cronbach's alpha of the two composite scores." (p. 12).

We calculated all three versions of HTMT for our sample. Table S2.3 shows the HTMT ratios based on covariances by Cheung *et al*. (2023). The other two versions of HTMT – based on correlations – produce a similar pattern and lead to the same conclusions. The recommended threshold value for HTMT is 0.85 (Clark and Watson, 1995; Henseler *et al*., 2015; Kline, 2011; Voorhees *et al*., 2016), while others propose a value of 0.9 (Gold *et al*., 2001; Teo *et al*., 2008). Out of the 66 HTMT ratios, 65 are below the threshold of 0.85, and 1 ratio (0.855) is only slightly above it; and hence discriminant validity is supported.

Table S2.3. HTMT (Cheung, 2023)

	COMMIT	PAY	FIT	AUTON	DEVEL	SUPER	COWOR	COMMU	ORGAN	MGT	CONSU	EMPL
COMMIT												
PAY	0.640											
FIT	0.822	0.677										
AUTON	0.614	0.495	0.697									
DEVEL	0.725	0.612	0.808	0.662								
SUPER	0.630	0.545	0.704	0.615	0.695							
COWOR	0.527	0.425	0.596	0.573	0.532	0.647						
COMMU	0.631	0.528	0.665	0.591	0.654	0.622	0.634					
ORGAN	0.643	0.590	0.654	0.621	0.611	0.641	0.619	0.736				
MGT	0.715	0.627	0.731	0.595	0.704	0.707	0.645	0.855	0.796			
CONSU	0.688	0.560	0.655	0.586	0.668	0.591	0.572	0.682	0.693	0.797		
EMPL	0.755	0.674	0.721	0.607	0.705	0.669	0.566	0.690	0.716	0.786	0.839	

Notes: COMMIT = Organizational commitment; PAY = Pay satisfaction; FIT = Job fit; AUTON = Job autonomy; DEVEL = Training and professional development; SUPER = Relationships with direct supervisors; COWOR = Relationships with coworkers; COMMU = Information and communication; ORGAN = Working conditions and work organization; MGT = Effectiveness of management in running a firm; CONSU = A firm's reputation in the consumer market; EMPL = A firm's reputation as employer in the labor market.
Source: Authors' own calculations.

Another method to test for discriminant validity is to assess the presence of multicollinearity in our regression model, that is, a situation in which two or more explanatory variables have a high correlation with one another. If

our composite measures were collinear (i.e., demonstrated insufficient discriminant validity), that would manifest itself in their high variance inflation factor (VIF). We calculated the VIFs for each of the 75 explanatory variables included in our regression model shown in Column (1) of Table 2 in the paper. For the 11 composite measures, the VIFs were in the acceptable range of 1-5, indicating a low correlation of those variables with other explanatory variables (all the 75 VIFs are shown in online Supplement # 4).

Overall, we conclude that our composite measures which intent to reflect the underlying latent constructs are reliable, are not measuring the same phenomena, and are truly distinct from each other.

References:

Campbell, D., & Fiske, D. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105.

Carmines, E., & Zeller, R. (1979). *Reliability and validity assessment*. Sage.

Cheung, G., Cooper-Thomas, H., Lau, R., & Wang, L. (2023). Reporting reliability, convergent and discriminant validity with structural equation modeling: a review and best-practice recommendations. *Asia Pacific Journal of Management*, forthcoming, 1-39. Published online January 30, 2023 https://doi.org/10.1007/s10490-023-09871-y

Churchill, G. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, 16(1), 64-73.

Clark, L., & Watson, D. (1995). Constructing validity: basic issues in objective scale development. *Psychological Assessment*, 7(3), 309-319.

Garson, G. (2002). Guide to writing empirical papers, theses, and dissertations. CRC Press.

Gold, A., Malhotra, A., & Segars, A. (2001). Knowledge management: an organizational capabilities perspective. *Journal of Management Information Systems*, 18(1), 185-214.

Henseler, J., Ringle, C., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43, 115-135.

John, O., & Benet-Martínez, V. (2000). Measurement: Reliability, construct validation, and scale construction. In H. Reis & C. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 339-369). Cambridge University Press.

Kenny, D. (2016). Multiple latent variable models: confirmatory factor analysis. Retrieved August 2023 from https://davidakenny.net/cm/mfactor.htm

Kline, R. (2011). *Principles and practice of structural equation modeling*. Guilford Press.

Koo, T., & Li, M. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155-163.

Lance, C., Butts, M., & Michels, L. (2006). The sources of four commonly reported cutoff criteria: what did they really say? *Organizational Research Methods*, 9(2), 202-220.

Nunnally, J. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

Roemer, E., Schuberth, F., & Henseler, J. (2021). HTMT2 - an improved criterion for assessing discriminant validity in structural equation modeling. *Industrial Management and Data Systems*, 121(12), 2637-2650.

Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, 25(1), 6-14.

Teo, T., Srivastava, S., & Jiang, L. (2008). Trust and electronic government success: an empirical study. *Journal of Management Information Systems*, 25(3), 99-132.

Voorhees, C., Brady, M., Calantone, R., & Ramirez, E. (2016). Discriminant validity testing in marketing: an analysis, causes for concern, and proposed remedies. *Journal of the Academy of Marketing Science*, 44, 119-134.

Online Supplement # 3. The 'affective organizational commitment' composite measure: relation to other questionnaires and CFA

The operational definition of affective organizational commitment used by S&S is based on the five questions available in the OBSZP survey. As a for-profit company, S&S relies only on its own survey questions in order to avoid intellectual property rights violations and possible licensing problems. Although the wording may be slightly different, the questions about affective organizational commitment in the OBSZP survey bear a close resemblance and/or are identical in meaning to the questions about affective organizational commitment that have been used in HRM research (van Rossenberg *et al*., 2022, Table 5, p. 10).

The first question – 'You feel an emotional connection with your firm' – reflects employees' feelings of psychological bond and belonging to the organization. Similar questions (e.g., I do not feel 'emotionally attached' to this organization; I do not feel a strong sense of belonging to my organization; I do not feel like 'part of the family' at my organization; *etc*.) were used by Allen and Meyer (1990), Bentein *et al*. (2005), Gellatly *et al*. (2006), Meyer *et al*. (1993), Moideenkutty *et al*. (2001), Rhoades *et al*. (2001), Saks (2006), Vandenberghe *et al*. (2004).

The second question – 'You are proud to work for this firm' – reflects employees' pride in organizational membership. Similar questions (e.g., I am proud to tell others that I am a part of this organization; I talk up this organization to my friends as a great organization to work for; *etc*.) were used by Marsden *et al*. (1993), Mowday *et al*. (1979), O'Reilly and Chatman (1986), Rayton (2006), Rhoades *et al*. (2001), Saks (2006).

The last three questions – 'You see your future with the firm for which you currently work'; 'You often think about changing jobs'; and 'If you had the opportunity to move to a competitor-firm, you would be willing to consider this option' – reflect employees' intent to remain with the organization. Similar questions (e.g., I would be very happy to spend the rest of my career with this organization; I would turn down another job for more pay in order to stay with this organization; I feel very little loyalty to this organization; There is not too much to be gained by sticking with this organization indefinitely; I often think about quitting this organization; I intend to search for a position with another employer within the next year; *etc*.) were used by Allen and Meyer (1990), Bentein *et al*. (2005), Gellatly *et al*. (2006), Marsden *et al*. (1993), Meyer *et al*. (1993), Mowday *et al*. (1979), Saks (2006), Vandenberghe *et al*. (2004).

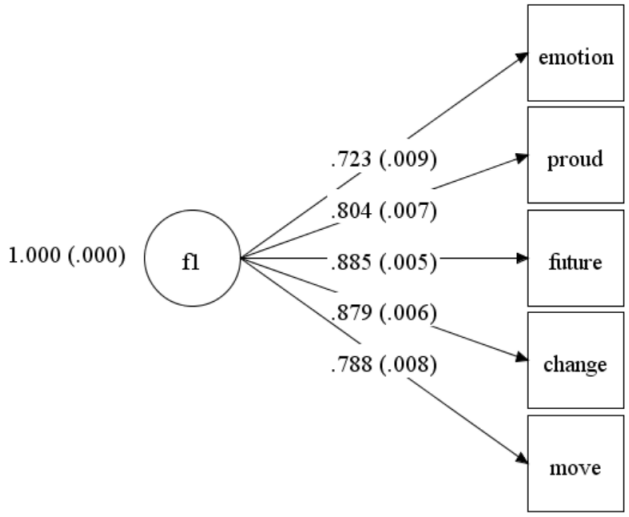
As stated in Section 3.1 of the paper, for this study we used the 12 predetermined dimensions (or survey scales, or composite measures) reflecting employees' subjective perceptions and satisfaction, which were formalized and are being used by S&S (see online Supplement # 1). Performing EFA or CFA on our data was beyond the scope of this study. First, as Harpe (2015) pointed out, "Scales that have been developed to be used as a group must be analyzed as a group, and only as a group. (…) Separating the items conceptually "breaks" the theoretical measurement properties of the aggregated scale as it was originally developed." (p. 840). Second, we wanted this study to be in line with S&S's approach, and the findings to be consistent and comparable with the company's past and future research.

However, at the request of one of the reviewers, we conducted CFA on the five questions comprising the 'employee affective organizational commitment' composite measure in order to confirm its unifactor structure and see how each variable (i.e., survey question) loads onto this specific latent construct. Mplus (Version 8) uses the WLSMV estimator with a probit link, as this is the most accurate estimator for testing models with ordinal data (e.g., Li, 2016). Figure S3.1 below shows standardized factor loadings and their standard errors.

The results of the CFA analysis show that the five variables (i.e., survey questions) are most closely associated with the 'employee affective organizational commitment' latent construct and that the one-factor measurement model adequately fits the data. All the five factor loadings are greater than 0.7, statistically significant, and have the expected (i.e., positive) sign. The model fit indices are as follows: CFI = 0.967, TLI = 0.933, SRMR = 0.043, RMSEA = n/a due to small degrees of freedom in our case (df = 5). Prior research suggests that RMSEA may not

be informative when fitting models with very small df. Shi *et al*. (2022, p. 179) find that "in comparison to RMSEA, population SRMR and CFI are less susceptible to the effects of df. In small df models, the sample SRMR and CFI could provide more useful information to differentiate models with various levels of misfit. (…) We recommend researchers use caution when interpreting RMSEA for models with small df and to rely more on SRMR and CFI." Kenny *et al*. (2015, p. 486) argue to not even compute the RMSEA for low df models.

Figure S3.1. Standardized factor loadings and their standard errors



Notes:
EMOTION = You feel an emotional connection with your company.
PROUD = You are proud to work for this company.
FUTURE = You see your future with the company for which you currently work.
CHANGE = You often think about changing jobs, [i.e., leaving your company and beginning to work for another]. (R)
MOVE = If you had the opportunity to move to a competitor-company, you would be willing to consider this option. (R)
Source: Authors' own calculations.

References:

Allen, N., & Meyer, J. (1990). The measurement and antecedents of affective, continuance and normative commitment to the organization. *Journal of Occupational Psychology*, 63(1), 1-18.

Bentein, K., Vandenberghe, C., Vandenberg, R., & Stinglhamber, F. (2005). The role of change in the relationship between commitment and turnover: a latent growth modeling approach. *Journal of Applied Psychology*, 90(3), 468-482.

Gellatly, I., Meyer, J., & Luchak, A. (2006). Combined effects of the three commitment components on focal and discretionary behaviors: a test of Meyer and Herscovitch's propositions. *Journal of Vocational Behavior*, 69(2), 331-345.

Harpe, S. (2015). How to analyze Likert and other rating scale data. *Currents in Pharmacy Teaching and Learning*, 7(6), 836-850.

Kenny, D., Kaniskan, B., & McCoach, D. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods and Research*, 44(3), 486-507.

Li, C. (2016). Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48, 936-949.

Marsden, P., Kalleberg, A., & Cook, C. (1993). Gender differences in organizational commitment: influences of work positions and family roles. *Work and Occupations*, 20(3), 368-390.

Meyer, J., Allen, N., & Smith, C. (1993). Commitment to organizations and occupations: extension and test of a three-component conceptualization. *Journal of Applied Psychology*, 78(4), 538-551.

Moideenkutty, U., Blau, G., Kumar, R., & Nalakath, A. (2001). Perceived organisational support as a mediator of the relationship of perceived situational factors to affective organisational commitment. *Applied Psychology*, 50(4), 615-634.

Mowday, R., Steers, R., & Porter, L. (1979). The measurement of organizational commitment. *Journal of Vocational Behavior*, 14(2), 224-247.

O'Reilly, C., & Chatman, J. (1986). Organizational commitment and psychological attachment: the effects of compliance, identification, and internalization on prosocial behavior. *Journal of Applied Psychology*, 71(3), 492-499.

Rayton, B. (2006). Examining the interconnection of job satisfaction and organizational commitment: an application of the bivariate probit model. *International Journal of Human Resource Management*, 17(1), 139-154.

Rhoades, L., Eisenberger, R., & Armeli, S. (2001). Affective commitment to the organization: the contribution of perceived organizational support. *Journal of Applied Psychology*, 86(5), 825-836.

Saks, A. (2006). Antecedents and consequences of employee engagement. *Journal of Managerial Psychology*, 21(7), 600-619.

Shi, D., DiStefano, C., Maydeu-Olivares, A., & Lee, T. (2022). Evaluating SEM model fit with small degrees of freedom. *Multivariate Behavioral Research*, 57(2-3), 179-207.

Vandenberghe, C., Bentein, K., & Stinglhamber, F. (2004). Affective commitment to the organization, supervisor, and work group: antecedents and outcomes. *Journal of Vocational Behavior*, 64(1), 47-71.

Van Rossenberg, Y., Cross, D., & Swart, J. (2022). An HRM perspective on workplace commitment: reconnecting in concept, measurement and methodology. *Human Resource Management Review*, 32(4), 100891.

Online Supplement # 4. Choice of multiple regression over SEM, OLS diagnostics & testing regression coefficients

Choice of multiple regression over SEM

For models with latent variables, there are two common approaches: structural equation modeling (SEM) and manifest regression analysis. While theoretically SEM is preferred to regression, in practice both SEM and regression have their advantages and limitations, and their proper application depends on the nature of the study (see, e.g., Gefen *et al*., 2000; Mai *et al*., 2018; Nunkoo and Ramkissoon, 2012). Gefen *et al*. (2000) analyzed the same dataset via three very different statistical techniques – covariance-based SEM, partial-least-squares-based SEM, and linear regression – and concluded that "there are situations where SEM tools are not called for" and "choosing an analysis method based correctly on the research objectives (…) is crucial" (p. 47). The reasons for selecting multiple regression over SEM in our study are as follows:

(a) Analyzing complex relationships *among all the variables in the model* was beyond the scope of this study. The study solely focuses on the association between one dependent variable (affective organizational commitment) and a set of independent variables (gender and other explanatory variables). As we do not need to solve the equation model with more than one dependent variable and the reciprocal (recursive) influence, there is no need to employ SEM.

(b) Regression analysis seems to be more appropriate, and SEM seems to be less suited for research which is exploratory in nature, as SEM is a confirmatory technique and a number of statistically valid models can be generated with the same data (Nunkoo and Ramkissoon, 2012, p. 796).

(c) SEM requires that the variables be measured on interval or ratio scale, which restricts its usefulness in studies using nominal/categorical variables, such as gender (as a matter of fact, the predominant majority of variables in our model are categorical) (Nunkoo and Ramkissoon, 2012, p. 797).

(d) SEM seems to be more stringent in parsimonious model testing and may suffer from testing large models with a large number of parameters and too many unexpected relationships (Nunkoo and Ramkissoon, 2012, p. 796).

(e) Finally, multiple regression analysis is "more appealing and user-friendly to researchers compared to those of SEM" because "the steps for conducting multiple regression analysis are straightforward and the measures of model fit are well established" (Nunkoo and Ramkissoon, 2012, p. 796).

Thus, we believe that multiple regression (which, by the way, is a special case of SEM) is an appropriate estimation tool in our case.

OLS assumptions, diagnostics & testing regression coefficients

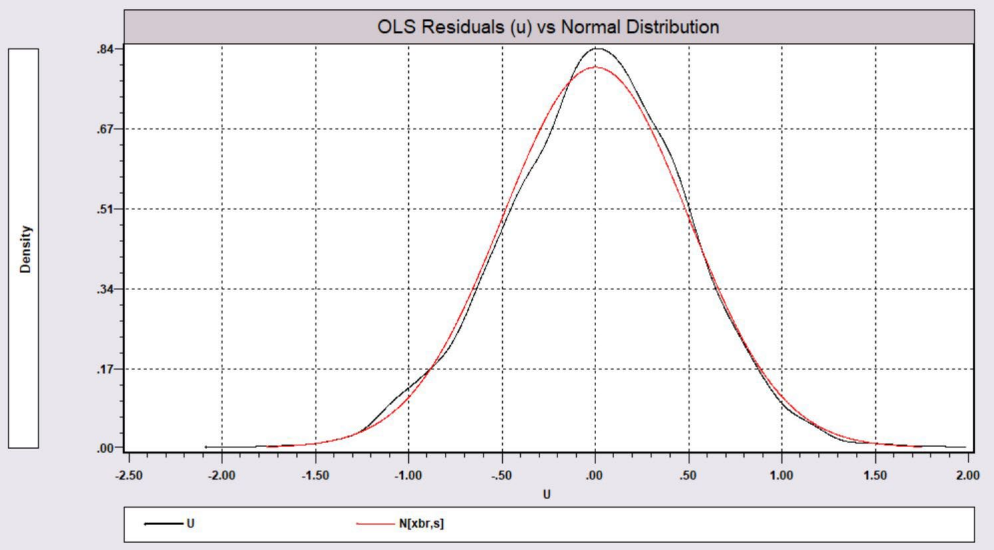
It is worth noting that many statisticians warn against using OLS for ordinal dependent variables, as the assumptions of OLS regression are likely to be violated. However, given certain requirements pertaining to the number of items in a composite scale (at least 5), slight skewness, large sample size, etc., it seems nevertheless possible to apply OLS to ordered data and recover true parameter estimates (Allen and Seaman, 2007; Carifio and Perla, 2007, 2008; Harpe, 2015; Mircioiu and Atkinson, 2017; Norman, 2010; Sullivan and Artino, 2013).

This online Supplement presents the tests for normality of error terms, multicollinearity, and heteroscedasticity. The OLS residuals are approximately normally distributed. The estimated variance inflation factor (VIF) scores are much lower than the typical threshold value of 5, indicating that multicollinearity does not adversely influence the regression results. The Breusch-Pagan test does not reject the null hypothesis of homoscedasticity. Although our data set does not exactly conform to OLS assumptions, the sample size is large enough and the deviations are slight enough, suggesting that the OLS method is appropriate.

Table S4.1. Normality of residuals (u) from the OLS regression of organizational commitment on the set of explanatory variables shown in Column (1) of Table II in the paper

```
------------------------------
Kernel Density Estimator for U
Kernel Function     =    Normal
Observations        =      2960
Points plotted      =      2960
Bandwidth           =   .089104
Statistics for abscissa values-
Mean                =   .000000
Standard Deviation  =   .489679
Skewness            =  -.101097
Kurtosis-3 (excess)=   .184128
Chi2 normality test=   .184904
Minimum             = -1.998807
Maximum             =  1.897276
Results matrix      =    KERNEL
------------------------------
```

```
----------------------------------------
Kolmogorov-Smirnov test of F(U         )
vs. Normal[     .00000,        .48968^2]
******* K-S test statistic =    .0380841
******* 95% critical value =    .0249973
******* 99% critical value =    .0299600
Normality hypothesis should be rejected.
----------------------------------------
```



Source: Authors' own calculations.

Breusch-Pagan test for heteroscedasticity for the OLS regression of organizational commitment on the set of explanatory variables shown in Column (1) of Table II in the paper

B-P test Chi squared [75] = 93.78241 p-value = .07012

The chi-square test statistic is 93.782 and the *p*-value for the test statistic is 7.0%, exceeding 5%. This indicates that we cannot reject the null hypothesis of homoscedasticity.

Source: Authors' own calculations.

Table S4.2. Variance inflation factor (VIF) & test for equality of coefficients for men and women

Variable	VIF	z-score (Ho: β woman – β man = 0)
Constant	0.000	-0.267
Woman	1.415	
ln(Age, years)	1.377	-0.303
Education College or higher	1.198	-0.611
Pay satisfaction	2.401	-0.208
Job fit	4.083	0.537
Job autonomy	2.190	0.396
Training and professional development	3.327	-0.316
Relationships with direct supervisors	2.665	-0.914
Relationships with coworkers	1.794	2.347
Information and communication	2.760	0.770
Working conditions and work organization	2.649	0.847
Effectiveness of management in running a firm	4.894	-0.412
A firm's reputation in the consumer market	2.087	1.293
A firm's reputation as employer in the labor market	2.650	-1.843
Private with majority Polish ownership	2.998	-0.309
Private with majority foreign ownership	3.183	0.131
Individual business activities	1.319	-1.188
Firm size 51-250 employees	1.791	0.927
Firm size 251-1500 employees	2.016	1.572
Firm size 1501 and more employees	2.151	1.911
Hired with a contract	1.194	0.530
Ln(Tenure at the current workplace, years)	1.383	0.710
Specialist	3.215	-0.459
Team leader	3.051	-0.823
Director and top manager	1.926	-1.003
Size of a city/town (4 dummies)	1.985; 1.826; 2.255; 3.320	-0.530; 0.286; -0.012; -0.614
Department (19 dummies)	1.854; 1.506; 1.166; 2.111; 3.257; 1.329; 1.179; 1.839; 1.401; 1.574; 1.548; 2.128; 1.078; 1.924; 1.666; 1.257; 1.121; 1.128; 1.226	-1.260; -1.095; -0.693; -0.423; -1.645; -2.229; -0.625; -1.112; -0.727; -1.694; -0.126; -0.578; -1.675; -1.573; 0.277; 0.037; -0.560; -1.586; -1.593
Industry (19 dummies)	2.112; 1.942; 2.030; 1.494; 2.408; 1.153; 1.702; 1.351; 1.691; 1.118; 1.733; 3.186; 1.111; 1.556; 2.955; 1.229; 1.214; 1.961; 1.500	0.393; -1.755; -0.469; 0.352; -0.171; 1.334; 0.613; 1.374; -0.795; 0.183; -0.464; -0.002; -0.152; -0.726; -0.676; -0.643; -0.448; -0.055; -0.298
Macroregion (6 dummies)	2.397; 1.958; 1.970; 1.659;	0.808; 1.255; 0.481; 0.402; 1.187;

Variable	VIF	z-score (Ho: β woman – β man = 0)
	1.512; 2.918	0.688
Covid1; Covid2	1.229; 1.678	0.436; -0.849

Notes: VIFs for all explanatory variables used in the regression model specified in Column (1) of Table II in the paper. Z-scores for the test of equality of coefficients for men and women shown in Columns (2) and (3) of Table II in the paper (see Paternoster *et al*., 1998, Formula 4 on p. 862).
Source: Authors' own calculations.

Table S4.3. Moderated regression

Variable	Coeff	StErr	t	p-value	95% CI lower	95% CI upper	R sq.	R-bar sq.
Pay satisfaction	-0.01278	0.01915	-0.67	0.5046	-0.05032	0.02476	0.74237	0.73557
Job fit	0.00443	0.01962	0.23	0.8213	-0.03403	0.04289	0.74233	0.73554
Job autonomy	0.00652	0.02057	0.32	0.7512	-0.03379	0.04683	0.74233	0.73554
Training and professional development	-0.00358	0.02103	-0.17	0.8649	-0.04481	0.03765	0.74233	0.73554
Relationships with direct supervisors	-0.00084	0.02103	-0.04	0.9683	-0.04205	0.04038	0.74233	0.73553
Relationships with coworkers	0.06502**	0.02764	2.35	0.0187	0.01085	0.11919	0.74282	0.73604
Information and communication	0.02385	0.02297	1.04	0.2992	-0.02117	0.06886	0.74242	0.73563
Working conditions and work organization	0.01751	0.02353	0.74	0.4569	-0.02862	0.06363	0.74237	0.73558
Effectiveness of management in running a firm	0.00188	0.02317	0.08	0.9353	-0.04353	0.04729	0.74233	0.73553
A firm's reputation in the consumer market	0.01758	0.02502	0.70	0.4824	-0.03146	0.06661	0.74237	0.73558
A firm's reputation as employer in the labor market	-0.02169	0.02259	-0.96	0.3370	-0.06596	0.02258	0.74241	0.73562
Education College or higher	-0.04160	0.05096	-0.82	0.4143	-0.14148	0.05828	0.74238	0.73559
Ln(Age)	-0.01021	0.08224	-0.12	0.9011	-0.17140	0.15097	0.74233	0.73553
Ln(Tenure)	0.02321	0.03627	0.64	0.5222	-0.04788	0.09431	0.74236	0.73557

Notes: The table shows the estimated coefficients on the the interaction terms between each of the variables and gender (with Woman=1, Man=0) added to the regression model specified in Column (1) of Table II in the paper, one interaction term at a time. For the model specification without an interaction term in Column (1) of Table II, R squared = 0.74233 R-bar squared = 0.73562.
Source: Authors' own calculations.

References:

Allen, I., & Seaman, C. (2007). Likert scales and data analyses. *Quality Progress*, 40(7), 64-65.

Carifio, J., & Perla, R. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of Social Sciences*,

3(3), 106-116.

Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education*, 42(12), 1150-1152.

Gefen, D., Straub, D., & Boudreau, M. (2000). Structural equation modeling and regression: guidelines for research practice. *Communications of the Association for Information Systems*, 4(1), 7.

Mai, Y., Zhang, Z., & Wen, Z. (2018). Comparing exploratory structural equation modeling and existing approaches for multiple regression with latent variables. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(5), 737-749.

Mircioiu, C., & Atkinson, J. (2017). A comparison of parametric and non-parametric methods applied to a Likert scale. *Pharmacy*, 5(2), 26.

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, 15, 625-632.

Nunkoo, R., & Ramkissoon, H. (2012). Structural equation modelling and regression analysis in tourism research. *Current Issues in Tourism*, 15(8), 777-802.

Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology*, 36(4), 859-866.

Sullivan, G., & Artino Jr, A. (2013). Analyzing and interpreting data from Likert-type scales. *Journal of Graduate Medical Education*, 5(4), 541-542.

Online Supplement # 5. The ordered probit model

There are a growing number of articles and statistical software that treat ordinal variables by default as ordinal, and not as continuous. For example, Mplus uses polychoric correlations and ordered probit to conduct EFA and CFA on ordinal data. Rayton (2006) applied a bivariate probit model and Saridakis *et al*. (2020) applied a two-stage probit model in their studies of organizational commitment.

Rayton, B. (2006). Examining the interconnection of job satisfaction and organizational commitment: an application of the bivariate probit model. *International Journal of Human Resource Management*, 17(1), 139-154.

Saridakis, G., Lai, Y., Muñoz Torres, R., & Gourlay, S. (2020). Exploring the relationship between job satisfaction and organizational commitment: an instrumental variable approach. *International Journal of Human Resource Management*, 31(13), 1739-1769.

Due to the ordinal nature of our dependent variable 'employee affective organizational commitment', we employ the ordered probit model. The level of commitment for individual i is symbolized by a latent variable (y_i^*). The basic ordered choice model is based on the latent regression (in what follows, we drop the subscript i to simplify the notations):

$$y^* = \beta' x + \varepsilon \qquad \text{(S5.1)}$$

where x is the vector of characteristics for individual i, β is the set of parameters to be estimated, and ε is the error term. If y^* were observed, β could be consistently estimated by OLS. However, for the type of data considered in this study, y^* is not observed. The observed counterpart to y^* is y, that is, a self-reported response on an ordinal 1,2,3,…,*J* = 1,2,3,4,5 scale. Hence, the observed discrete dependent variable y is related to y^* as follows:

$$\begin{aligned} y &= 1 \text{ if } y^* \le \mu_1 \\ &= 2 \text{ if } \mu_1, < y^* \le \mu_2 \\ &= 3 \text{ if } \mu_2, < y^* \le \mu_3 \\ &= 4 \text{ if } \mu_3, < y^* \le \mu_4 \\ &= 5 \text{ if } y^* > \mu_4 \end{aligned} \qquad \text{(S5.2)}$$

with μ_j, j = 1,2,…, *J-1* = 1,2,3,4 being additional parameters to be estimated. In ordered probit models, identification is achieved by fixing $\mu_1 = 0$ and assuming that the error term follows a standard normal distribution, that is, $\varepsilon_i \sim N(0,1)$. Estimates of β and μ are obtained by maximum likelihood.

A few remarks are in order. First, the estimated parameters β in the ordered probit model do not represent marginal effects. The estimated coefficients show the impacts of particular variables on the *index function*, and not on probability. A positive value of the estimated parameter implies that the entire distribution of y^* moves to the right when the value of the associated variable increases, and a negative value of the estimated parameter suggests the opposite (only for the highest and lowest ordered categories the sign and magnitude of β unequivocally predict the impact on the estimated probabilities). Second, since the ordered probit model is a nonlinear probability model, its estimated coefficients are always attenuated, but we can compare the sizes of coefficients within the same model because the attenuation bias is the same for all of them.

Table S5. Estimation results (ordered probit)

Variables	All	Men	Women
INDIVIDUAL DEMOGRAPHIC CHARACTERISTICS			
Gender Woman (Ref: Man)	0.152***		
Ln(Age, years)	0.718***	0.789***	0.629***
Education College or higher (Ref: High school or lower)	-0.195***	-0.194***	-0.246**

Variables	All	Men	Women
SUBJECTIVE PERCEPTIONS AND SATISFACTION			
Pay satisfaction	0.253***	0.253***	0.248***
Job fit	0.967***	1.006***	0.984***
Job autonomy	0.028	0.009	0.040
Training and professional development	0.236***	0.249***	0.247***
Relationships with direct supervisors	0.045	0.085*	-0.022
Relationships with coworkers	0.058	0.006	0.125**
Information and communication	-0.167	-0.208	-0.156
Working conditions and work organization	0.108***	0.080	0.177**
Effectiveness of management in running a firm	0.243***	0.233***	0.271***
A firm's reputation in the consumer market	0.133***	0.106**	0.191***
A firm's reputation as employer in the labor market	0.275***	0.303***	0.222***
INDIVIDUAL JOB AND WORKPLACE CHARACTERISTICS			
Firm ownership (Ref: Public sector)			
Private with majority Polish ownership	-0.116	-0.137	-0.088
Private with majority foreign ownership	-0.283***	-0.352***	-0.213*
Individual business activities	0.124	0.210	0.039
Firm size (Ref: < 50 employees)			
51-250 employees	-0.069	-0.148*	0.023
251-1500 employees	0.013	-0.100	0.189*
1501 and more employees	-0.112	-0.246***	0.097
Type of employment: Hired with a contract (Ref: Other types of employment)	0.213***	0.233***	0.209
Ln(Tenure at the current workplace, years)	0.285***	0.248***	0.405***
Hierarchical position (Ref: Rank-and-file)			
Specialist	-0.025	0.015	-0.063
Team leader	0.067	0.105	-0.003
Director and top manager	0.149	0.268*	-0.050
Size of a city/town (Ref: < 20,000 residents)			
21,000 – 100,000	-0.027	0.031	-0.123
101,000 – 200,000	-0.068	-0.127	-0.006
201,000 – 500,000	-0.171*	-0.181	-0.183
501,000 and more	-0.118	-0.112	-0.167
Department (19 dummies)	yes	yes	yes
Industry (19 dummies)	yes	yes	yes
Macroregion (6 dummies)	yes	yes	yes
COVID-RELATED QUESTIONS			
Covid's impact on work life	0.012	-0.012	0.038
Employer is taking proper care of employees	0.024	0.052	-0.005
Threshold parameters			
Mu2	1.978***	2.024***	2.004***
Mu3	3.895***	3.966***	3.958***
Mu4	6.044***	6.221***	6.018***
Log-likelihood	-2486.8	-1534.8	-914.8
McFadden pseudo R-sq.	0.408	0.413	0.420
N variables (including the intercept and threshold parameters)	79	78	78
N observations	2960	1864	1096
Wald test of no difference in the two coefficient vectors (without the three threshold coefficients)		78.769 (p-value = 0.361)	

Notes: ***, **, * Significance at 1%, 5%, 10% level. Robust standard errors. See Section 3 of the paper for the description of variables used in estimations.
Source: Authors' own calculations.