

Emotional Labour as an Unpriced Cost of Human Capital in Seasonal Tourism Work: An Instrument-Screening Pilot from the Podhale Region

Kamil Opar¹, Marcin Surówka²

¹University of Applied Sciences in Nowy Targ,
Poland

²Krakow University of Economics,
Poland

Abstract— Frontline service in tourism is a paradigm case of emotional labour, an input the wage does not price separately yet one whose hidden costs — presenteeism, exhaustion, somatic symptoms — plausibly depreciate human capital. This paper reports the instrument-screening stage of that inquiry, not a test of the economic argument. A cross-sectional pilot surveyed sixty-four frontline workers in the Podhale region on five-point scales. Reliability is reported with statistics appropriate to scale length. The exhaustion, psychological-safety and self-regulation scales were adequately consistent (Spearman-Brown 0.79, 0.79, 0.81); the somatic-symptom and presenteeism scales were not (0.32 and 0.02), and the five-item surface-acting scale reached an alpha of only 0.30. That scale also failed a validity check: its association with exhaustion was negative, contrary to meta-analytic evidence, and item analysis locates the reversal in a single item read normatively rather than in restriction of range. The burden variables were reported at high levels and pay adequacy at low levels, but the burden distributions are ceiling-bound, so their levels cannot be interpreted. Within a family of twenty-three tests one association survives familywise correction, and it is a coherence check, not a finding. The yield is a specification of the requirements for the main study.

Keywords— emotional labour, surface acting, presenteeism, human capital, seasonal tourism employment, measurement validity

I. INTRODUCTION

In tourism and hospitality the service is produced in direct contact with the guest, and the demeanour of the frontline worker is part of the product. Meeting the guest with composure irrespective of what the worker feels is a job requirement rather than a disposition, and it is the original setting in which emotional labour was described (Hochschild, 1983): the

regulation of felt emotion to satisfy organizationally mandated display rules (Grandey and Sayre, 2019). It is performed for remuneration, yet the remuneration is set for the visible tasks of the role and not for the regulation itself.

Emotional labour is thus an input the wage does not price separately and conventional records of hours and output do not capture, yet whose costs are real. The worker bears them as diminished well-being and, over time, as the cumulative burden of sustained regulation (McEwen, 1998); the firm bears them as presenteeism, that is presence at work with reduced capacity (Johns, 2010), as degraded service quality and as turnover. Read through the economics of human capital, this amounts to a depreciation of the resource on which a service firm depends (Becker, 1964). Seasonal mountain tourism sharpens the problem: demand is compressed into a short peak during which staffing is thin and income time-bound, conditions under which working through fatigue and failing to return the following season are both documented (Radlińska and Gardziejewska, 2022; Tokarz-Kocik et al., 2023). Two features of this setting are less examined than the general association between emotional labour and exhaustion: the framing of the accumulated burden as human-capital depreciation with economic consequence, and the seasonal micro-enterprise labour market of mountain tourism.

The contribution claimed here is a measurement contribution: the paper is the instrument-development stage of a longer inquiry rather than a test of the economic argument that motivates it. That argument is developed in Section 2 because it determines what has to be measured; it is not adjudicated by the data in Section 4. The constructs at its centre are precisely those the pilot failed to measure adequately; the burden



distributions on which any descriptive claim would rest are ceiling-bound; and the design is cross-sectional and wholly self-reported. Reporting the failures in full is therefore the substance of the contribution rather than an admission attached to it.

Three core hypotheses are examined. H1: surface acting is positively associated with emotional exhaustion and somatic symptoms. H2: surface acting is positively associated with presenteeism. H3: emotional exhaustion is positively associated with the intention to leave and negatively with the intention to return in the following season. Three relations are exploratory: group differences by contact intensity and contract type (E1); the association between perceived pay adequacy and the intensity of surface acting (E2); and the association of psychological safety and self-regulation with exhaustion (E3). In a pilot these function as directional expectations against which the instrument can be checked, since a scale yielding an association contrary to a robust meta-analytic finding is thereby put in doubt.

II. THEORETICAL FRAMEWORK

1) Emotional labour and its cost

Emotional labour describes work in which the management of feeling is sold for a wage and displayed as part of the role (Hochschild, 1983). The construct was refined into two strategies: in deep acting the worker attempts to feel the required emotion, so display and feeling converge; in surface acting the worker displays an emotion not felt, so they diverge (Grandey, 2000). Surface acting, and the dissonance it entails, are consistently associated with exhaustion and reduced well-being (Grandey, 2003; Hülsheger and Schewe, 2011).

Displaying an emotion that is not felt requires regulating a spontaneous response. In Gross's process model this is response-focused regulation, that is suppression applied after the emotion has been generated, and it is the form most consistently associated with physiological cost and impaired outcomes (Gross, 1998). The stronger claim sometimes attached to it, the ego-depletion account, according to which such effort draws down a depletable pool of self-control, is retained only in a weak form, because a multi-laboratory preregistered replication found the depletion effect indistinguishable from zero (Hagger et al., 2016). The argument here requires only that suppression is effortful and repeated without recovery, which the demands-and-resources tradition establishes independently (Bakker and Demerouti, 2007). Where the demand is chronic, repeated activation of the stress-response systems is held to accumulate as physiological wear, conceptualized as allostatic load (McEwen, 1998) — an interpretive frame for measurement that the main study intends to add, since no physiological measurement was undertaken here and no result below bears on any biological quantity.

The strain arises from the gap between display and feeling and accumulates rather than striking in single episodes; in frontline tourism guest contact is continuous through the peak, which is held to limit recovery. Resources can attenuate it: the

capacity to regulate one's own emotional responses varies between workers, and at team level psychological safety (Edmondson, 1999) offers a route by which strain can be disclosed. Both are examined as potential buffers.

2) Hidden costs and human-capital depreciation

Emotional labour is remunerated only indirectly: the wage attaches to the visible duties of the post and not to the regulatory effort (Grandey and Sayre, 2019). The theory of equalizing differences holds that jobs carrying onerous conditions should command a compensating wage premium (Rosen, 1986), but whether the market in fact compensates emotional labour is an empirical question these data cannot settle: a perceptual survey records how workers assess their pay, not the wage differentials obtaining across the market.

Costs not reflected in the wage reappear as hidden costs to the firm, of which presenteeism is the clearest instance, whose cost can exceed that of absence because the worker is present and paid but performs below capacity (Johns, 2010; Miraglia and Johns, 2016). In a seasonal operation the worker who is present but depleted during the peak is the one on whom the guest's experience most depends. The consequence is best expressed in the language of human capital: the capability a worker brings to the role is a stock in which worker and firm have both invested (Becker, 1964), and sustained emotional labour without recovery depletes it, with exhaustion and departure as its expected markers. What is threatened is the durability of the workforce, not only present output.

Seasonality concentrates the demand for emotional labour in a short peak, with three consequences. The burden is compressed into an intense episode rather than distributed across the year. Presenteeism is likely to be effectively forced where income is earned in a limited window and staffing is stretched. The retention problem, in turn, takes a specific form — failure to return in the following season — a loss of human capital at the moment it has been accumulated (Radlińska and Gardziejewska, 2022; Guidetti, Pedrini and Zamparini, 2021). In a thin local labour market these workers are also the most readily recruited by neighbouring establishments, so compression accumulates the cost while the thinness of the market weakens the incentive to prevent it — which is why the case for provision organized above the level of the firm has an economic and not merely a social basis (Acemoglu and Pischke, 1999).

III. MATERIALS AND METHODS

1) Design and instrument

The study is a cross-sectional pilot conducted by computer-assisted web interviewing, for instrument screening and estimation of the direction and approximate magnitude of associations. Screening is used deliberately in place of validation: validation in the psychometric sense requires evidence of convergent and discriminant validity and, ordinarily, a confirmatory factor model, whereas with five of six scales carrying two items each no factor model is estimable and no external criterion was available.

Respondents answered a short self-report questionnaire on five-point scales, from one for strong disagreement to five for strong agreement; the instrument is short by construction, since completion time is a constraint for frontline workers. It measures surface acting and emotional dissonance (Grandey, 2000); emotional exhaustion, somatic symptoms and presenteeism; guest-contact intensity and peak-season workload; perceived pay adequacy, the fit between pay and emotional demands, and the intentions to leave and to return; and psychological safety (Edmondson, 1999) and self-regulation. The full item list is in the supplementary materials.

Validated instruments exist for every construct measured: the Emotional Labour Scale (Brotheridge and Lee, 2003), the acting scales of Grandey (2003), the Stanford Presenteeism Scale (Koopman et al., 2002) and the Oldenburg Burnout Inventory (Demerouti et al., 2003). They were not used at this stage because none has been adapted to a compressed-season mountain-tourism labour market or validated in Polish on a seasonal tourism population, and their items presuppose a continuous employment relation and a stable reference period: an item asking about the last month means something different to a worker whose year contains a ten-week peak and a long trough. Adapting such an instrument requires prior knowledge of which demands are salient, and the purpose of this pilot was to obtain that knowledge first.

Items within each scale were worded in a common direction, which keeps the instrument transparent for a frontline respondent but does not guard against response style, treated below as a finding rather than a caveat. Cronbach's alpha is reported for the five-item surface-acting scale; for the five two-item scales it is not the appropriate statistic, since with two items alpha is a deterministic function of the single inter-item correlation, and the Spearman-Brown coefficient is the recommended estimate (Eisinga, Grotenhuis and Pelzer, 2013). Both are reported for each two-item scale. This does not rescue the failing scales, but it reports them in the form that identifies the problem: a near-zero correlation between two items presumed to measure one thing.

2) Sample and analysis

The target population comprises frontline workers in accommodation and food-service enterprises in the Nowy Targ and Tatra districts. Sampling was purposive and directed at guest-facing roles, with a target of fifty to sixty completed responses. Establishment managers and owners known to the authors distributed the questionnaire link to guest-facing staff, and regional tourism organizations circulated it among member firms; because distribution was delegated, no response rate can be computed. Participation was voluntary and responses anonymous. Sixty-four complete responses were obtained, with no missing data.

The pilot was designed around seasonally contracted staff, and the realized sample is not composed that way: forty-seven respondents (73.4 per cent) hold year-round contracts and seventeen (26.6 per cent) seasonal ones. The unit of exposure in this argument is the compressed season and not the contract: a year-round worker in these establishments is exposed to the same peak, and what differs is the contractual horizon over

which that exposure is borne; Section 4.5 reports the check. Combined accommodation-and-food establishments and thermal-bath operations account for seventy-five per cent of the sample, and such establishments trade through the year with a pronounced peak rather than a closed low season. The study can therefore describe what this sample reported about peak-season emotional demand and flag contract-group differences as candidates for a powered comparison, but cannot estimate the effect of the contract itself.

The core hypotheses are examined by Spearman rank correlation, with the effect size read from the coefficient itself and with exact probabilities and ninety-five per cent confidence intervals in place of significance asterisks; intervals are obtained by Fisher transformation with the correction appropriate to rank correlations. Exploratory group differences are examined by the Mann-Whitney U test, with the qualification that partitioning a sample of this size yields subgroups whose comparison is indicative only, and multiple regression is not estimated. The analysis comprises a family of twenty-three confirmatory tests — the fifteen correlations among surface acting, exhaustion, somatic symptoms, presenteeism and the two turnover items, three exploratory correlations and five group comparisons — to which both the Holm step-down correction (Holm, 1979) and the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995) are applied, with the corrected family in the supplementary materials. Every test bearing on H1–H3 or E1–E3 belongs to the family; convergent-validity checks and item-level coefficients are diagnostic and reported outside it. Otherwise the narrative of a pilot is liable to be built around whichever tests fall below the conventional threshold, which is indistinguishable from selection on the outcome. Conversely, a non-significant result is uninformative rather than evidence of absence: power reaches eighty per cent at a rank correlation of about 0.34 against an uncorrected threshold, but only at about 0.50 against the strictest Holm threshold.

IV. RESULTS

1) Sample characteristics

Sixty-four frontline workers completed the questionnaire; their characteristics are given in Table 1. Contract type is confounded with age and tenure: eleven of the seventeen seasonally contracted respondents were aged twenty-four or under, against nine of the forty-seven year-round respondents, so any contract-group difference below may reflect career stage.

TABLE 1. CHARACTERISTICS OF THE RESPONDENTS.

Characteristic	Category	n	%
Gender	Woman	45	70.3
	Man	19	29.7
Age	24 or under	20	31.3
	25–34	18	28.1
	35–44	17	26.6
	45 or over	9	14.1
Role	Reception	19	29.7
	Waiting / food service	17	26.6

Characteristic	Category	n	%
	Guiding activities	3	4.7
	Other guest-facing	25	39.1
Establishment	Accommodation	5	7.8
	Food service	11	17.2
	Combined	35	54.7
	Thermal bath / leisure	13	20.3
Contract	Seasonal	17	26.6
	Year-round	47	73.4
Tenure	Up to 1 season	28	43.8
	2–3 seasons	36	56.3
District	4 seasons or more	22	34.4
	Nowy Targ	42	65.6
	Tatra		

Source: own elaboration (N = 64).

2) Scale reliability

Internal consistency varied markedly, which is itself a central result of a pilot (Table 2). Three scales reached acceptable consistency — emotional exhaustion, psychological safety and self-regulation, with Spearman-Brown coefficients of 0.79, 0.79 and 0.81 resting on inter-item correlations between 0.65 and 0.68. The two presenteeism items correlate at $r = 0.01$ ($\rho = -0.003$, $p = 0.98$): they are statistically independent, and averaging them produces a composite with no common variance to interpret. The confidence intervals in Table 2 qualify these verdicts: for the somatic scale the interval runs from -0.13 to 0.58 , so the defensible conclusion is that its reliability is not established rather than that the scale failed, whereas for presenteeism an interval of -0.61 to 0.40 around a point estimate of 0.02 admits no such reading.

TABLE 2. INTERNAL CONSISTENCY OF THE SCALES.

Scale	Items	Cronbach's alpha	Inter-item r	Spearman-Brown	95% CI
Surface acting / emotional dissonance	5	0.30	0.08 (mean)	n/a (k = 5)	$-0.02-0.54$
Emotional exhaustion	2	0.79	0.66	0.79	$0.65-0.87$
Somatic symptoms	2	0.32	0.19	0.32	$-0.13-0.58$
Presenteeism	2	0.02	0.01	0.02	$-0.61-0.40$
Psychological safety	2	0.79	0.65	0.79	$0.65-0.87$
Self-regulation	2	0.79	0.68	0.81	$0.66-0.87$

Source: own elaboration (N = 64). For two-item scales the Spearman-Brown coefficient is the operative estimate (Eisinga, Grotenhuis and Pelzer, 2013); alpha is raw and intervals are F-based.

The surface-acting scale requires separate treatment, because its failure is not merely that of a short scale. Its five items share almost no common variance: the mean inter-item correlation is 0.079 and, of the ten item pairs, three are negative. Both SA1 ($M = 4.52$, $SD = 0.50$) and SA3 ($M = 4.78$, $SD = 0.42$) were answered at four or five by every respondent, and SA1 is, as Section 4.4 shows, the item responsible for the anomalous sign of the scale-level association with exhaustion. The item-level matrix is in the supplementary materials.

3) Descriptive statistics

The surface-acting composite, emotional exhaustion,

presenteeism, guest-contact intensity and peak-season workload were all reported at high levels, with means from 4.28 to 4.67 , while perceived pay adequacy was low (1.72) and the perceived correspondence between pay and emotional demands lower still (1.38); Table 3 reports the distributions.

These distributions require a qualification that governs how the whole of Section 5 should be read. The high means on the burden variables are accompanied by severe restriction of range and marked negative skew: the surface-acting composite has a standard deviation of 0.46 with an observed minimum of 3.0 , two of its items were answered at four or five by every respondent, and item skew runs from -0.06 to -2.16 . Every item was worded in the same direction. Content-free acquiescence is not, however, the explanation: respondents disagreed heavily with the equally positively keyed pay items (1.6 and 0.0 per cent at the agreement end), which a general tendency to agree could not produce. What cannot be separated from the high burden means is a self-presentation keyed to the same narrative, since the design contains no balanced item and no criterion outside the questionnaire. The levels of the burden variables are therefore not interpretable, whereas the pay and psychological-safety distributions are not ceiling-bound and can be read.

TABLE 3. MEANS AND STANDARD DEVIATIONS OF THE PRINCIPAL VARIABLES.

Variable	M	SD	Min	Max	% ≥ 4
Surface acting	4.30	0.46	3.0	5.0	78.1
Emotional exhaustion	4.28	0.90	1.0	5.0	85.9
Somatic symptoms	3.78	0.84	1.5	5.0	57.8
Presenteeism	4.39	0.63	3.0	5.0	81.2
Guest-contact intensity	4.64	0.90	1.0	5.0	95.3
Peak-season workload	4.67	0.74	2.0	5.0	93.8
Perceived pay adequacy	1.72	0.58	1.0	4.0	1.6
Pay–emotional-demand fit	1.38	0.52	1.0	3.0	0.0
Intention to leave	2.75	0.91	1.0	5.0	20.3
Intention to return	3.69	1.17	1.0	5.0	57.8
Psychological safety	3.09	1.15	1.0	5.0	37.5
Self-regulation	3.90	0.77	2.0	5.0	81.2

Source: own elaboration (N = 64). The final column reports the proportion at the agreement end.

4) Core associations

Table 4 reports the eleven correlations bearing on H1–H3; the full family is in the supplementary materials. For H1, the surface-acting composite was positively associated with somatic symptoms ($\rho = 0.29$) while its association with exhaustion was negative ($\rho = -0.19$, $95\% \text{ CI } -0.43 \text{ to } +0.07$); for H2 it was positively associated with presenteeism ($\rho = 0.31$). For H3, emotional exhaustion — the one reliable scale among these — was associated with neither the intention to leave ($\rho = -0.23$) nor the intention to return ($\rho = 0.08$), so H3 was tested and not supported.

TABLE 4. SPEARMAN RANK CORRELATIONS AMONG THE CORE VARIABLES
(CORRELATIONS BEARING ON H1–H3).

Pair	rho	p	95% CI	Scale reliability
Surface acting – Emotional exhaustion	–0.19	0.124	–0.43 to +0.07	SA unvalidated
Surface acting – Somatic symptoms	+0.29	0.021	+0.03 to +0.51	Neither validated
Surface acting – Presenteeism	+0.31	0.012	+0.06 to +0.53	Neither validated
Surface acting – Intention to leave	+0.12	0.331	–0.14 to +0.37	SA unvalid.; 1 item
Surface acting – Intention to return	+0.02	0.890	–0.24 to +0.28	SA unvalid.; 1 item
Emotional exhaustion – Somatic symptoms	+0.02	0.883	–0.24 to +0.28	SOM not estab.
Emotional exhaustion – Presenteeism	+0.22	0.077	–0.04 to +0.46	PRES not reliable
Emotional exhaustion – Intention to leave	–0.23	0.072	–0.46 to +0.04	EX ok; 1 item
Emotional exhaustion – Intention to return	+0.08	0.519	–0.18 to +0.33	EX ok; 1 item
Somatic symptoms – Presenteeism	+0.37	0.003	+0.12 to +0.57	Neither validated
Intention to leave – Intention to return	–0.63	<0.001	–0.76 to –0.44	Single items

Source: own elaboration (N = 64). Intention to leave and intention to return are single items.

The scale-level association between the surface-acting composite and emotional exhaustion is not merely weak but negative. The meta-analytic literature reports a positive association of moderate size across three decades of studies (Hülsheger and Schewe, 2011), and a scale that reverses one of the most consistently replicated findings in the field is very unlikely to be measuring surface acting. Of the five items, SA1 correlates –0.51 with the exhaustion composite while the remaining four correlate between –0.13 and +0.04, so SA1 alone accounts for the negative coefficient. Restriction of range does not explain this: SA3 is the more restricted item, since 78 per cent of respondents chose the maximum against 52 per cent on SA1, yet its association with exhaustion is +0.04, whereas SA1, which divides the sample almost evenly between the top two categories, carries the reversal. The more plausible reading is that SA1 records a normative statement about how one ought to conduct oneself with guests rather than a report of dissonant display. Removing it would not repair the scale, since the remaining items correlate with exhaustion at essentially zero.

The consequence is that H1 was not tested, since a hypothesis test presupposes a measure of the construct named in the hypothesis. The same reasoning extends to H2 and to the pay association in Section 4.5: both rest on the same composite, so neither is a test of the relation it names, and both are reported as properties of the instrument rather than as findings about the constructs. It is also why correcting for attenuation is not pursued: computed from the unrounded reliabilities it returns coefficients outside the admissible range (3.67 and 4.16). A second internal check points the same way. Emotional exhaustion and somatic symptoms correlate at 0.02, where the literature would lead one to expect a substantial positive

association; here too the failure is attributable to the somatic scale rather than the exhaustion scale, whose two items correlate at 0.66 and which behaves as expected against another quantity in the instrument: exhaustion correlates +0.39 with peak-season workload ($p = 0.001$, 95% CI +0.15 to +0.59). That is the pilot's only convergent signal, and it comes from the same method.

5) Exploratory analyses and multiplicity

Perceived pay adequacy was not positively associated with the surface-acting composite; the association was negative ($\rho = -0.31$, $p = 0.013$), with the caveat above about what that composite measures. For E1, seasonal and year-round workers did not differ detectably in exhaustion ($U = 480$, $p = 0.205$) or presenteeism ($U = 466$, $p = 0.285$), but seasonally contracted workers reported a higher intention to leave (3.35 against 2.53; $U = 587$, $p = 0.003$, rank-biserial $r = 0.47$), a lower intention to return (3.18 against 3.87; $U = 256$, $p = 0.024$, $r = 0.36$), and somewhat more adequate pay (2.00 against 1.62; $U = 511$, $p = 0.044$). E3 was not supported: neither psychological safety nor self-regulation was associated with exhaustion ($\rho = 0.08$ and 0.02). The comparison by guest-contact intensity could not be performed: fifty of the sixty-four respondents chose the top category. The contract groups differed by less than a tenth of a scale point on peak-season workload and guest-contact intensity, and neither difference approached significance; this is not evidence of equivalence, but nothing in it suggests the demand differs by contract.

Two further diagnostics belong here. A single-factor test on all twenty-two items gives the first unrotated factor 19.9 per cent of the variance, well below the conventional fifty-per-cent threshold; this counts against common-method variance as a general explanation, though the test is weak at this sample size. Correction for multiplicity also changes what may be claimed. Across the family of twenty-three tests, only the association between the intention to leave and the intention to return survives Holm control, and it is a coherence check on two single items rather than a substantive finding. Under Benjamini-Hochberg at a false-discovery rate of 0.05, two further results are retained: the higher intention to leave among seasonally contracted workers ($q = 0.023$) and the association of somatic symptoms with presenteeism ($q = 0.023$), the latter between two scales whose reliability is not established. At 0.10 the associations of the surface-acting composite with presenteeism, somatic symptoms and pay adequacy, and the lower intention to return among seasonal workers, are also retained.

V. DISCUSSION

1) What the pilot established

The pilot met its primary purpose, which was to screen the instrument rather than to confirm a theory. The exhaustion, psychological-safety and self-regulation scales function; the presenteeism scale does not, the somatic scale is inconclusive, and the surface-acting composite fails as a measure of its construct, since it reverses the best-established association in this literature through one item read normatively rather than

through restriction of range. The three problematic scales are exactly those for which validated instruments were available and were not used, which settles how the main study should proceed. The descriptive picture must correspondingly be stated with care: high reported burden alongside low perceived pay adequacy is what an unrelieved and unpriced demand would look like (Grandey and Sayre, 2019), and also what a self-presentation keyed to that narrative would look like.

2) Seasonality, presenteeism and retention

Among the group comparisons, one is both robust to a false-discovery correction and interpretable: seasonally contracted workers reported a stronger intention to leave than year-round workers ($q = 0.023$); the weaker intention to return is retained only at a false-discovery rate of 0.10. Read as human-capital depreciation (Becker, 1964), this is the retention margin on which a seasonal operation is most exposed, since accumulated capability is thought to be lost between seasons (Radlińska and Gardziejewska, 2022; Guidetti, Pedrini and Zamparini, 2021). Exhaustion and presenteeism did not differ detectably between the groups, and the comparison is underpowered; the high reported presenteeism is consistent with accounts in which working while ill tracks constraints on absence as much as ill health (Miraglia and Johns, 2016), with the caveat that the presenteeism composite has no interpretable common variance. For the year-round majority the referent of the return item is also not self-evident, since for them the coming season is a phase of an ongoing employment relation rather than a decision point. Contract is also confounded with age and tenure.

3) The question of compensation

The result on pay is the one most likely to be over-read. Under the theory of equalizing differences an onerous demand should attract a compensating premium (Rosen, 1986); here perceived pay adequacy was negatively associated with the surface-acting composite ($\rho = -0.31$) and low in absolute terms. Four readings are available and the design cannot discriminate among them. The demand may be genuinely uncompensated. The association may be common-method variance aggravated by negative affectivity (Podsakoff et al., 2003), though the single-factor test gives that account no support, and a uniform negativity bias should also depress pay-adequacy ratings among the more exhausted, whereas exhaustion correlates positively with them ($\rho = +0.34$, $p = 0.006$). The order may be the reverse: a worker who considers themselves underpaid may withdraw authentic engagement while still meeting the display rules (Siegrist, 1996). Or the measure may not address the theory at all, a compensating differential being a property of the wage structure rather than of one worker's satisfaction with one wage.

4) Resources, limitations and the main study

Neither psychological safety nor self-regulation was associated with exhaustion, although both were measured reliably. Under the demands-and-resources account resources attenuate strain by offsetting demands (Bakker and Demerouti, 2007); that pattern does not appear here. This should be read against the ceiling problem, since with exhaustion reported at four or five by eighty-six per cent of respondents there is little variation for a buffer to explain. It nonetheless tempers any

expectation that individual competence alone would offset a demand that appears structural, without removing the case for developing such competence.

The limitations are those of a pilot, and the foremost is measurement: the presenteeism scale has no interpretable common variance, the somatic scale has a reliability interval too wide to establish, and the surface-acting composite reverses the sign of a robustly established association, so H1 and H2 were not tested rather than partly supported. Survivorship: the questionnaire offered a tenure category of up to one season, but no respondent selected it, so every respondent had by definition returned at least once and the central retention construct has been measured on a sample from which everyone who failed to return was already excluded. First-season entrants, for whom attrition is highest, are absent entirely, so the intentions reported here are estimated on persisters and are very likely more favourable than in the population of interest; the main study must recruit within the season, through channels independent of continuous attachment to an establishment. Response style: every item ran in one direction and the burden distributions are ceiling-bound, so no claim about the level of those constructs can be defended. Multiplicity: twenty-three tests were conducted on sixty-four respondents and one survives familywise correction. Construct ambiguity and confounding: the intention-to-return item has no clear referent for the year-round majority, contract type is confounded with age and tenure, and the design is cross-sectional and wholly self-reported (Podsakoff et al., 2003).

The pilot makes the main study concrete by specifying the measurement work required first, and that work is not the revision of the failing scales but their replacement: the surface-acting scale with a translated and culturally adapted short form of an established instrument (Brotheridge and Lee, 2003); the presenteeism items with an adapted Stanford Presenteeism Scale (Koopman et al., 2002), whose items measure reduced output while present rather than the decision to attend; and the somatic scale with an established inventory. The three reliable scales can be retained alongside an established burnout measure (Demerouti et al., 2003) run in parallel for calibration, and all scales will carry balanced polarity with reverse-keyed markers. The relations worth powering are the negative association between pay adequacy and surface acting and the seasonal difference in turnover and return intention. A longitudinal design, with measurement before and after the peak and the pay-perception measure taken first, would separate accumulation of strain within a season from stable differences between workers; complementing self-report with supervisor ratings and administrative records of return would reduce common-method variance; and sampling across several districts, with contract types in balance and first-season workers included, would support the subgroup comparisons the present sample could only flag.

VI. CONCLUSIONS

This pilot examined, among sixty-four frontline tourism workers surveyed in the Podhale region, whether emotional labour is accompanied by the hidden costs that would mark it as an unpriced burden on human capital. Its contribution is a screened instrument, part of which works and part of which must be replaced. Of six scales, three are usable — emotional exhaustion, psychological safety and self-regulation, with two-item reliabilities of 0.79 to 0.81 — and the exhaustion scale additionally behaves as expected against peak-season workload. Three are not, and the failure of the surface-acting composite is one of validity rather than of consistency, since it inverts the best-established association in the field through a single item that restriction of range does not explain.

The descriptive picture is compatible with the study's premise without constituting evidence for it: the burden distributions are at ceiling, so their levels cannot be interpreted, although the low pay-adequacy ratings are not subject to that objection. H1 and H2 were not tested, because the composite that names them does not measure the construct; H3, which rests on the reliable exhaustion scale, was tested and not supported. Within a family of twenty-three tests one association survives familywise correction, so none of these patterns is offered as an established finding. The evidence is perceptual, cross-sectional and drawn from a small regional sample composed entirely of workers who had already returned at least once, so the study indicates rather than demonstrates.

Its value lies in establishing what can and cannot yet be measured, and in a transferable warning: in a workforce of this kind, single-direction items yielded burden distributions we could not interpret, and an item about display rules was read as a norm rather than as a report. Whether firms in a thin seasonal labour market can sustain a compensating premium, or whether the cost is displaced onto the worker, is the question the main study is placed to pursue.

Supplementary Materials: item list; surface-acting item-level matrix; corrected family of twenty-three tests.

Acknowledgments and funding: no external funding.

Author Contributions (CRediT). K.O.: conceptualization, methodology, investigation, formal analysis, data curation, software, writing — original draft. M.S.: conceptualization, methodology, validation, writing — review and editing, supervision.

Conflicts of Interest: the authors declare no conflict of interest. The practical motivation for the study arose in connection with a vocational-training project in the tourism-services sector at the first author's institution; that project did not fund this article and had no role in its design, analysis or reporting.

Ethics: the study involved a voluntary and anonymous survey of adult workers about their own working experience. Two items ask about self-reported physical complaints; no identifying data were collected and no response can be linked to an individual. Under the applicable national and institutional rules, formal ethics-committee approval is not required for anonymous, non-interventional survey research of this kind. All

procedures follow the Declaration of Helsinki.

Data Availability: the anonymized item-level dataset and the questionnaire will be deposited in a public repository on publication.

VII. REFERENCES

- Acemoglu, D. and Pischke, J.-S. (1999). The structure of wages and investment in general training. *Journal of Political Economy*, 107(3), pp. 539-572. <https://doi.org/10.1086/250071>
- Bakker, A. B. and Demerouti, E. (2007). The Job Demands-Resources model: state of the art. *Journal of Managerial Psychology*, 22(3), pp. 309-328. <https://doi.org/10.1108/02683940710733115>
- Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. New York: Columbia University Press for the NBER.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57(1), pp. 289-300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Brotheridge, C. M. and Lee, R. T. (2003). Development and validation of the Emotional Labour Scale. *Journal of Occupational and Organizational Psychology*, 76(3), pp. 365-379. <https://doi.org/10.1348/096317903769647229>
- Demerouti, E., Bakker, A. B., Vardakou, I. and Kantas, A. (2003). The convergent validity of two burnout instruments: a multitrait-multimethod analysis. *European Journal of Psychological Assessment*, 19(1), pp. 12-23. <https://doi.org/10.1027/1015-5759.19.1.12>
- Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44(2), pp. 350-383. <https://doi.org/10.2307/2666999>
- Eisinga, R., Grotenhuis, M. te and Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, 58(4), pp. 637-642. <https://doi.org/10.1007/s00038-012-0416-3>
- Grandey, A. A. (2000). Emotion regulation in the workplace: a new way to conceptualize emotional labor. *Journal of Occupational Health Psychology*, 5(1), pp. 95-110. <https://doi.org/10.1037/1076-8998.5.1.95>
- Grandey, A. A. (2003). When the show must go on: surface acting and deep acting as determinants of emotional exhaustion and peer-rated service delivery. *Academy of Management Journal*, 46(1), pp. 86-96. <https://doi.org/10.5465/30040678>
- Grandey, A. A. and Sayre, G. M. (2019). Emotional labor: regulating emotions for a wage. *Current Directions in Psychological Science*, 28(2), pp. 131-137. <https://doi.org/10.1177/0963721418812771>
- Gross, J. J. (1998). The emerging field of emotion regulation: an integrative review. *Review of General Psychology*, 2(3), pp. 271-299. <https://doi.org/10.1037/1089-2680.2.3.271>
- Guidetti, G., Pedrini, G. and Zamparini, L. (2021). Assessing perceived job quality among seasonal tourism workers: the case of Rimini, Italy. *Tourism Economics*, 27(8), pp. 1629-1649. <https://doi.org/10.1177/1354816620927524>
- Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H. et al. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science*, 11(4), pp. 546-573. <https://doi.org/10.1177/1745691616652873>
- Hochschild, A. R. (1983). *The Managed Heart: Commercialization of Human Feeling*. Berkeley: University of California Press.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), pp. 65-70.
- Hülsheger, U. R. and Schewe, A. F. (2011). On the costs and benefits of emotional labor: a meta-analysis of three decades of research. *Journal of Occupational Health Psychology*, 16(3), pp. 361-389. <https://doi.org/10.1037/a0022876>

- Johns, G. (2010). Presenteeism in the workplace: a review and research agenda. *Journal of Organizational Behavior*, 31(4), pp. 519-542. <https://doi.org/10.1002/job.630>
- Koopman, C., Pelletier, K. R., Murray, J. F. et al. (2002). Stanford Presenteeism Scale: health status and employee productivity. *Journal of Occupational and Environmental Medicine*, 44(1), pp. 14-20. <https://doi.org/10.1097/00043764-200201000-00004>
- McEwen, B. S. (1998). Stress, adaptation, and disease: allostasis and allostatic load. *Annals of the New York Academy of Sciences*, 840, pp. 33-44. <https://doi.org/10.1111/j.1749-6632.1998.tb09546.x>
- Miraglia, M. and Johns, G. (2016). Going to work ill: a meta-analysis of the correlates of presenteeism and a dual-path model. *Journal of Occupational Health Psychology*, 21(3), pp. 261-283. <https://doi.org/10.1037/ocp0000015>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y. and Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), pp. 879-903. <https://doi.org/10.1037/0021-9010.88.5.879>
- Radlińska, K. and Gardziejewska, B. (2022). The seasonal labor hoarding in tourist enterprises—choice or necessity? *Sustainability*, 14(12), 6995. <https://doi.org/10.3390/su14126995>
- Rosen, S. (1986). The theory of equalizing differences. In: O. C. Ashenfelter and R. Layard, eds., *Handbook of Labor Economics*, vol. 1. Amsterdam: North-Holland, pp. 641-692. [https://doi.org/10.1016/S1573-4463\(86\)01015-5](https://doi.org/10.1016/S1573-4463(86)01015-5)
- Siegrist, J. (1996). Adverse health effects of high-effort/low-reward conditions. *Journal of Occupational Health Psychology*, 1(1), pp. 27-41. <https://doi.org/10.1037/1076-8998.1.1.27>
- Tokarz-Kocik, A., Bera, A., Drela, K. and Malkowska, A. (2023). The impact of the COVID-19 pandemic on the labour market in the hotel industry: selected conditions in Poland. *Sustainability*, 15(6), 4886. <https://doi.org/10.3390/su15064886>