AI Human Resource Tools and Employee Performance The Mediating Role of Job Burnout

Yule Liu*

School of Management, Hebei University, Baoding, 071002, China

*1356326971@qq.com

Abstract

When AI tools are used in human resource management, they can offer real-time feedback and amplify the requirement to evaluate using data. This study will investigate the potential of this mediating role between the two variables-perceived exposure to these tools and task performance for the employees-using the statistical design. There are six items in the questionnaire and the items refer to the past four weeks. There are 342 records (out of 400 submitted) that have completed answers on all 18 substantive items and are analyzed in the main analysis. Ordinary least squares regression with HC3 robust standard errors and 5,000 percentile bootstrap resample analyses results indicate a positive association of exposure to burnout (a = 0.339) and a negative association of burnout with performance (b = −0.266). The indirect effect is −0.090, with a 95% bootstrap interval of [−0.132, −0.055]. However, it has an indirect effect, which is small (0.029, p=0.515) and also has a total effect which is not significant (−0.062, p=0.161). Estimates obtained by using the alternative scoring and by covariate adjustment are similar and negative. The results indicated an indirect association of the findings to burnout that was not necessarily connected to a competitive mediation, an efficiency benefit, or causal cause. The findings suggest that focusing on employee experience is crucial when rolling out AI HR systems, in addition to their impact on performance.

Keywords

Artificial intelligence, human resource management, job burnout, employee performance, questionnaire, mediation analysis.

1. Introduction

Human resource systems with AI incorporate computational performance evaluation, employee feedback, workforce planning as well as suggesting what can be done for them so far as development is concerned. Most of the time their organizational relevance is discussed in terms of speeding up the processing or consistency in decisions made. But these systems can be a primary means by which they are constellated as workers, workers who are appraised and impactful to their careers. That's why it's important to have an understanding of not just the opportunities the tools offer, but also the psychological experiences that they elicit.

Tambe, Cappelli and Yakubovich [1] point out that despite this promise of AI in HRM, there are still some conditions in place that must be fulfilled, and this gives rise to the gap. At work, information does not necessarily lead to optimizations in performance when carrying out tasks. Exposure to additional evaluation demand may be a source of burnout and burnout or disengagement may lead to poor ongoing work quality. Such a mediation framework enables them to be considered jointly, instead of assuming technology is necessarily a good or bad thing.

This study involves the analysis of a questionnaire data set of 400 records, and a negative indirect pathway test involving job burnout. It also examines the assumption that – once

burnout is taken into account – exposure is positively directly related to performance. The differentiation is important, since direct relations might be unclear even if the indirect association is discernible. The analysis includes 342 records that are complete for analysis, includes measurement diagnostics and assesses sensitivity to covariate adjustment and to handling missing response treatment. One of its contributions is the crystal clear representation of the results of the examined relationships, including the lack of the posited direct relationship.

2. Theoretical Framework and Hypotheses

2.1. Exposure to AI Human Resource Tools

AI HR Exposure: How well employees feel they are monitored, assessed, given feedback, and may be suggested for tasks, development or personnel decisions by systems. It's a system where you're subject to it and are directly involved in it. This is a more individual level gauge of technology adoption than a firm level indicator of the technology being used and it is less voluntary than an indicator of general purpose AI use. Without a human manually entering instructions into software, evaluation of an employee can be done algorithmically.

Algorithmic research on management has proved to be of great importance to differentiate management functions and exposure. Parent-Rocheleau et al. [2] developed a questionnaire on the themes of monitoring, goal setting, scheduling, performance rating and compensation. The present questionnaire's item sources are not provided by their measure, but rather only conceptualized. Here the six exposure items were developed by the researcher and evidence for another instrument or tool cannot be shifted to the items here. In particular, there is no way to independently verify if a product, deployed or not, makes use of machine learning.

2.2. Resources and Demands

Resources can be related to the job and employee energy functioning as per the job demands resources theory [3, 4]. Accessible feedback and more clearly-stated expectations could function as assets with AI HR systems. Higher levels of testing, lack of knowing results of testing, and responding to system recommendations may act as demands. These are not mediators that were measured within the current data but rather implementation characteristics (theoretical). Being exposed does not necessarily mean employees had a specific resource and demand.

Algorithmic control research also indicates that by the use of technology, work can be directed and evaluated in other ways [5]. This angle acts as an incentive for looking at burnout, and not just taking it that a better administrative process would lead to better worker outcomes. Concurrently, performance may be empowered by a pertinent improvement direction or creating a significant assistance proposal. The model thus provides for a direct causal path of a positive relationship as well as a causal path of a negative relationship through burnout, leaving the existence of either or both only to the information in the data.

2.3. Hypotheses and Effect Decomposition

Previous to studying the relationship between AI HR exposure and task performance, we ask our readers to consider H1, which postulates a positive association between AI HR exposure and task performance, taking burnout into account. This conditional direct relationship is an overview of paths not measured within the mediator. Since efficiency is not measured separately, even a positive direct coefficient alone could not measure time saving and a specific mechanism of efficiency. H2 suggests that there is a positive relationship between exposure and burnout. After accounting control for exposure, H3 proposes that burnout is negatively associated with performance. H4 suggests that there will be an indirect negative relationship between exposure and performance with burnout as the mechanism.

The Models are: Y = iY + c'X + bM + eY; M = iM + aX + eM; (i is exposure; a, b, c' and e are unknown parameters). A is indirectly associated with B: ab. We measure the total relationship as the regression of the Y on the X such that, the assumption that the same cases and covariates are used in the model, c = c' + ab. For evidence-based/competitive Mediation, both a positive direct part and indirect negative part would be necessary. An estimate for c' that is just positive is not enough.

The arrows refer to the suggested theoretical ordering. They do not address how the effects of poor performance impact subsequent monitoring or how effects of workload impact all three constructs. Due to the fact that all substantive questions concern the same four weeks, the statistics in the study are only calculating the statistical association in one wave. Longitudinal mediation studies have suggested against assuming the cross sectional estimates reflect a process and change over time [6].

3. Research Design and Data Preparation

3.1. Questionnaire and Eligibility

A questionnaire involving employees is utilized for the study, which is to be completed regarding perceived exposure to AI in HR, a recent work experience and an experience of task completion. The participation screen includes questions about the respondent's willingness to participate, age, current participation in the labor force, and the amount of time the respondent has been in his/her current role, as well as if he/she has experienced a particular tool in the last four weeks. Exposure means accessing the systems or being an object of evaluation, analysis or recommendation of these systems. 100% of the records submitted (400 records) show compliance with the eligibility criteria.

The instrument also differentiates between the tools and regular attendance and spreadsheets, as well as the popular personal AI writing tool, ChatGPT. The background questions cover industry, occupational group, occupational management position, length of time at work unit, hours worked in a normal week, how one was exposed, and system operation/function. They do not have enterprise or team identifiers. This means that you cannot estimate the clustering of organizations, compare the sample with a specified organizational sampling frame or check the product characteristics against enterprise records.

3.2. Scoring and Missing Responses

Exposure is quantified on a 5 degree "agreement" scale 0 (completely inconsistent) to 4 (completely consistent) in terms of the employee's situation as in X1–X6. For burnout M1–M6 (performance Y1–Y6), 5 levels of frequency were used ranging from never to almost always. Chinese data is examined using the same coding the questionnaire used; The present article presents the data in English translations of the questionnaires. All items are scored in the direction listed; each construct is measured by six items, the mean score of which is used to calculate the score for the construct. This means that the constructs have a 1–5 response metric.

There are 60 cells in the workbook that are marked unable to judge, out of the 7,200 substantive item cells, which is 0.83%. In 58 records, they have had one such response (and two have had two such responses). These are considered as missing and not scored as zero or at the midpoint. There are 342 records with everyone responding to all 18 items and those are selected to provide the primary analysis, which represents 85.5% of the records submitted. This percentage is not a survey response rate-a type of analytical retention rate is in use. There is NO manipulation of observed scores for obtaining "better" fits or significance levels of the model.

The file does not have any duplicate respondents and the complete case subset does not contain any duplicate 18 item response vectors or respondents with the same responses to all 18 items.

None of these are used for individual authenticating of response and there are no indicators for attentive completion. Response times and denominators of recruitment are not available for analysis and therefore speed screening and response rate calculation are not carried out. There are no optional comments to suggest what the responses would have been at those places on the scale if they had not been written.

3.3. Analytical Sample Composition

Table 1. Industry composition of the complete-case analytical sample

Industry	n	Percent
Information technology	119	34.8
Manufacturing	88	25.7
Retail and e-commerce	55	16.1
Financial services	34	9.9
Professional services	26	7.6
Other	20	5.8

Note: Percentages use n = 342. The categories are mostly descriptions of the analytical sample, and not meant to be representative of populations.

Among complete cases, 84 respondents report management responsibilities and 258 do not. The numbers 101, 110 and 131 indicate direct tool operation, analysis/evaluation of tool operation, and both, respectively. 97 are exposed less than a month, 125 exposed for 1–3 months, 70 exposed for more than 3–6 months and 50 more than 6 months. Variation in exposure experience occurs in the sample, but each of the sample items/retained respondents participated in passing the exposure screen.

4. Measurement Assessment and Estimation

4.1. Construct Coverage and Internal Consistency

Items in each section of the six exposure items address aspects of work data analysis, automated assessment, system delivered feedback, task/target suggestions, development recommendations, and managerial dependence of system results. Burnout items include feeling like running out of gas, feeling tired when starting the job, feeling like you can't recover, psychological distance, lack of enthusiasm and apathy. Performance items include timeliness, quality, and output required as well as organization of the task, any rework and accuracy in the aspects of job tasks. Self-reported task performance (rating) vs. an independent rating of performance (productivity).

Table 2. Descriptive statistics and internal consistency of six-item composites

Construct	Mean	SD	Alpha	CITC range
AI HR exposure X	2.891	0.935	0.933	0.787-0.814
Job burnout M	3.583	0.884	0.930	0.788-0.808
Task performance Y	3.115	0.745	0.886	0.683-0.730

Note: CITC denotes corrected item total correlation. These scores are higher if there was more exposure, more burnout and higher level of task performance, respectively. All statistics use complete cases.

Exposure, burnout and performance have a coefficient alpha of 0.933, 0.930 and 0.886, respectively. Corrected item total correlations range from 0.787 to 0.814, 0.788 to 0.808, and 0.683 to 0.730, respectively. Alpha is used to measure the internal consistency for this sample, but is not valid for items created/designed by the researcher [7]. High consistency does not necessarily mean that individual applications are interchangeable because there are unique management functions that will be addressed in each case.

4.2. Correlations and Preliminary Construct Separation

Table 3. Correlations between the three composite scores

Construct	X	M	Y
AI HR exposure X	1.000	0.358	-0.077
Job burnout M	0.358	1.000	-0.303
Task performance Y	-0.077	-0.303	1.000

Note: Entries are Pearson correlations. X = AI HR exposure; M = burnout; Y = task performance. The matrix is descriptive rather than a causal model.

Their absolute-correlation heterotrait monotrait ratios are 0.385 for the "exposure–burnout", 0.092 for the "exposure–performance" and 0.334 in the "burnout–performance" group. These diagnostics hint that the item blocks are not close to being equal, and come as a first approximation from a HTMT point of view [8]. They do not conclusively provide evidence to assess the discriminant validity or have the ability to replace confirmatory testing for measurement. The biggest factor for all 18 items is 36.45% of the variance of the standardized items, which would not be interpreted as evidence that common method bias is not a problem.

4.3. Estimation Procedure

Modeling of the mediator equation, conditional outcome equation and total relationship are Ordinary least square estimates. Large sample normal approximations (two-sided) are used for the regression p values and HC3 heteroskedasticity robust standard errors are used. All coefficients are unstandardized, i.e., not meant to be compared across analyses. In the case of indirect effects the 2.5th – 97.5th percentile of 5,000 complete-row bootstrap samples is used to create a 95% interval using both equations [9]. The main program's random number seed is 20261019. A row resampling approach assumes independence of observations; it is not possible to account for firm level dependence without having the cluster identifiers.

5. Main Results

5.1. Exposure and Burnout Paths

The p value is < 0.001, a= 0.339, and HC3 standard error of 0.048 and it is a positive relationship with burnout. There is about a one-third of a point increase in the burnout measure for each one-point difference in exposure on the questionnaire measure. This association is coherent with H2 but the coefficient does not indicate which is the source of the burden in the system. The amount of variance due to exposure only mediated model is 12.85% of sample burnout variance.

The results of the conditional performance equation indicate that the relationship between burnout and performance is negative with b = −0.266, an HC3 standard error of 0.046 and a p of < 0.001. This corroborates the association of H3 in the sample being observed. 9.29% of the variance in performance is accounted for by the conditional model. Thus, the majority of the

variation observed is outside the boundaries of the model-that is, the burnout explanation may not be sufficient for understanding employee performance.

Table 4. Unadjusted mediation model estimates for 342 complete cases

Component	Estimate	HC3 SE	p	95% bootstrap CI
Exposure to burnout a	0.339	0.048	<0.001	[0.249, 0.432]
Burnout to performance b	-0.266	0.046	<0.001	[-0.355, -0.178]
Direct effect c'	0.029	0.044	0.515	[-0.057, 0.116]
Indirect effect ab	-0.090	n/a	n/a	[-0.132, -0.055]
Total effect c	-0.062	0.044	0.161	[-0.143, 0.023]

Note: CI = 95% percentile Bootstrap confidence interval based upon 5,000 resamples. The regression p values are based on HC3 standard errors and normal approximations. This is because there is no indirect effect p value so you will find the interval.

5.2. Indirect Association and Hypothesis Assessment

The estimated indirect effect is −0.090, with a 95% bootstrap interval of [−0.132, −0.055]. It has a negative direction and interval excluding zero, which are in line with H4. The magnitude (value) refers to a statistical separation of the mean-score system of the questionnaire. Not the percent of employee work output lost, one of the clinical results of the burnout or organizational ROI. The rounded values of c', ab and c will not sum to zero because they are just estimates of the decomposition identity; the unrounded values of the estimate components are: c' = 0.029, ab = −0.090, and c = −0.062.

The direct exposure coefficient is positive but small, c' = 0.029, with p = 0.515 and an interval of [−0.057, 0.116]. So, H1 is not provided for. The total relationship is also nonsignificant, c = −0.062, p = 0.161, with an interval of [−0.143, 0.023]. The exposure only outcome model accounts for 0.60% of the variance of the outcomes of performance. The scores for each of these periods can be positive or negative, which means there is not a clear benefit or disadvantage of being exposed to AI as an HR.

Competitive mediation – the hypothesized positive direct component does not receive statistical support – should not be described as evidenced by the pattern. Nor does the lack of statistical significance, or zero of a direct coefficient indicate that its population counterpart is zero or that complete mediation exists. The negative indirect association under this model suggests that direct and total associations are unknown. This is because a confirmable total effect is not required as a precondition for coming up with an estimate of the product, but also does not turn the product into a causal mechanism.

6. Robustness and Sensitivity Analyses

6.1. Adjustment for Recorded Work Characteristics

Then you have an adjusted model where the factors of management responsibility, organizational tenure, weekly working hours and duration of AI exposure are included. These variables can pertain to either employee experience and performance, but changes do not necessarily result in no confounding. Categories are added in the form of 12 indicator columns, not arbitrary numbers. Regardless of the nature of the responsibility, 40–44 hours per week, and 1–3 months of exposure, reference groups are tenure-limited (1–3 years of tenure). The three equations have common entry of the same covariates.

The results for the burnout adjusted for exposure are 0.343 for the burnout to performance coefficient and −0.269 for the burnout to burnout coefficient. The adjusted indirect effect is −0.092, with an interval of [−0.135, −0.055]. The direct estimate remains small at 0.028 (p = 0.541), and the total estimate is −0.065 (p = 0.150). Thus conditioning on these recorded characteristics will not affect the main interpretation. The selected adjustment set is a specification on the observational sensitivity, and does not imply that all the variables are captured pretreatment confounders.

6.2. Alternative Burnout Scores and Missing Response Treatment

Table 5. Sensitivity of the indirect association to alternative specifications

Specification	n	Indirect ab	95% bootstrap CI	Direct p
Primary complete cases	342	-0.090	[-0.132, -0.055]	0.515
Adjusted complete cases	342	-0.092	[-0.135, -0.055]	0.541
Exhaustion items M1-M3	342	-0.084	[-0.125, -0.049]	0.604
Disengagement M4-M6	342	-0.083	[-0.123, -0.049]	0.629
At least five items per block	398	-0.088	[-0.126, -0.056]	0.417

Note: All rows are created with 5,000 percentile bootstraps. There are two unadjusted models – one for exhaustion (M1–M3) and one for disengagement (M4–M6). In each of the six item blocks at least five items must be answered for it to count as a proration.

An indirect estimate for the six-item mediator replacing the three exhaustion items is −0.084 with an interval of [−0.125, −0.049]. Using the three disengagement items gives −0.083, with an interval of [−0.123, −0.049]. Each is an exploratory grouping of content and not an independently-anchored subscale. Since the two groups are not fit at the same time, the resulting estimates cannot be summed together and cannot be viewed as independent (controlling for the other group).

An Exploratory Scoring Sensitivity is an average of the items available in each of the blocks when at least 5 of 6 are answered. It retains 398 records and gives an indirect effect of −0.088, with an interval of [−0.126, −0.056]. The direct coefficient is 0.033 (p = 0.417). This gets rid of a reliance on scoring with absolute crystal clear ‘complete’ answers, and preferring to use invented values to fill in the unknown answers. It does not put forth the idea that items are randomly missing and alters the items in the scores of some respondents. The result – 342 cases – is primary.

This means that the sign of the indirect association agrees with the primary and sensitivity analyses and a positive statistically supported direct association is not suggested. None of their agreement takes away self-report bias, measurement limitations, potentially organizational dependence or reverse causation. The focus of these analyses is merely to examine whether a specific statistical relationship is present or not and not to provide causal identification of the effects of AI in HR.

7. Discussion

7.1. What the Findings Establish

These data suggest a correlation between increased exposure to AI in the HR function, increased burnout and decreased self-reported task performance. It has negative scores for both its primary and alternative specifications. This aligns with the demand-based model of EE: EE systems that increase the use of data for evaluation and management might be experienced

parallel to job energy drain and decreased energy and enthusiasm. But the effects of evaluation pressure and/or perceived transparency are not measured separately, and are thus only offered as a speculation – perhaps for future research.

The results provide less support for the initial framework's resources components. It was hoped that there would be a positive direct coefficient, but it is too small and inaccurate. Thus, it is wrong to assume that the findings of the study mean that AI makes things more efficient, but burnout makes up for that. What is currently known is that there is a burnout related indirect association and the rest of the exposure to performance relationship is unresolved. The data do not differentiate between these possibilities – efficiency benefits, heterogeneity of benefits, poor self-reported measure of task performance, or measurement error.

This distinction affects the nature of the contribution – how it should be presented. Not a uniform "j" curve effect is shown in the study. It does, however, demonstrate that when there is a simple overall performance relationship, a psychologically relevant association between the skills may exist even though this relationship is weak. Reporting indirect effect estimates as well as direct and total estimates (when they are not significant) avoids a selective reporting of results and provides a clear indicator for future efforts that will want to replicate this study.

7.2. Implications for Organizational Practice

When considering AI HR systems, managers need to take into account employee experience and not just the operational metrics. Task quality, completion time, correction rates and employee reported fatigue could all be a part of a deployment review. Assessments of useful feedback and development recommendations should be made based on whether or not there is an AI product and not on whether the feedback or recommendations are useful. Existing associations are worthy of consideration with regard to employee burden, but do not provide any information on what would be the gain for an intervention in improving performance.

For those organizations that use an automated scoring process for decision making, they may look to see if their employees understand what they are being evaluated on, they can get explanations of the criteria, and that they have a way to ask for a person to review their score. They can also test to see if a lot of feedback and alerts end up being a waste of time, or if the feedback and alerts are making it easier or harder for them to get back on track. Here, I suggest implementation questions, based on the theoretical framework. This questionnaire does not prove their effectiveness as remedies, and system designing should be done in cooperation with the employees and not just based on a cross sectional coefficient.

7.3. Implications for Future Research

The next study should take the separation of applications and organizational settings into account. An analysis tool for development can be different from the score applied to decide the ranking of the employees and even in that case, they can both be part of the composite today. Independent verification of system functions would help weed out basic "rule based" automation from AI inferential. The proposed resource and demand explanations would be more precisely compared using direct measures of feedback usefulness, measures of workload, transparency and autonomy.

A longitudinal design would include a baseline measure of the exposure and a baseline measure of burnout and/or performance, followed by a measure of burnout at a later point, and/or subsequent supervisor- (or independent) measure of performance. Appropriate multi-level/cluster robust analysis would be possible with company and team identifiers. It would be more useful if temporal separation were greater, but there would still need to be defensible assumptions about confounding to make causal mediation useful. Experimental or well-designed rollout comparisons could be more powerful evidence than taking another data collection with the same one wave questionnaire with a larger sample.

8. Limitations and Conclusion

The scope of the conclusions is limited by three things: First, consider the fact that it does not give chronological order, as the single wave design and common four week recall period used do not. The findings related to poor performances could result in increased monitoring, and vice-versa, in terms of workload that went unobserved, exposure/burnout and performance might be affected. Secondly, the researcher developed items have good internal consistency but need to be construct validated. Perception of exposure is no measure of product technology and self-assessment of performance is not an indicator of actual productivity. Thirdly, the complete case selection and lack of organizational identifiers means that representativeness and dependency checks are limited. The proration analysis does not speak to these more general limitations – only to the one chosen scoring.

But in these limits, the study reveals a negative indirect relationship of ab = −0.090 (95% bootstrap interval [−0.132, −0.055]) based on 342 complete questionnaire records, through burnout. The association stands up when adjusted for and alternate scored. There is no direct positive support for the positive direct hypothesis and the total relationship is unclear. It is therefore a relevant statistical pathway when the HR is considered, and the notion that AI HR tools can either boost efficiency, have a negative impact on overall performance, or induce HR burnout are less substantiated.

References

[1] Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review*, 61(4), 15–42. https://doi.org/10.1177/0008125619867910

[2] Parent-Rocheleau, X., Parker, S. K., Bujold, A., & Gaudet, M.-C. (2024). Creation of the algorithmic management questionnaire: A six-phase scale development process. *Human Resource Management*, 63(1), 25–44. https://doi.org/10.1002/hrm.22185

[3] Demerouti, E., Bakker, A. B., Nachreiner, F., & Schaufeli, W. B. (2001). The job demands-resources model of burnout. *Journal of Applied Psychology*, 86(3), 499–512. https://doi.org/10.1037/0021-9010.86.3.499

[4] Bakker, A. B., & Demerouti, E. (2017). Job demands-resources theory: Taking stock and looking forward. *Journal of Occupational Health Psychology*, 22(3), 273–285. https://doi.org/10.1037/ocp0000056

[5] Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, 14(1), 366–410. https://doi.org/10.5465/annals.2018.0174

[6] Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods*, 12(1), 23–44. https://doi.org/10.1037/1082-989X.12.1.23

[7] Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. https://doi.org/10.1007/BF02310555

[8] Henseler, J., Ringle, C. M., & 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. https://doi.org/10.1007/s11747-014-0403-8

[9] Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40(3), 879–891. https://doi.org/10.3758/BRM.40.3.879