Human resources analytics in practice: a knowledge discovery process

Dr. María Jesús Belizón*

ICADE, Universidad de Comillas & University College Dublin

mjbelizon@comillas.edu

Delia Majarín

Accenture

deliamajarin@gmail.com

Dr. David Aguado

Universidad Autónoma de Madrid & Instituto de Ingenieria del Conocimiento

david.aguado@uam.es

(*) Corresponding author

This article has been accepted for publication in the European Management Review in August 2023. Copy for private circulation.

Human Resources Analytics in practice: a knowledge discovery process

Abstract

Existing scholarship offers a comprehensive understanding of the concept and purposes of human resource analytics (HRA). However, how HRA is carried out in practice in organisations is still under-researched. We examine the practice of HRA through a systematic review across three disciplines, namely, human resource management, business analytics and management information systems while using a process lens: the knowledge discovery process (KDD) model. A hundred and three high-end quality manuscripts were analysed. Our findings show that the scope of HRA is expanding both in the use of HR and business data and certain sophisticated statistical techniques. However, much needs to be done to uncover the measurable impact of HRA on HR and business outcomes.

Keywords: HR Analytics; people analytics; process lens; analytics projects; data-driven decision making.

Introduction

There has been certain scepticism concerning the use of human resource analytics (HRA) and its potential for meaningful impact on HR practices and operations (Angrave et al., 2016; Belizón & Kieran, 2022; Capelli, 2017; Edwards et al. 2022), inter alia, because of a lack of both theoretical foundations and empirical evidence within the emerging field of HRA over the past ten years (Edwards et al. 2022; Margherita, 2021; Van der Laken et al., 2018)

There are three relevant primary questions that scholarship is likely to address when a new field starts to blossom: What, Why and How? In the HR Analytics field: (i) what is HRA? (ii) why is HRA important? and (iii) how do we carry out HRA? What HRA is has extensively been defined in the HR literature (Angrave et al. 2016; Marler and Boudreau, 2017; Minbaeva, 2018). More recently, Belizón & Kieran (2022) view HRA "as the practice of data-driven HR decision-making". They continue explaining that "this practice involves addressing a strategic or operational HR concern by making use of data (HR, business and/or external data) and encompasses the following process: identification of a business problem, research design for the specific HR issue under scrutiny, data management, data analysis, data interpretation and communication, a subsequent action plan and evaluation" (p. 2). Their contribution as well as Ellmer & Reichel (2021)'s conceptualise HRA in practice, both as legitimacy and epistemic alignment *processes* respectively.

How HRA is carried out in practice and ultimately how HRA impacts HR decision-making and the business are questions that still need further clarity (Angrave et al., 2016; Capelli, 2017; Edwards et al. 2022; Ellmer & Reichel, 2021; Huselid, 2018). As Angrave et al., (2016) note "without a focus on praxis, the fog and confusion around analytics is a block to action. Can academics do more to burn through this fog and assist the HR profession to upskill to a new world of strategic analytics-driven HR?" (p. 8). The question at this point then is: how does the HRA process present and how do organisations respond to its different stages? We believe that a key element in HR embracing HRA as a natural supporting function in data-led decision making is to comprehend how HRA is carried out in practice. A recent opening article to a special issue

provides two challenges in this endeavour: one is the fact that HR analytics is not a recognised keyword, making scholarly searches all the more exasperating; second, they point out to the increasing literature review articles of limited scope due to the low number of empirical articles considered in them (Edwards et al., 2022). We add a third reason that can also help mitigate these challenges: a need for cross disciplinary enquiry. There is consensus that data analytics is cross disciplinary in nature (Mortenson et al., 2015) and that its application into the HR function still builds on disciplines other than HR such as business analytics, management information systems (MIS) or data engineering (Angrave et al., 2016; Belizón & Kieran, 2022; Ellmer & Reichel, 2021).

Therefore, in this paper we map the practical process of HRA across disciplines. Drawing from HR, MIS and business analytics scholarships we examine how organisations 'do' HRA in practice and how that process is executed. We, therefore, aim: (i) to identify the different steps of the practical process of HRA and (ii) to build these steps from theoretical and empirical foundations of data-led HR decision-making across disciplines.

This paper, therefore, contributes to existing literature by systematically examining how HRA is carried out in practice, that is, through a process lens. A process narrative allows us to capture the degree of complexity of the HRA process as a whole and also of each stage in particular. Additionally, a step-by-step process lens provides the space to study how organisations have navigated the different stages, as well as the inevitable dynamics throughout the entire process. In sum, this is the first academic article that addresses how HRA is carried out in practice as a process.

HR Analytics in practice: a knowledge discovery process

Increasingly, management phenomena, particularly those resting at the intersection of theory and practice are examined through a process lens and narratives (Langley, 1999; Aldrich, 2001; Salk, 2010; Godfroid and Labie, 2023). This approach not only allows scholars to systematise information but also to theorise in a more granular manner, and it has proven effective

in exploring, observing and identifying the different components, activities and events of a particular phenomenon in management studies (Langley & Truax, 1994; Langley, 1999). The study of HRA as a process is not new (Aguado, 2018; Belizón & Kieran, 2022; Boudreau & Cascio, 2017; Ellmer & Reichel, 2021; Falletta & Combs, 2020; Rasmussen & Ulrich, 2015). As it has transpired to be extremely useful in the nuanced study of multifaceted phenomena, scholars in the HRA field have deployed this prism in order to investigate aspects such as the development of institutional legitimacy of the HRA function, seen as the combined byproduct of cognitive, technological and socio-political legitimacies (Belizón & Kieran, 2022).

Attempts to construct step-by-step models for the practical application of HRA in the HR literature already exists (Aguado, 2018; Boudreau & Cascio, 2017; Cascio & Boudreau, 2010; Falletta & Combs, 2020). A simple but nonetheless useful framework that Boudreau & Cascio (2017) put forward to HR professionals for a greater adoption of HRA is the LAMP model. LAMP stands for Logic, Analytics, Measures and Process. Logic is the articulation of the link and impact between HRA and the business. Analytics involves the appropriate use of statistical techniques. Measures revolve around the systematic generation of useful HR metrics and Process refers to the deployment of communication and project management methodologies for meaningful impact. More recently and drawing primarily from HR scholarship, Falletta & Combs (2020) embarked in a much more granular identification of a HRA process seen as a seven-step cycle or process with a view to helping organisations develop analytical capabilities and ultimately, developing their HRA functions. These steps are to (i) determine stakeholder requirements; (ii) define HR research and analytics agenda; (iii) identify data sources; (iv) gather data; (v) transform data; (vi) communicate intelligence results and (vii) enable strategy and decision-making.

These frameworks share certain similarities as they all capture certain elements of the analytical process. However, although valid and informative, they are primarily based on limited HR scholarship and some practitioners' experiences. They fall short in two fronts. First, they do not draw from a comprehensive and systematic body of knowledge from the HR literature. Second, they exclude contributions from two other scholarly fields that are building blocks of HRA, namely, business analytics and MIS. For these reasons, we visualise the need for a

systematic literature review, encompassing work across disciplines in order to study HRA in practice.

Instead of relying on existing limited HRA frameworks, we turn to the business analytics scholarship in order to find a conventional and consensual analytics model with which we can sign post the HRA process. We find that the Knowledge Discovery Process (KDP), also termed Knowledge Discovery in Databases (KDD), has been used extensively in the fields of medicine, engineering, production, e-business, software, marketing and sales (Anand & Buchner, 1998; Cabena et al., 1998; Cios et al., 2005; Cios et al., 2007; Fayyad et al., 1996; Lee & Cho, 2020). Schmidt & Nan Sun, 2018). These conceptualisations of the KDP have been primarily academic in nature and much less a hybrid between academic scholarship and industry expertise. Yet, KDP has also been applied to purely industry models such as the CRISP-DM (Cross Industry Standard Process for Data Mining) model proposed by industry experts and funded by the EU (Shearer, 2000). KDP entails the sine-qua-non pre-existence of a business issue that needs to be solved (Cios et al., 2007). The KDP model also highlights the different levels of effort and capabilities appropriate for each step. We selected the Knowledge Discovery Process (KDP) framework proposed by Cios et al., (2007) due to its hybrid nature, combining both solid theoretical underpinnings and industry expertise. This framework in particular proposes a methodological development in six stages, namely, "understanding of the problem domain, understanding of data, preparation of the data, data mining, evaluation of the discovered knowledge, and use of the discovery knowledge" (Cios et al., 2007; p. 22). The authors view this process as sequentially executed but iterative in nature, including 'feedback loops' and review iterations exercises. Our choice for the KDP framework does not constitute a total departure from previous models. However, we believe that KDP constitute a solid, sequential and systematic way to represent the analytical process.

In our attempt to examine how organisations respond to the HRA process, we proceed to lay out the HRA process in alignment with the stages of the KDP proposed by Cios et al., (2007). In relation to the first stage of the KDP model 'understanding of the problem domain', we set out to answer the question 'What workforce issues are identified?'. Regarding the second KDP stage

'understanding of the data', we respond to 'What type of data is used in these HRA projects?' The third KDP state entails the 'preparation of the data' and our question reads as follows: 'How are technologies deployed to access and manipulate data in HRA projects?' In our attempt to shed light onto the fourth stage pertaining to 'data mining' per se, we pose: 'What analytical techniques are used in HRA projects?'. Regarding the fifth KDP stage, 'evaluation of the discovered knowledge', we identify the different strategies used for outcomes assessment by asking: 'how are statistical outputs interpreted?'. Finally, referring to the last KDP stage, 'use of the discovered knowledge' we set out to explore 'how HRA insights are being used for decision-making in the HR function together with the main limitations inherent in HRA projects'. Table 1 illustrates the six stages and corresponding questions.

Table 1: The KDP model and implications for HRA

Stages	Focus	Key questions
Stage 1: How is the problem domain understood?	Domain experts work towards defining the business problem and goals. They possess extensive domain-specific and business acumen and are knowledgeable of key issues, capabilities, and potential solutions. A description of the project is prepared, including its limitations and key people involved. In this phase, there is a preliminary understanding of the data and analytics tools that might be needed.	What workforce issues are identified?
Stage 2: How is the data understood?	This phase encompasses identifying HR and business metrics, the population, sample size needed and how the data will be collected. It seems valuable at this stage to control for potential data problems such as completeness, duplicated cases, missing values, possibility of gathering all data needed, etc.	What type of data is used?
Stage 3: How is the data prepared?	This step includes data collection and preparation. Pilot sampling, complete sampling, data cleaning, preliminary descriptive analysis, correlations and significance testing. This may entail interrogation of a diverse range of databases, and subsequent merger.	How are technologies deployed to access and manipulate data?
Stage 4: How is the data mined?	In this phase the data analyst/scientist carries out the data analysis according to plan in order to meet the project objectives. The data mining can involve descriptive and informative statistical techniques as well as predictive techniques and algorithms.	What analytical techniques are used?
Stage 5: How is the discovered knowledge evaluated?	This step is dedicated to the interpretation of the statistical output and elucidating whether the results are relevant to the business problem and how significant it is. A blend of business and technical expertise is needed at this phase in order to define the impact of the discovered knowledge. Statistical limitations are also taken into account at this point.	How are statistical outputs outcomes interpreted?
Stage 6: How is the discovered knowledge used?	The final stage involves translating the results to business terminology, devising 'how to tell the story' (storytelling) to relevant stakeholders and developing an action plan that would address the project aims. A plan to monitor and review the implementation of interventions or nudges is elaborated.	How are HRA insights used for decision making?

Source: Adapted from Cios et al. (2007).

Addressing these questions drawing from scholarship across three disciplines offers a novel contribution both from an academic and industry perspective. Hence, the present work constitutes a road map of the specific process stages followed in HRA projects.

Research Process

The search for the literature was carried out following the traditional procedure and methodology for systematic literature reviews (Rob & Denyer, 2012). As Edwards et al., (2022) points out, limited existing empirical contributions along with a shortage of qualitative case studies of HRA projects in the emerging HRA field can substantially constrain the development of a robust literature review. For these reasons and in order to ensure methodological validity, we embark upon a comprehensive search encapsulating both theoretical and empirical published literature across three disciplines: human resource management, business analytics and MIS. We would like to note at this point that not every journal article will reveal how organisations respond to each one of the six process stages. However, these articles in combination will certainly contribute to understand the process specifically and as a whole.

With these criteria in mind, we undertook a systematic search in Web Of Science (WOS) using key word combinations that would allow us to find the greatest number of possible articles in the HR analytics scholarship primarily across three mainstream disciplines, namely, human resource management, business analytics and MIS. We purposely combined a diverse range of terms generally used to refer to HR Analytics (i.e. workforce planning, people analytics, workforce analytics, talent analytics, HR Analytics and human capital analytics), along with keywords referring to human resource management (i.e. human resource management, human capital, workforce, employee) and business analytics (i.e. analytics, big data, data mining, data science, business analytics, business intelligence). Our final keyword strategy was the following:

((((TS=analytics OR TS="big data" OR TS="data mining" OR TS="data science" OR TS="business analytics" OR TS="business intelligence" OR TS="artificial intelligence") AND

(TS=HR OR TS="human resources" OR TS=workforce OR TS="human capital" OR TS=HRM)) AND (LA==("ENGLISH")) OR TS="People analytics" OR TS="workforce

analytics" OR TS="talent analytics" OR TS="HR analytics" OR TS="human resources analytics" OR TS="human capital analytics" AND (LA==("ENGLISH")))

Having included only those articles published through the English language, we obtained a first outcome of 2,441 articles. Subsequently, we filtered by journal articles and date, including those articles published between 2010 and 2022. Finally, we obtained a total of 2,007 articles. The three authors proceeded to read all titles in order to carry out a first filter, allowing us to narrow down the volume by discarding those articles that fell outside our criteria. Our criteria included all articles that would help us understand the KDP stages of HRA projects individually and as a part of a whole. Articles published in the HRA space that did not relate to the KDP process were therefore excluded. We selected 248 articles that initially pertained to HR Analytics topics according to our criteria. We, then, proceeded to read the abstract in order to determine which one of these articles can be considered a suitable study for this analysis. We identified 163 studies, which we have classified into five clusters primarily based on the methodology used in these articles. This classification, which was agreed upon by consensus, comprises the following types of articles:

a. Theoretical and literature review articles in the HRA field (Theoretical);

b. Empirical articles based on a qualitative methodology primarily exploring the adoption and impact of HRA (Empirical: qualitative);

c. Empirical articles studying how private or state-owned organisations or industries obtain HR analytics insights through the traditional KDP sequence (Empirical: KDP);

d. Articles developing novel mathematical models that are applied to specific HR domains and that make use of HR data to be tested (Mathematical model);

e. Articles developing technological tools or applications that are applied to specific HR domains and that are validated with HR data (Technological application).

The 163 articles were assessed at least by two of the authors acting as judges, and those articles that proved problematic were reviewed by the three authors. By problematic we mean articles that having been included in our initial search did not talk to any of the stages of the KDP framework. Finally, reviewing carefully each of these articles we selected a final sample of 103

articles, all above 3 points in CiteScore in order to guarantee certain scholarly quality. There was one exception, namely, Rasmussen & Ulrich (2015) [1], included due to the relevancy of their work. Figure 1 shows the results found by classifying selected articles based on our analysis criteria. The full categorisation of the peer-reviewed journal articles analysed in this sample can be found in Appendix 1.

Figure 1: Systematic Literature Search Prisma Flowchart

---
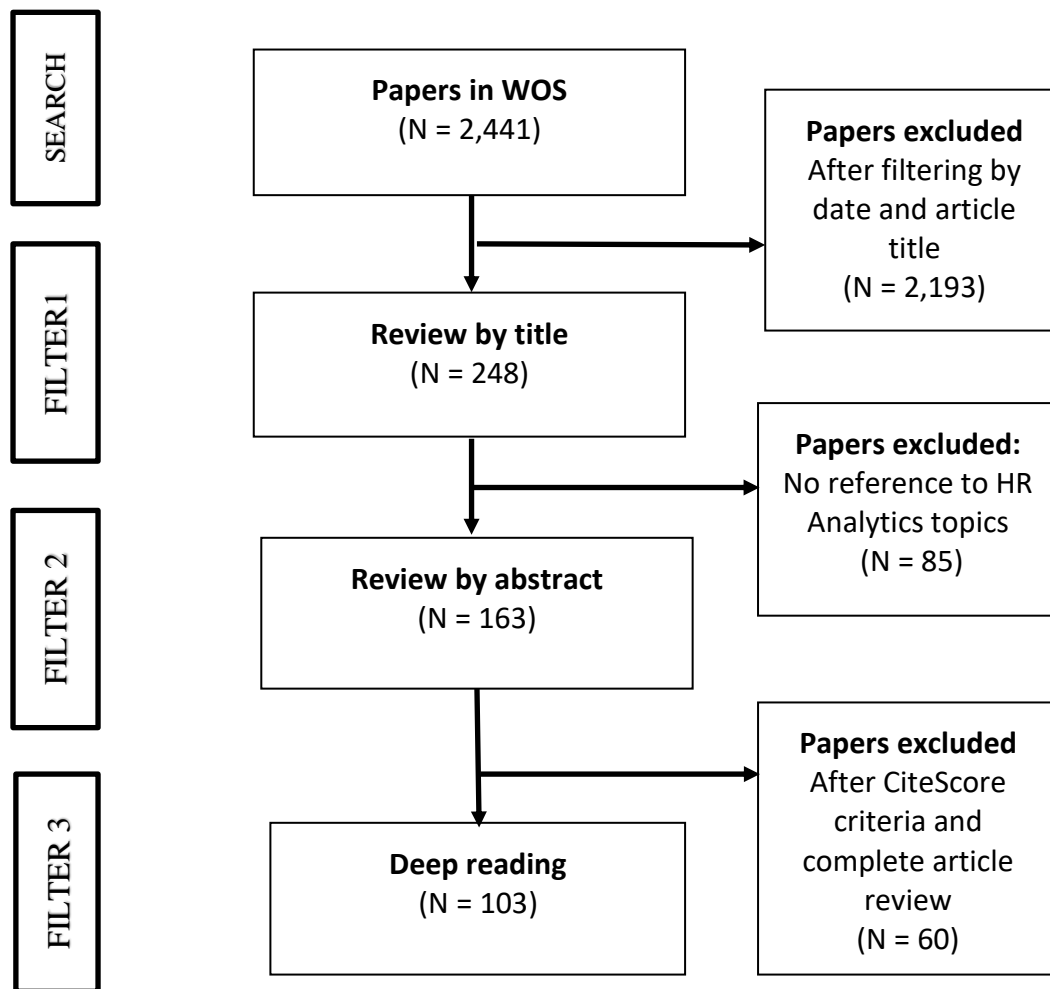
[1] This paper was also suggested to be included by the reviewers.

Empirical Analysis

1. **What workforce issues are identified?**

How the problem domain for HRA is understood in organisations is crucial and addresses the manner in which core workforce issues are identified and made suitable for HR analysis. The operational objectives of the different studies analysed fluctuate pretty consistently among four top HR challenges, namely, employee turnover / attrition, workforce planning and scheduling, talent acquisition / recruitment and employee performance.

Turnover / Attrition is the topic most discussed focusing on predicting turnover and identifying the reasons employees may have to leave the company so that leadership can take action (Alshareef et al., 2020; Fan et al., 2012; Ballinger et al., 2016, Liu & Raghuram, 2021; Rombaut & Guerry, 2018; Saradhi & Palshikar, 2011). Additionally, some articles analyse how to foster employee retention avoiding the so-called brain drain (Rombaut & Guerry, 2020; Schofield, 2018) or discuss and compare different attrition algorithms to increase model discrimination (Speer, 2021).

Workforce planning ranks second place with a strong emphasis on either capabilities or scheduling. Some articles focus on resource and people allocation for projects based on job roles, skills and capabilities to match the organisation needs (Andre et al., 2011; Berk et al., 2019; Fareri et al., 2020) while others focus more on workforce scheduling to optimise the shift assignments (Becker, 2020; Simeunovic et al., 2017). Workforce planning is a recurrent challenge in most of organisations, especially consultancy firms where resources have higher rotation and projects are very diverse and demand very specific skills (Hargaden & Ryan, 2015; Llort et al., 2019; Shahbazi et al., 2019).

Talent acquisition and employee recruitment also appears as frequent as performance in our literature review, with various HR Analytics applications focused on making the recruitment process more efficient and objective by trying to reduce human biases (Bohlouli et al., 2017; Faliagka et al., 2012; Garg et al., 2021; Guo et al., 2021; Hickman et al., 2021; Maghsoodi et al., 2020; Michelotti et al., 2021; Necula et al., 2019; Pessach et al., 2020). Additionally, a

handful of articles are concerned with the following HR topics: employee engagement (Garg et al., 2021; Guo et al., 2021; Van der Togt & Rasmussen, 2017), training (Ginda et al., 2019), readiness to change (Shah et al., 2017), employee / customer satisfaction (Barnes et al., 2020; Kang et al., 2019), compensation (Jafari et al., 2020), wellbeing (Greasley & Thomas, 2021), and there is also one article focused on skills analysis to identify market competitors' workforces (Liu et al., 2020). There are also two articles that we categorise as "general HR" as they focus on applications affecting several HR domains. For instance, one of them proposing an ontology to evaluate human factors by modelling key performance indicators (KPIs) associated with digital footprints (Gelbard et al., 2018).

Employee performance constitutes the third *hot topic* and it is explored at individual level (Hermawati & Nasharuddin, 2017; Luo et al., 2019; Nicolaescu et al., 2020), team level (Wang & Cotton, 2018) and organisational level. For instance, the work of Simón & Ferreiro (2018) suggests that certain aspects of the workforce that are worth exploring through HRA come to the fore because they have been identified and are seen as antecedents of organisational performance. Thus, the authors relate the different HR KPIs to certain sales performance indicators in Inditex, the Spanish multinational fashion retail group. Generally, the studies analysed are good at identifying those workforce aspects at the centre of their business problem. However, they are not as good at identifying those organisational effectiveness KPIs on which the HRA project intends to impact. In this sense, there are very few explicit works that directly link HRA with operational performance (Berk et al., 2019; S Luo et al., 2019; Simón & Ferreiro, 2018; Wang & Cotton, 2018). A summary of these workforce issues is captured in Table 2.

Table 2. HR problem domains addressed with HRA

| Workforce planning and scheduling |
| *Skills allocation for projects* |
| *Workforce scheduling optimisation* |
| Employee attrition and turnover |
| *Employee retention* |
| Employee performance |

Talent acquisition/Recruitment
Recruitment efficiency optimisation
Labor market skills analysis
Employee engagement
Employee voice channels
Training and development
Readiness to change
Customer satisfaction
Compensation
Wellbeing
General HR

2. What type of data is used in HRA projects?

The analysis of the varied nature of the data used in HRA projects shows a very interesting picture in which the use of HR traditional data collected through longstanding procedures (i.e. Shah et al., 2017) coexists with the analysis of work-related data obtained through simulation procedures (i.e. Hargaden & Ryan, 2015). Our analysis points to the existence of six principal groups of data used: (a) data that reflects knowledge, skills, abilities and other characteristics (KSAO's); (b) employee work-related records; (c) employee outcomes; (d) human resources outcomes (key performance indicators and HR practices); (e) organisational outcomes (operational data and organisational performance); and (f) external data.

Employee KSAOs related data includes (1) data on the skills and technical skills necessary for carrying out the tasks needed in organisations (Bohlouli et al.,2017; Fareri et al., 2020; Hargaden & Ryan, 2015; Necula et al., 2019; Nicolaescu et al., 2020; Shahbazi et al., 2019). (2) data related with employee psychological capital and traits (Andre et al., 2011; Faliagka et al., 2012; Guo et al., 2021; Hickman et al., 2021; Maghsoodi et al., 2020; Michelotti et al., 2021; Shah et al., 2017). (3) Leadership effectiveness data (Hermawati & Mas, 2017; Pessach et al., 2020) used for example to explore transglobal leadership in order to study its influence on the performance of managers based in cooperatives in Indonesia (Hermawati & Mas, 2017). (4) data related to demographics, education and experience (e.g. Alshareef et al., 2020; Fan et al., 2012; Faliagka et al., 2012; Maghsoodi et al., 2020; Pessach et al., 2020; Rombaut & Guerry, 2018);

and (5) data related with the employee position in the informal network of the organisation (Ballinger et al., 2016; De Laat & Schreurs, 2013).

Work-related records are also used in several of these studies (Alshareef et al., 2020; Becker, 2020; Esmaeilzadeh et al., 2016; Fan et al., 2012; Jain et al., 2020; Liu & Raghuram, 2021; Llort et al., 2019; Luo et al., 2019; Maghsoodi et al., 2020; Schofield et al., 2018; Shahbazi et al., 2019; Rombaut & Guerry, 2018; Rombaut & Guerry, 2020; Simeunovic et al., 2017; Wang & Cotton, 2018). This type of data includes records on work schedule, workforce availability, workload, client organisation designation, employee location, delivery location, projects, average monthly hours, salary, work home distance, sickness and objectives. The use of email repository in a business school by Raman et al., (2018) is worth mentioning, through which they identified behavior than can help in predicting attrition.

A third type of data used is concerned with employee outcomes. It includes data related to the employee attitudes and perceptions linked to factors such as salary, job promotion, organisational commitment and loyalty, job satisfaction, pay satisfaction, social support, wellbeing or engagement (Alshareef et al., 2020; Faliagka et al., 2012; Fan et al., 2012; Greasley & Thomas, 2020; Hermawati & Mas, 2017; Jain et al., 2020; Kang et al., 2019; Nicolaescu et al., 2020; Shah et al., 2017; Van der Togt & Rasmussen, 2017; Yuan et al., 2021). And data related to employee behavior such as performance, organisational citizenship behavior and turnover (Alshareef et al., 2020; Fan et al., 2012; Hermawati & Mas, 2017; Nicolaescu et al., 2020)

Another type of data is concerned with HR outcomes that includes HR KPIs and data related to HR practices. For example, Simón & Ferreiro (2018) use aggregate measurements relating to absence, average workday, voluntary turnover, permanent contracts, workday reductions, job tenure and age. Whereas Rombaut et al., (2020) use data related to the grade that an employee has received different retention strategies: rewards through compensation, recognition, training and flexibility (Arfae et al., 2022). A differential aspect of these two cases is in the analysis level of measurement they have used: i.e. group level in the case of Simón & Ferreiro (2018), and individual level in the case of Rombaut et al., (2020). Additionally,

Schiemann et al., (2018) include employee turnover costs (e.g. hiring and training) and employee productivity as part of their HR metrics.

In relation to organisational outcomes metrics we found the use of data relating to the operations of the organisation (Berk et al., 2019; Hargaden & Ryan, 2015; Simón & Ferreiro, 2018); security risk and sales (Van der Togt & Rasmussen, 2017); bill rates, labour costs, client request (location and skills demanded), transfer costs (Berk et al., 2019); opening hours, store location, size, years in operation, efficiency (Simón & Ferreiro, 2018); projects by quarter, project extensions, skills required by project, and hours of skills required by project (Hargaden & Ryan, 2015); client experience (Barnes et al., 2020, Kang et al., 2019) and profit levels (Schiemann et al., 2017).

Additionally, we found two studies that use data external to the organisation. Simeunovic et al., (2017) uses daily weather information to obtain better results in forecasting the demand for professional services produced in the offices of the Public Utility Company billing department and thus optimise the workforce scheduling process. Nicolaescu et al., (2020) uses data from job websites whereas Conway et al., (2019) uses Twitter to try to inference, unsuccessfully, employee voice in social media, Guo et al., (2021) extracted engagement data from Glassdoor in an employer branding use case and Platanou et al., (2017) uses network data coming from open sources such as LinkedIn, Twitter or YouTube to propose a framework for text analysis. This is in line with Strohmeier & Piazza (2013) who argued that around 10% of the data used in their review sample were obtained through web mining or social media scrapping.

How about big data?

The actual use of data under the umbrella-term 'Big Data' deserves special consideration. Most HR data included in the literature cannot be described as big data. However, new technological devices are allowing organisations to keep track of high volumes of employee data. For example, Ontrup et al., (2022) propose six types of data from digital platforms that speaks to the proactivity of work teams, some of which are viewed as big data: (1) word-related information on communication regarding content (what is being said?) and tone (how are things said?); (2)

16

quantitative data entered by employees; (3) meta-data (information about data); (4) passive data (digital trace data of technology use from a specific user); (5) behavior (data of movement, posture and gesture); and (6) biometrics (biological data such as health data, fingerprints, voice, etc.). The inclusion of measures such as biometrics and behavior undoubtedly requires a serious and cautious reflection on the ethical and privacy issues associated with the use of this type of data. In this vein, Leonardi & Contractor (2018) emphasised the idea of using data that reflects the interactions between individuals and groups, beyond the measurement of individual biometrics through the use of analytical technologies as Organisational Network Analysis (ONA).

A fundamental issue in this context is the lack of data quality reported by organisations (Tambe et al., 2019, Dahlbom et al., 2020), which, among other issues, has an important impact on the development of machine learning models and algorithms. This is due to the fact that when low-quality data is used the resulting learning algorithms will also turn out to be of poor quality (as they say 'garbage in, garbage out'). On the other hand, oftentimes there is solid data quality coupled with scarce observations, which also hampers the statistical quality of the machine learning-based algorithms that are developed. In any case, regarding the data challenges in HRA, there is always an opportunity in using data applied to human resources needs or problems. In this line, Zhang et al., (2021) emphasises the use of "small data" to generate "smart insights". Table 3 condenses the types of data used in HRA projects.

Table 3. Data used in HRA projects

Data on knowledge, skills, abilities and other employee characteristics (KAO's)	HR outcomes (KPIs and HR practices)
Professional and technical skills Psychometrics Leadership effectiveness Demographics, education and work experience Employees' position and influence in informal networks	Absenteeism Average workday Voluntary turnover Type of contracts Workday reductions Job tenure Employee turnover costs (hire & train) Employee productivity
Employee work-related records	Organisational outcomes
Work schedule Workforce availability Workloads	Operational data Security risk Sales volume

Client organisation designation	Bill rates
Employee location	Labour costs
Delivery location	Client request (location and skills)
Projects	Transfer costs
Average monthly hours	Opening hours
Salary	Store location
Work-home distance/Commuting distance	Store size
Sickness	Store years in operation
Employee's objectives	Store efficiency
Email repository	Projects by quarter
	Projects extensions
	Client experiences
	Profit margin levels
Employee outcomes (employees' perception)	**External data**
Salary levels	Weather information
Promotions	Job websites
Organizational commitment and loyalty	Social networks websites
Social support	
Employee wellbeing	
Employee engagement	
Employee outcomes (organisation's perception)	
Employee performance	
OCB	
Employee turnover	

3. How is the data prepared? How are technologies deployed to access and manipulate data in HRA projects?

Regarding how technology is used for data access and manipulation, there are several approaches. Firstly, nearly half of our sample uses existing data and basic computer technology to access proprietary databases in which the management of medium and large volumes of data is commonplace. Strohmeier & Piazza (2013) show evidence to argue that organisations make use of HRIS as repository of their data but also of purpose-built HR data bases or data lakes for specific projects. These solutions do not have to be an extension or plugged into the main HRIS or proprietary system. Therefore, data manipulation does not prove a complex undertaking as the samples found are relatively small. In this vein, we observe more than 3 million records in the case of Wang & Cotton (2018) or Kifor et al., (2021) and more than 5 million records in the case

of Luo et al., (2019). More modest data volumes are also used, such as the case of Pessach et al., (2020) using 700,000 observations in recruitment during the decade between 2000 and 2010. Only two articles use a considerable sample size, as it is the case of Fareri et al., (2020)'s sample from the Department of Transportation in the US (USDOT), which includes 17,800 employees; and the case of Avrahami et al., (2022) that uses a dataset with information on 700.000 employees recruited by the organisation over a period of ten years. Smaller sizes are indeed employed in Rombaut & Guerry (2020)'s sample of 1,606 employees from a Belgium company, Falluci et al., (2020)'s sample of 1,435 employees, in Simón & Ferreiro (2018) in their analysis of 244 stores or in Safarishahrbijari (2018) using the email repository of only 126 faculty members but analysing panel data gathered in the previous six years. What becomes clearer is that data volume is directly associated to the level of analysis under investigation, the nature of the study, the size of the organisation (i.e. size of the organisation in terms of number of employees, teams, managers, stores, countries, etc.) and the timeframe covered in the study.

Secondly, we identify cases that collected ad-hoc data and used specific technologies for it with the aim to validate a mathematical model or a new technological application. For example, Fareri et al., (2020) used technology for analysing text contained in Whirpool's records of employees' CV and job descriptions, employing Natural Language Processing (NLP) technologies to label them. Another example within this cluster is Gelbard et al., (2018), who developed an ontology through text analysis in order to measure core KPIs in a specific email corpus from Enron. They focused on assessing KPIs around creativity, innovation, service quality, efficiency and effectiveness, and extra-role behaviors from 250,000 emails exchanged among 151 employees. Other articles extract the data from public data sources like Twitter or LinkedIn using web scrapping techniques (Conway et al., 2019; Guo et al., 2021; Platanou et al., 2021). What transpires from our sample is that most cases using data purposely extracted from other sources other than HR proprietary systems are geared towards testing and validating mathematical models or technological applications. Thirdly, a relevant proportion of our sample collected data through traditional questionnaires.

Finally, two articles employed data exclusively from secondary sources. This is the case of Liu et al., (2020) analysing approximately 90,000 employees from the Standard & Poor 100 Index companies and the case of Bechter et al., (2022) that used data coming from the European company survey (ECS) across 28 EU member states with an estimation of 20,411 firms included.

We observe that data used to test and validate mathematical models applied to HR domains generally do not draw from HR metrics stored in proprietary systems while the latter is mainly used for HR projects which follow the KDP sequence. Table 4 synthesises how data is sourced across different technological infrastructure.

Table 4. Data sources and data integration

Mainstream proprietary HRIS data
Ad-hoc data collection through questionnaires and surveys, web scrapping, or HR platforms data incorporated into a purpose-built HR database or data lake
Ad-hoc secondary sources merged and incorporated into a purpose-built HR database or data lake
Purpose-built HR data warehouses or HR data incorporated to the central business intelligence data warehouse

4. What analytical techniques are used in HRA projects?

In his literature review, Zhang et al., (2021) identified eleven big data methods that can be used effectively in HRM research: structural equation modelling (SEM) trees, classification analysis, genetic algorithms, prescriptive analytics, latent semantic analysis (LSA) and differential language analysis (DLA), singular value decomposition (SVD), the experimental approach of event studies, Social Network Analysis (SNA) approach, clustering techniques and deep learning techniques. These techniques are slightly more complex than those reviewed by

Strohmeier & Piazza (2013) a decade earlier, which included univariate and multivariate analysis such as decision trees, neural network, discrimination analysis, cluster analysis or regressions.

When searching for the analytical techniques used in these studies, we observed that a good deal of articles move between descriptive and linear models, namely, regression models (Ballinger et al., 2016; Bohlouli et al., 2017; Du & Li, 2020; Fallucchi et al., 2020; Jain et al., 2020; Jafari et al., 2020; Necula et al., 2019; Schofield et al., 2018; Simón & Ferreiro, 2018; Rombaut & Guerry, 2018; Rombaut & Guerry, 2020, Saradhi & Palshikar, 2011), especially found in KDP articles and turnover case studies. Structural Equation Models (SEM) are also popular when surveys need to be analysed and constructs validated (Alshareef et al., 2020; Hermawati & Mas, 2017; Liu and Raghuram, 2021; Shah et al., 2017; Wang & Cotton, 2018).

The second most used models are those descriptive techniques combined with some form of optimisation (Andre et al., 2011; Becker, 2020; Berk et al., 2019, Chen et al.,2022; Corominas et al., 201; Llort et al., 2019; Shahbazi et al., 2019; Pessach et al., 2020), linear programming (Hargaden & Ryan, 2015; Llort et al., 2019) and genetic algorithms applied to workforce scheduling (Algethami et al., 2019). They appear to be popular when applied to workforce planning case studies where the main goal is to identify who will need to be moved to a vacancy considering project constraints or even some use-cases in this field are considered as "prescriptive" as they calculate different workforce planning scenarios and recommend the most "optimised" scenario so that decisions can be taken efficiently (Berk et al., 2019).

Clustering or naïve bayes model occupied the third place in KDP articles but also in technology applications or mathematical models for turnover challenges, talent acquisition or workforce planning where a classification may help (Bohlouli et al., 2017; Esmaeilzadeh et al., 2016; Fan et al., 2012; Fallucchi et al., 2020; Fareri et al., 2020; Gelbard et al., 2018; Maghsoodi et al., 2020; Liu et al., 2020).

Additional advanced techniques are appearing more and more frequently in recent articles (from 2018 to 2020). Natural language processing (NLP) is generally applied to extract valuable insights from surveys about employee and customer satisfaction (Barnes et al., 2020; Guo et al., 2021) or to identify and classify employees' skills in workforce planning or talent acquisition

challenges (Fareri et al., 2020; Faliagka et al., 2012; Gelbard et al., 2018; Liu et al., 2020). We have also found that organisational network analysis (ONA) is applied to represent and calculate distances between employees' skills (Simeunovic et al., 2017), predicting employee compensation (Jafari et al., 2020) or for visualising informal professional development networks (De Laat & Schreurs, 2013).

In relation to more advanced analytical techniques, we also recognised some articles with prediction goals applied to performance (Nicolaescu et al., 2020), compensation (Jafari et al., 2020), talent acquisition (Pessach et al., 2020), learning paths (Ginda et al., 2019) or predicting who the successful recruitment candidate will be according to the company context or even predicting labour market competition using machine learning techniques as multilayer perceptron (a type of artificial neural network) or convolutional neural network (Liu et al., 2020). Comparing the performance of advanced analytical techniques like deep learning with 'humbler' approaches like general linear models is also common. Arfae et al., (2022) compares the performance of neural networks models vs decision trees for analysing the impact of different training paths in employee performance, being decision trees selected as preferred algorithm for this use-case. Yahia et al., (2021) compare a set of classification predictors for understanding which one is more beneficial in classifying churners and non-churners using decision trees, logistic regression, support vector machine models, random forest and three different deep learning models. Pratham & Ow (2020) analysed the effectiveness of Bayes' theorem compared with artificial neural network for predicting performance in candidates.

We found a substantial lacuna when it comes to identifying the programming language used in the analysis of HR data, with only Garg et al., (2021) mentioning the use of python libraries for recruitment and selection purposes in order to match job position with suitable candidates. In any case, as Dahlbom et al., (2020) point out, there is a significant difficulty in moving from basic analytical techniques to the most advanced ones. Finally, in addition to benefiting from the techniques coming from the field of machine learning, HRA can use and benefit from the techniques and methodologies used in the field of insider econometrics (Larsson & Edwards, 2021). Table 5 gathers those analytical techniques employed in HRA projects.

Table 5. Statistical techniques and HRA projects

Structural equation modelling (SEM)
Regression analysis
Classification analysis (clustering techniques, Naïve Bayes model)
Genetic algorithms
Decision trees
Natural language processing (NLP)
Latent semantic analysis (LSA)
Differential language analysis (DLA)
Singular value decomposition (SVD)
Organisational / Social network analysis (O/SNA)
Discrimination Analysis
Machine learning (ML) techniques
Multi-layer perceptron ANN
Convoluted neural networks

5. How are statistical outputs carried out and how are the outcomes interpreted?

The evaluation and assessment of the information found within the data with respect to the overall HRA project are crucial issues. Reinforcing the results of the systematic review of forecasting modelling published by Safarishahrbijari (2018) where less than a half of articles have evaluated the validity of the developed models, many of our reviewed papers did not thoroughly evaluate the statistical output and subsequently, did not unravel the potential implications and impact on the particular organisation. Instead, they tended to focus only on statistical indicators when contrasting a trained model on the data collected. This was the case of Shah et al., (2017) who used regression analysis to test the impact that variables such as salary have on job satisfaction and how this mediated the relationship with readiness to change.

Referring to statistical validation, we would like to emphasise that besides the usual multivariate significance tests, articles using Machine Learning techniques normally include the analysis of confusion matrixes where training data and testing data are validated (Jafari et al., 2020). In contrast, statistical validation in other studies is carried out by comparing results with those of a Gold Standard or 'judge' or a person asked to validate whether the results of the analytical model make any sense. Bohlouli (2017)'s work developed a model to predict the candidate's quality and their suitability for jobs based on their competency profile and used a manager to verify that the evaluation was correct.

A third evaluation strategy was to replicate the study and the results obtained in other samples or scenarios. Simón & Ferreiro (2018) predicted performance in Inditex stores where they decided to 'test the waters' by running parallel analyses for all the stores throughout the entire chain in Spain. In addition, they have planned an initiative to collect more data from different countries and continue to learn and improve the model.

Apart from the theoretical contributions on this issue in recent literature reviews (Fernandez & Gallardo-Gallardo, 2020; McCartney & Fu, 2022), a few empirical qualitative studies shed some light on the challenges of statistical output translation and storytelling abilities needed to communicate a HRA project's outcomes effectively (Belizón & Kieran, 2022; Ellmer and Reichel, 2021; Fu et al. 2022; McCartney & Fu, 2022). A misalignment of epistemological objects among HRA stakeholders and a lack of cognitive legitimacy of HRA within HR and the broader organisation are significant antecedents of the disconnect between HR Analysts and other HRA and HR professionals and, indeed, the trigger of additional institutional work towards transmitting results and putting forward a convincing action plan (Ellmer & Reichel, 2021; Belizón & Kieran, 2022). Fu et al.; (2022) digs even deeper to uncover how HR Analysts need to engage in both showcasing and curbing to get their point across successfully. Table 6 encapsulates how statistical outputs is interpreted.

Table 6. Validation and interpretation of results in HRA projects

Multivariate significance tests
Analysis of confusion matrixes
Comparing and assessing results with a 'Gold Standard' or with a 'judge'
Replicating with different samples, employee segments or scenarios

6. How are HRA insights used for decision-making?

Even though Chalutz Ben-Gal (2019) argued that most HRA theoretical and empirical articles demonstrated medium or high levels of return on investment (ROI), in the 103 articles analysed we found little reference to how findings are used to inform data-driven decision-making in HR, or in other functions. Nearly all of the sample regurgitated the potential of their study in terms of theoretical and/or managerial implications. Most studies mainly allude to the possibility and theoretical benefits of reusing their methodology, as in the case of Wang & Cotton (2018) "this methodology could also be applied to the study of how strategic and support roles and individual careers evolve and change over time and the effect this has on team performance, team identity, and shared cognition (p. 775)". In this vein, only two papers dare to name financial and cost-effective strategies as potential outcomes of generating action plans based on their findings but these have not been measured or even explained further (Faliagka et al., 2012; Shanbazi et al., 2019). Perhaps, as Holwerda (2021) argues, companies which truly leverage HRA in a tangible way simply will not share this confidential information with competitors for obvious reasons. Or maybe this premise is somewhat too optimistic.

There were only five exceptions to this general rule. Chief among these is Aral et al. (2012) that through a longitudinal study of 189 firms, found how HRA, pay for performance and information technology has a complementary impact on business productivity. Xu & Li (2021) studied nurses' satisfaction before and after the application of optimisation models to workforce scheduling. The results showed that the target value of the nurse scheduling model was 43.43%

lower than the previous manual scheduling target value, the salary cost was reduced by 10.8% and the nurse´s satisfaction with the shift increased by 35.24%. Another exception is Simón & Ferreiro (2018), who provide a very detailed account of the extent to which the HR department took stock of the findings for driving practices and processes. Firstly, the HR department designed a knowledge-sharing process to share good practices from successful and productive stores to those less so. Secondly, they explained that other measures were taken, however varied, across the two chains included in the study. An illustration of this was the formation of discussion groups in one of the chains to reflect on the findings and brainstorm different action plans that could be effective and viable. This academic-industry collaboration had an impact and continuity on the company, and they appear to have been committed to collecting data in several countries in order to be able to apply what they have learnt from their previous work. Rasmussen & Ulrich (2015) offer a similar example of an impactful analytical model to identify what shapes performance differentials across rigs. This organisation employs longitudinal data on leadership effectiveness, competencies, and health and safety indicators in order to measure impact on operational performance and customer satisfaction. Finally, Schiemann et al. (2018) uncover a case study of a real fast-food organisation where the People Equity framework is applied to measure the impact of human capital on HR and organisational outcomes, specifically employee turnover, employee productivity, restaurants sales volume and restaurant profitability.

In any case, different studies reveal two fundamental issues for the use of insights generated by HRA. First, the idea that the mere existence of data does not guarantee that their analysis can become actionable knowledge for the organisation to generate value (Levenson & Fink, 2017). In this vein, Schiemann et al., (2018) argue that vast amounts of information are overwhelming for managers and that simplifying HR metrics through a model is a more user friendly and effective way to make HR more data-driven and to generate a measurable impact. As Ulrich & Dulebohn (2017) suggest, the connection between HR practices and operations and organisational outcomes has significant difficulties, which is why it is necessary to look for intermediate impacts of HR work. These intermediate impacts may revolve around HR outputs

and targets such as individual skills (e.g., competence, commitment, contribution), organisational capabilities, and leadership quality.

This begs the question of how we can turn good data into good decisions (Ulrich & Dulebohn, 2017). HRA generates value for the organisation to the extent that professionals are capable of transforming valuable epistemic objects (i.e. questions) into technical objects (i.e. solutions) (Ellmer & Reichel, 2021). Ultimately, the central mechanism through which HRA builds value is the epistemic alignment that includes the performance of three practices: (a) boundary spanning; (b) customisation dashboard; and (c) speaking a language of numbers (Ellmer & Reichel, 2021). In addition, the orientation of the HRA team also has an influence on the usability of the insights. As Jorden et al., (2021) shows, when HRA teams have a strong orientation towards customisation (meeting their internal clients' needs) and action (making data insights available for decision making), there is greater scope for impact. Second, the existing evidence about the necessary involvement of managers to generate changes in the organisation from the insights produced through HRA. The influence of HRA on organisational performance is mediated by the decision-making carried out by managers (Fu, 2022; Margherita, 2022; McCartney & Fu, 2021; Samson, & Bhanugopan, 2022; Ulrich & Dulebohn, 2017). They are the ones who decide to take action (or not) based on the evidence that data analysis provides.

In sum, the broad problem of HRA impact has been extensively discussed in several theoretical and empirical articles. Some of these are concerned with impact on decision making (Bechter et al., 2022; Boudreau and Cascio, 2017; Strohmeier et al., 2022), while others focused on factors affecting the lack of impact such as the lack of people's capabilities, technology integration, data quality or leadership support, which eventually hamper HRA effectiveness, legitimacy and adoption (Belizón & Kieran, 2022; Ellmer & Reichel, 2021; Giermindl et al., 2022; Gurusinghe et al., 2021; McCartney & Fu, 2022; Peeters et al., 2020; Shet et al., 2021). Special merit should be credited to those academia and industry collaboration articles which bring the research rigor, the case study exploration and the empirical analysis to life in very tangible and powerful ways (Schiemann et al., 2018; Simón & Ferreiro, 2018).

Finally, although the ethical dimension of HRA appear transversally in some stages of the process, its importance is more salient when it comes to impact (Speer, 2021; Tambe et al., 2019; Vassilopoulou, et al., 2022). Early on, Strohmeier & Piazza (2013) demonstrate that organisations are purposedly avoiding discrimination when preparing and treating data, and selecting techniques, highlighting the idea of certain legal and ethical awareness on the part of HRA professionals. This is corroborated by review and empirical articles that go beyond data privacy pointing out issues such as cyber-risk potential in HRA, particularly in the area of data security (García-Arroyo & Osca, 2019; Tursunbayeva et al., 2018) or the creation of data ethics protocols within HRA functions (Belizón & Kieran, 2022; Tursunbayeva et al., 2021).

A series of articles have also expressed serious ethical concerns existing in highly automated and AI-driven HRA environments: (a) the transparency of algorithms (Angrave et al., 2016); (b) potential unfairness biases (Dahlbom et al., 2020; Speer, 2021; Vassilopoulou, et al., 2022); and (c) employee privacy and reactions to decisions made assisted by algorithms (Chatterjee, et al., 2022; Tambe et al., 2019). Additionally, the progressive datafication of the workplace may promote the naive thought that the data produced constitutes a good description of the work context and the phenomenon to be studied leading to a potentially misaligned institutionalisation of metrics (Gal et al., 2020). This can be mitigated, among other things, by always including qualitative contextual information and listening to the business acumen present in the organisation (Newman et al., 2020). Table 7 captures how impact has been operationalised in HRA projects.

Table 7. Impact of HRA projects

Benefits of reusing the same methodology
Decreased labor cost
Increased of shift satisfaction
Formation of discussion groups to share good practices after HRA project
Decreased employee turnover
Rise in employee productivity

| Increased sales volume |
| Improved restaurant profitability |

Discussion and future research avenues

Our systematic literature review has embraced scholarship across three main disciplines where HR Analytics studies have been published, namely, human resource management, business analytics and management information systems. This has given us the opportunity to produce a comprehensive view of the state-of-the-art research evidence in the growing space of HR Analytics to address an important issue: HR Analytics in practice and how this process is carried out. In order to tackle this question, taking a cross-disciplinary approach was indispensable. We believe there are relevant novel insights we will now discuss, which will also lead us to potential future research avenues. Through a systematic review, building on the KDP framework and with a clear focus on praxis, we have uncovered how organisations address the different stages of the HRA process.

Delving into the HR issues and concerns to which the HRA projects respond, performance, workforce planning, talent management and employee engagement appear to be top of the list in most organisations (Davenport et al., 2010; Garg et al., 2021; Van der Togt & Rasmussen, 2017). Different reasons may explain this focus. First, one of the current major HR challenges seems to be employee attraction and retention (CIPD, 2021). This entails the identification of capabilities in nature, quality and quantity that will render an optimal workforce in terms of performance while aligned to business needs (Davenport et al., 2010; Minbaeva, 2018). Second, performance and workforce planning metrics are more accessible than others (Berk et al., 2019; Hargaden & Ryan, 2015; Luo et al., 2018; Wang & Cotton, 2018). Third, as the way we work evolves in a post Covid-19 world, HRA has served as an effective tool to navigate the transition from onsite to hybrid work (Belizón & Kieran, 2022).

What is striking perhaps is the number of potential HR issues that are not present in the existing literature. This, indeed, is not necessarily a reflection of current managerial practice. For instance, given the relevance of diversity and inclusion (D&I) in scholarship as well as practice (Cachat-Rossetet al. 2021; Kelan, 2022), we have not found evidence of such projects in our review of the literature. The rationale behind this can be manifold and we can only speculate. On the one hand, recent regulatory requirements enforced in some countries to guarantee a certain level of transparency in relation to gender representation on boards and senior-level teams has plausibly made D&I reporting a priority for HRA teams. On the other, if we turn to practitioners' fora or HRA textbooks, there are also examples of slightly more sophisticated D&I analysis, such as the scrutiny of intersectionality in teams whereby organisations could easily explore the levels of employee engagement, perceived fairness and co-workers support of female ethnic minorities across teams and how this could predict attrition or lead to greater possibilities for promotion (Edwards & Edwards, 2019). This disconnect between academia and practice may be due to a natural and understandable reluctance in organisations to share their D&I data with third parties, including scholars, in their attempt to minimise any potential reputational or legal risk attached to their D&I status quo (Edwards et al. 2022)[i].

Another broad space where HRA is only starting to support is HR process improvement or what practitioners term HR transformation (Belizón & Kieran, 2022). Some experts have advocated for the specific assessment and enhancement of HR practices and processes through the introduction and evaluation of process metrics (West, 2019). If we take recruitment and selection as an example of a HR process, how can we make it more efficient in its execution? The introduction of metrics across the recruitment process is key to this efficiency, namely, volume of applicants, speed, cost and quality are core aspects of the process that can be quantified. For this to happen in a more transparent way, organisations may need to develop process blueprints and metrics for all key HR practices and processes. And indeed, involvement of all stakeholders in the assessment of process efficiency remains crucial. An example of this would be the impact of HR Business Partners and line managers voice in the development of HR data and metrics

dashboards and how organisations have used them to deliver tailored solutions for their business needs (Angrave et al., 2016; Belizón & Kieran, 2022; Ellmer & Reichel, 2021).

An examination of the type of data used in HRA unearthed six essential types of data: (i) those corresponding to employees KSAO's, (ii) employee work-related records, (iii) employee outcomes, (iv) human resource outcomes, (v) organisational outcomes, and (vi) external data. Two important points are worth noting. On the one hand, the richness of the data used: not only data obtained exclusively at the individual level are used, but at group and organisational level (e.g. Simón & Ferreiro, 2018). On the other hand, the ample richness of data collection. Far from only using data from HRIS the studies analysed showed a range of sources, from internal sources related to operations (e.g., Berk et al., 2019), to open texts posted by employees and companies onto social job sites to the generation of data for analyses from statistical modelling procedures (Hargaden & Ryan, 2015) and the analysis of employee interaction networks (Ballinger et al., 2016). When it comes to untapped data, we observe a lacuna in relation to HR continuous improvement data, project management platforms data, voice content and tone data, meta-data, and other passive data. Little or nothing is found using behavioural data (movement, posture, facial expression), or biometrics. We assume that naturally it is likely that ethical and privacy considerations are behind this outcome. Additionally, the wealth of the type of data used seems to show that methods and tools traditionally associated with Big Data have a place in the development of the HRA. Here we highlight the potential use of historical and longitudinal data in organisations.

When exploring how technological tools are deployed for data management, we found that organisations are interrogating in-house or proprietary HRIS with a large number of data points, which require IT skills beyond the traditional HR capability and skillset (Berk et al., 2019; Bohlouli et al., 2017; Simón & Ferreiro, 2018). Some organisations dared to use novel techniques such as natural language processing (NLP) and organisational network analysis (ONA), sourcing their data from either websites and email repositories, or from open questions incorporated in employee engagement surveys (Fareri et al., 2020; Gelbard et al., 2018; Safarishahrbijari, 2018).

This co-exists with the deployment of online or traditional hard copy questionnaires (Shah et al., 2017).

Although there have been sceptical views around the existence of big data in HR (Capelli, 2017), our findings, while acknowledging this overwhelming evidence, also point to a slight shift towards big data when it comes to longitudinal HR data and text and sentiment analysis (Avrahami et al., 2022; Luo et al., 2019; Pessach et al., 2020; Wang & Cotton, 2018). There is some consensus in the HRM literature around the limited nature of big data in HR, however, more research is needed in the area of unstructured and passive data. On the one hand, unstructured data provides HR with a vehicle for greater sample sizes, bordering the big data phenomenon (Angrave et al., 2016; Capelli, 2017). One note of caution here would be that HRA projects should be designed within social and ethical boundaries whereby only truly necessary and reasonably justified data is collected (Tursunbayeva et al., 2018; Tursunbayeva et al., 2022). On the other hand, there is certainly an abundant scarcity of HRA studies analysing and deploying passive data. Far from suggesting the use of CCTV or video capture, which we did not find in our sample, the analysis of passive data gathered by digital platforms such as project management tools (e.g. JIRA, Wrike, etc.) or Internet of Things (IoT) devices can become a viable source of HR data when manipulated in an aggregated manner. Finally, a greater integration of HR data within business data lakes or data warehouses continues to be a challenge but a necessary undertaking for a successful adoption of impactful HRA (Belizón & Kieran, 2022; Rasmussen & Ulrich, 2015).

Reviewing how statistical techniques are employed to analyse HR data, frequently these are techniques used in high quality academic research in the fields of I/O Psychology and HRM, such as SEM (Shah et al., 2017; Zhang et al., 2021), regression models (Bellinger et al., 2016; Schofield et al., 2018; Simón & Ferreiro, 2018), Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA) (Shah et al., 2017) and also complex models of mathematical optimisation (Berk et al., 2019; Bohlouli et al., 2017; Hargarden & Ryan, 2015; Luo et al., 2018). However, it is worth noting that we have also found the use of more dvanced

techniques such as Natural Language Processing (NLP), Organisational Network Analysis (ONA) or even machine learning techniques such as Artificial Neural Networks, when the case study required it (Arfae et al., 2022; Liu et al., 2020). We could think that HRA is not so mature as other functions in the organisation like marketing or business intelligence because it falls short of advanced analytics (Dahlbom et al., 2020). Linear models are easier models to interpretate and explain than the "black box" models of sophisticated algorithms, which made many executives uneasy, especially when the resulting recommendations conflicts with conventional wisdom (Cheng & Hackett, 2021). We find that linear models still play a key role understanding the "root causes" we need to transmit to our stakeholders (Strohmeier & Piazza, 2013; Valencia, 2018).

The analysis carried out on how the data obtained in the HRA process are evaluated and interpreted shows a lack of thoroughness in assessing statistical outputs and connecting them with HR or organisational outcomes (Ellmer and Reichel, 2021; Fu et al., 2022 Safarishahrbijari, 2018). Many of the articles analysed directly do not evaluate the statistical analysis outcome, and additionally, when this evaluation is carried out it only focuses on statistical indicators without engaging in generating managerial implications. However, statistical significance does not imply a direct potential change in organisational effectiveness or the organisation's strategy (Levenson, 2018). What seems to be crucial here is data interpretation and storytelling on the part of HR Analysts (Algrave et al., 2016; Ellmer and Reichel, 2021; Fu et al., 2022; Greasley & Thomas, 2020). The lack of understanding, common logic and language between HR Analysts and HR professionals makes the communication of results challenging and oftentimes, sterile. HR Analysts are, however, engaging in serious institutional work by learning to showcase HRA projects' insights using storytelling skills and curbing when necessary to navigate their relationships with other HRA team members or HR professionals (Fu et al., 2022).

In relation to impact, we have sought to uncover how HRA outcomes are used in decision-making for HR processes, practices and programs, where we found a substantial void. This could be due to a variety of reasons. Transposing HR Analytics results into an approachable action plan can be challenging both in terms of legitimacy and cost binding implications (Belizón & Kieran,

2022; Ellmer & Reichel, 2021). Often there is a considerable disconnect between the implications of a particular project and the necessary steps that need to be taken to put them into practice. Both can originate from a possible lack of credibility of the HR Analytics potential impact within the HR function. In relation to HR legitimacy, this has been a historical legacy that HR functions and HR professionals have been dragging their heels on for years (Pholer & Willness, 2014; Heizmann & Fox, 2019). One of the reasons behind this lack of legitimacy is precisely a substantial challenge linking HR practices and business outcomes. Oftentimes HR metrics are poorly defined and constructed and therefore, their validity and reliability may hinder their objective impact on business outcomes. Additionally, there seems to be a disconnect between HR metrics and business metrics. Just a handful of the studies included here gave evidence of a direct or indirect connection between HR and business metrics (Aral et al., 2012; Schiemann et al., 2018; Simón & Ferreiro, 2018; Xu & Li, 2021), and therefore, business impact was rare.

We suggest that HRA scholars and practitioners may focus on mapping HR impact through those HR metrics directly associated to strategic business activities or priorities (Margherita, 2021). As linking any management practice with organisational performance can become excessively farfetched due to the aggregated nature of organisational performance, it seems to become more manageable from a metric viewpoint to establish it through 'operational performance measures' (Crook et al., 2011; p. 445). Hence, future research avenues can concentrate on the links between HRA impact on HR or operational performance, and eventually on HR and business strategy (Angrave et al., 2016; Minbaeva, 2018). The increasing need for academic-industry collaborations aiming at projects midway between research and practice can serve as a vehicle for high quality and impactful work in management studies, including HRA (Avenier & Cajaiba, 2012; Lawler & Benson, 2020; Kakabadse and Morley, 2021). Simón & Ferreiro (2018) becomes the poster child of these collaborations, bringing together practitioners' needs and the academic hat to design, validate and action robust and scientifically proven HRA projects with a view to impacting business outcomes.

Finally, there is a growing awareness of the ethical impact data-driven HR is having on organisations (García-Arroyo & Osca, 2019; Tursunbayeva et al., 2018; Tursunbayeva et al.,

2022) ranging from people risks and security considerations, data privacy and ethical protocols, algorithmic transparency and potential unfairness biases (Angrave et al. 2016; Belizón & Kieran, 2022; Dahlbom et al., 2020; Speer 2021; Tambe et al., 2019). Again, a greater attention should be given to praxis when it comes to ethical implications. Most of the existing literature is theoretical in nature, perhaps neglecting what is currently occurring in organisations.

In terms of limitations, our research has attempted to bring together studies from three different disciplines, namely, HRM, MIS and business analytics. This has not come without challenges. Journal articles from these fields often take different structure and focus and one could argue that comparability is indeed inadequate to a certain extent. We found this to be the case when analysing those articles from MIS and business analytics literature developing mathematical models and technological applications. This limitation can be mitigated in the future by cross disciplinary collaboration. Additionally, even though systematic literature reviews are useful to synthesise management research in an effective manner (Williams et al., 2021), there are certain limitations associated to this sort of literature review. One of them, which applies to ours, is the limited scope of available relevant studies (Sampaio & Mancini, 2007; Siddaway et al., 2019; Williams et al., 2021), namely, literature that covers a comprehensive view of the complete HRA process. Finally, the lack of consensus on HR Analytics as a keyword (Edwards et al., 2022) may have also limited our search, and therefore our keyword inclusion was carefully crafted.

Nonetheless, this article constitutes the first literature review that contributes to HRA scholarship by providing insights on HRA in practice. Our unfolding of the KDP model as it applies to HRA projects can help scholars to understand the HRA process more comprehensively and can also offer practitioners a breadth of evidence of what has worked in the past and the possibilities for the future.

We conclude by setting down some implications for practice. First, this review article constitutes the first step-by-step study of the HRA process, unfolding the intricacies related to each of these process stages. This roadmap for HRA can be useful for HR and HRA professionals not only as a solid methodology but also to understand the complexities, dynamics and possible choices of the process in all its stages. Second, it can help HR professionals to understand HRA

in a more comprehensive manner. Unlike any other HR activity, HRA encapsulates six different stages, each with a distinct focus and set of abilities, needed of an overarching vision and leadership. Third, the article also showcases systematically those techniques that have been used in the past and that the existing literature has been able to capture. Finally, this article can serve as a process guide to identify the diverse range of technical skillsets needed, not just in the data analysis phase but even earlier at the data cleaning and management stage. Thus, it can provide a realistic overview of the skills and capabilities needed in each stage of the process.

References

Aguado. (2018) HR analytics : teoría y práctica para una analítica de recursos humanos con impacto [HR analytics: theory and practice for impactful human resources analytics]. ESIC: Madrid.

Aldrich, H.E. (2001). Who wants to be an evolutionary theorist? *Journal of Management Inquiry*, 10(June): 115–127.

*Algethami, H., Martínez-Gavara, A. & Landa-Silva, D. (2019) Adaptive multiple crossover genetic algorithm to solve workforce scheduling and routing problem. *Journal of Heuristics, 25*(4-5), 753–792. Available from: https://doi.org/10.1007/s10732-018-9385-x

*Alshareef, A. G., Wraith, D., Dingle, K. & Mays, J. (2020) Identifying the factors influencing Saudi Arabian nurses' turnover. *Journal of Nursing Management*, *28*(5), 1030–1040. Available from: https://doi.org/10.1111/jonm.13028

Anand, S. S. & Büchner, A. G. (1998) *Decision support using data mining*. Financial Times Management.

*Andersen, M. K. (2017). Human capital analytics: the winding road. *Journal of Organizational Effectiveness: People and Performance*, 4(2), 133-136. Available from: https://doi.org/10.1108/JOEPP-03-2017-0024

*Andre, M., Baldoquín, M. G. & Acuña, S. T. (2011) Formal model for assigning human resources to teams in software projects. *Information and Software Technology*, *53*(3), 259–275. Available from: https://doi.org/10.1016/j.infsof.2010.11.011

*Angrave, D., Charlwood, A., Kirkpatrick, I., Lawrence, M. & Stuart, M. (2016) HR and analytics: why HR is set to fail the big data challenge. *Human Resource Management Journal, 26*(1), 1-11. Available from: https://doi-org.ucd.idm.oclc.org/10.1111/1748-8583.12090

*Aral, S., Brynjolfsson, E. & Wu, L. (2012) Three-Way Complementarities: Performance Pay, Human Resource Analytics, and Information Technology. *Management Science*, *58*(5), 913–931. Available from: https://doi.org/10.1287/mnsc.1110.1460

*Arfaee, M., Bahari, A. & Khalilzadeh, M. (2022) A novel prediction model for educational planning of human resources with data mining approach: a national tax administration case study. *Education and Information Technologies*, *27*(2), 2209–2239. Available from: https://doi.org/10.1007/s10639-021-10699-6

Avenier, M.J. and Cajaiba, A.P. (2012), The Dialogical Model: Developing Academic Knowledge for and from *Practice. European Management Review*, 9(4), 199-212. Available from: https://doi.org/10.1111/j.1740-4762.2012.01038.x

*Avrahami, D., Pessach, D., Singer, G. & Ben-Gal, H. C. (2022) A human resources analytics and machine-learning examination of turnover: implications for theory and practice. *International Journal of Manpower*, *43*(6), 1405-1424. Available from: http://dx.doi.org/10.1108/IJM-12-2020-0548

*Ballinger, G. A., Cross, R. & Holtom, B. C. (2016) The right friends in the right places: Understanding network structure as a predictor of voluntary turnover. *The Journal of Applied Psychology, 101*(4), 535–548. Available from: https://doi.org/10.1037/apl0000061

*Barnes, S. J., Mattsson, J., Sørensen, F. & Jensen, J. F. (2020) Measuring employee-tourist encounter experience value: A big data analytics approach. *Expert Systems with Applications, 154*, 113450. Available from: https://doi.org/10.1016/j.eswa.2020.113450

*Bastian, N. D., Lunday, B. J., Fisher, C. B. & Hall, A. O. (2020) Models and methods for workforce planning under uncertainty: Optimizing U.S. Army cyber branch readiness and manning. *Omega, 92*, 102171. Available from: https://doi.org/10.1016/j.omega.2019.102171

*Bechter, B., Brandl, B. & Lehr, A. (2022) The role of the capability, opportunity, and motivation of firms for using human resource analytics to monitor employee performance: A multi-level analysis of the organisational, market, and country context. *New Technology, Work, and Employment*, 1-27. Available from: https://doi.org/10.1111/ntwe.12239

*Becker, T. (2020) A decomposition heuristic for rotational workforce scheduling. *Journal of Scheduling*, *23*(5), 539–554. Available from: https://doi.org/10.1007/s10951-020-00659-2

*Belizón, M.J. & Kieran, S. (2022) Human resources analytics: A legitimacy process. *Human Resource Management Journal*, *32*(3), 603–630. Available from: https://doi.org/10.1111/1748-8583.12417

*Ben-Gal, H.C. (2019) An ROI-based review of HR analytics: practical implementation tools. *Personnel Review*, *48*(6), 1429–1448. Available from: https://doi.org/10.1108/PR-11-2017-0362.

*Berk, L., Bertsimas, D., Weinstein, A. M. & Yan, J. (2019) Prescriptive analytics for human resource planning in the professional services industry. *European Journal of Operational Research, 272*(2), 636–641. Available from: https://doi.org/10.1016/j.ejor.2018.06.035

*Bohlouli, M., Mittas, N., Kakarontzas, G., Theodosiou, T., Angelis, L. & Fathi, M. (2017) Competence assessment as an expert system for human resource management: A mathematical approach. *Expert Systems with Applications, 70*, 83–102. Available from: https://doi.org/10.1016/j.eswa.2016.10.046

*Boudreau, J. & Cascio, W. (2017) Human capital analytics: why are we not there? *Journal of Organizational Effectiveness: People and Performance, 4*(2), 119-126. Available from: https://doi-org.ucd.idm.oclc.org/10.1108/JOEPP-03-2017-00211

Cabena, P., Hadjinian, P., Stadler, R., Verhees, J. & Zanasi, A. (1998) *Discovering Data Mining: From Concepts to Implementation*. Prentice Hall Saddle River: New Jersey.

Cappelli, P. (2017). There's no such thing as big data in HR. *Harvard Business Review*, *2*, 2-4. Available from: https://hbr.org/2017/06/theres-no-such-thing-as-big-data-in-hr

Cascio, W. & Boudreau, J. (2010) *Investing in people: Financial impact of human resource initiatives.* Pearson Education LTD: New Jersey.

*Chatterjee, S., Chaudhuri, R. & Vrontis, D. (2021) Dark side of instant messaging: an empirical investigation from technology and society perspective. *Aslib Journal of Information Management*, *74*(5), 928-955. Available from: https://doi.org/10.1108/AJIM-04-2021-0101

*Chen, C. C., Wei, C. C., Chen, S. H., Sun, L. M. & Lin, H. H. (2022) AI Predicted Competency Model to Maximize Job Performance. *Cybernetics and Systems*, *53*(3), 298-317. Available from: http://dx.doi.org/10.1080/01969722.2021.1983701

Cheng, M.M. & Hackett, R. D. (2021) A critical review of algorithms in HRM: Definition, theory, and practice. *Human Resource Management Review*, *31*(1), 100698. Available from: https://doi.org/10.1016/j.hrmr.2019.100698

Cios, K.J., Pedrycz, W., Swiniarski, R. W. & Kurgan, L. A. (2007) *Data Mining: A Knowledge Discovery Approach*. Springer: New York.

CIPD (2022) People Profession 2022: UK and Ireland survey report. *Chartered Institute of Personnel and Development*. Available from: https://www.cipd.co.uk/knowledge/strategy/hr/people-profession-survey

*Conway, E., Rosati, P., Monks, K. & Lynn, T. (2019) Voicing job satisfaction and dissatisfaction through Twitter: employees' use of cyberspace. *New Technology, Work and Employment*, *34*(2), 139-156. Available from: https://doi.org/10.1111/ntwe.12135

*Corominas, A., Lusa, A. & Olivella, J. (2012) A detailed workforce planning model including non-linear dependence of capacity on the size of the staff and cash management. *European Journal of Operational Research*, *216*(2), 445-458. Available from: https://doi.org/10.1016/j.ejor.2011.06.027

Crook, T.R., Todd,S.Y., Combs,J.G., Woehr, D.J. & Ketchen, D.J. (2011) Does human capital matter? A meta-analysis of the relationship between human capital and firm performance. *Journal of Applied Psychology*, *96*(3), 443-456. Available from: https://doi.org/10.1037/a0022147

*Dahlbom, Siikanen, N., Sajasalo, P. & Jarvenpää, M. (2020) Big data and HR analytics in the digital era. *Baltic Journal of Management*, *15*(1), 120–138. Available from: https://doi.org/10.1108/BJM-11-2018-0393

Davenport, T. H., Harris, J. G. & Morison, R. (2010*) Analytics at work: Smarter decisions, better results*. Harvard Business Press: Boston.

*De Laat, M. & Schreurs, B. (2013) Visualizing informal professional development networks: Building a case for learning analytics in the workplace. *American Behavioral Scientist*, *57*(10), 1421-1438. Available from: http://dx.doi.org/10.1177/0002764213479364

*Du, L. & Li, Q. (2020) A Data-driven approach to high-volume recruitment: Application to student admission. *Manufacturing & Service Operations Management*, *22*(5), 942-957. Available from: http://dx.doi.org/10.1287/msom.2019.0779

Edwards, M.R. & Edwards, K. (2019) *Predictive HR analytics: Mastering the HR metric*. Kogan Page Publishers: London.

Edwards, M.R., Charlwood, A., Guenole, N. & Marler, J. (2022) HR analytics: An emerging field finding its place in the world alongside simmering ethical challenges. *Human Resource Management Journal*, 1-11. Available from: https://doi.org/10.1111/1748-8583.12435

*Ellmer, M. & Reichel, A. (2021) Staying close to business: The role of epistemic alignment in rendering HR Analytics outputs relevant to decision-makers. *International Journal of Human Resource Management*, *32*(2), 2622–2642. Available from: https://doi.org/10.1080/09585192.2021.1886148

*Esmaeilzadeh, M., Abdollahi, B., Ganjali, A. & Hasanpoor, A. (2016) Evaluation of employee profiles using a hybrid clustering and optimization model Practical study. *International Journal of Intelligent Computing and Cybernet, 9(3), 281-236.* Available from: https://doi.org/10.1108/IJICC-01-2016-0004

*Faliagka, E., Tsakalidis, A. & Tzimas, G. (2012) An integrated e-recruitment system for automated personality

   mining and applicant ranking. *Internet Research, 22(5) 551-568.* Available from:

   https://doi.org/10.1108/10662241211271545

Falletta, S. V. & Combs, W. L. (2020) The HR analytics cycle: a seven-step process for building evidence-based and

   ethical HR analytics capabilities. *Journal of Work-Applied Management*, *13*(1) 51-68. Available from:

   https://doi.org/10.1108/JWAM-03-2020-0020

* Fallucchi, F., Coladangelo, M., Giuliano, R. & William De Luca, E. (2020) Predicting Employee Attrition Using

   Machine Learning Techniques. *Computers (Basel), 9*(4), 86. Available from:

   https://doi.org/10.3390/computers9040086

*Fan, C.Y., Fan, P.S., Chan, T.Y. & Chang, S.H. (2012) Using hybrid data mining and machine learning clustering

   analysis to predict the turnover rate for technology professionals. *Expert Systems with Applications,*

   39(10), 8844–8851. Available from: https://doi.org/10.1016/j.eswa.2012.02.005

*Fareri, S., Fantoni, G., Chiarello, F., Coli, E. & Binda, A. (2020) Estimating Industry 4.0 impact on job profiles and

   skills using text mining. *Computers in Industry*, *118*, 103222. Available from:

   https://doi.org/10.1016/j.compind.2020.103222

Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996) The KDD process for extracting useful knowledge from

   volumes of data: Data mining and knowledge discovery in databases. *Communications of the ACM,*

   39(11), 27–34. Available from: https://doi.org/10.1145/240455.240464

*Fernandez, V. & Gallardo-Gallardo, E. (2021) Tackling the HR digitalization challenge: key factors and barriers to

   HR analytics adoption. *Competitiveness Review*, *31*(1), 162–187. Available from:

   https://doi.org/10.1108/CR-12-2019-0163

*Fu, N., Keegan, A. & McCartney, S. (2022) The duality of HR analysts' storytelling: Showcasing and

   curbing. *Human Resource Management Journal,* 1-26. Available from: http://dx.doi.org/10.1111/1748-

   8583.12466

*Gal, U., Jensen, T. B. & Stein, M. K. (2020) Breaking the vicious cycle of algorithmic management: A virtue ethics

   approach to people analytics. *Information and Organization*, *30*(2), 100301. Available from:

   http://dx.doi.org/10.1016/j.infoandorg.2020.100301

*Garcia-Arroyo, J. & Osca, A. (2021) Big data contributions to human resource management: a systematic review. *The International Journal of Human Resource Management*, *32*(20), 4337-4362. Available from: https://doi.org/10.1080/09585192.2019.1674357

*Garg, S., Sinha, S., Kar, A. K. & Mani, M. (2021) A review of machine learning applications in human resource management. *International Journal of Productivity and Performance Management*, *71*(5), 1590. Available from: http://dx.doi.org/10.1108/IJPPM-08-2020-0427

*Gelbard, R., Ramon-Gonen, R., Carmeli, A., Bittmann, R. M. & Talyansky, R. (2018) Sentiment analysis in organizational work: Towards an ontology of people analytics. *Expert Systems, 35*(5), e12289. Available from: https://doi.org/10.1111/exsy.12289

*Giermindl, L. M., Strich, F., Christ, O., Leicht-Deobald, U. & Redzepi, A. (2022) The dark sides of people analytics: Reviewing the perils for organisations and employees. *European Journal of Information Systems*, *31*(3), 410-435. Available from: https://doi.org/10.1080/0960085X.2021.1927213

*Ginda, M., Richey, M. C., Cousino, M. & Börner, K. (2019a) Visualizing learner engagement, performance, and trajectories to evaluate and optimize online course design. *PloS One, 14*(5), e0215964. Available from: https://doi.org/10.1371/journal.pone.0215964

Godfroid, C. & Labie, M. (2022) The process of organizational identification in social enterprises: The role of coalitions. *European Management Review*, 1– 11. https://doi.org/10.1111/emre.12551

*Greasley, K. & Thomas, P. (2020) HR analytics: The onto-epistemology and politics of metricised HRM. *Human Resource Management Journal*, *30*(4), 494-507. Available from: https://doi.org/10.1111/1748-8583.12283

*Guo, F., Gallagher, C. M., Sun, T., Tavoosi, S. & Min, H. (2021) Smarter people analytics with organizational text data: Demonstrations using classic and advanced NLP models. *Human Resource Management Journal*, 1-16. Available from: https://doi.org/10.1111/1748-8583.12426

*Gurusinghe, R. N., Arachchige, B. J. & Dayarathna, D. (2021) Predictive HR analytics and talent management: a conceptual framework. *Journal of Management Analytics*, *8*(2), 195-221. Available from: https://doi.org/10.1080/23270012.2021.1899857

*Hargaden, V. & Ryan, J. (2015) Resource Planning in Engineering Services Firms. *IEEE Transactions on Engineering Management 62(4), 578-590.* Available from: https://doi.org/10.1109/TEM.2015.2457675

Heizmann, H. & Fox, S. (2019) O Partner, Where Art Thou? A critical discursive analysis of HR managers' struggle for legitimacy, *The International Journal of Human Resource Management, 30*(13), 2026-2048. Available from: https://doi-org.ucd.idm.oclc.org/10.1080/09585192.2017.1314974

*Hermawati, A. & Mas, N. (2017) Mediation effect of quality of work life, job involvement, and organizational citizenship behavior in relationship between transglobal leadership to employee performance. *International Journal of Law and Management*, *59*(6), 1143–1158. Available from: https://doi.org/10.1108/IJLMA-08-2016-0070

*Hickman, L., Saef, R., Ng, V., Woo, S. E., Tay, L. & Bosch, N. (2021) Developing and evaluating language-based machine learning algorithms for inferring applicant personality in video interviews. *Human Resource Management Journal*, 1-19. Available from: https://doi.org/10.1111/1748-8583.12356

*Holwerda, J. A. (2021) Big data? Big deal: Searching for big data's performance effects in HR. *Business Horizons*, *64*(4), 391-399. Available from: https://doi.org/10.1016/j.bushor.2021.02.006

*Huselid, M. A. (2018) The science and practice of workforce analytics: Introduction to the HRM special issue. *Human Resource Management, 57*(3), 679-684. Available from: https://doi-org.ucd.idm.oclc.org/10.1002/hrm.21916

*Jafari, A., Rouhanizadeh, B., Kermanshachi, S. & Murrieum, M. (2020) Predictive Analytics Approach to Evaluate Wage Inequality in Engineering Organizations. *Journal of Management in Engineering,* 36(6). Available from: https://doi.org/10.1061/(ASCE)ME.1943-5479.0000841

*Jain, N., Tomar, A. & Jana, P. K. (2020) A novel scheme for employee churn problem using multi-attribute decision making approach and machine learning. *Journal of Intelligent Information Systems, 56, 279–302.* Available from: https://doi.org/10.1007/s10844-020-00614-9

*Jörden, N. M., Sage, D. & Trusson, C. (2022) 'It's so fake': Identity performances and cynicism within a people analytics team. *Human Resource Management Journal, 32(3), 524-539.* Available from: https://doi.org/10.1111/1748-8583.12412

Kakabadse, N.K. & Morley, M.J. (2021) In search of stewardship: Advancing governance research*. European Management Review,* 18(3), 189– 196. Available from: https://doi.org/10.1111/emre.12493

*Kang, R., Kunkel, S. T., Columbo, J. A., Goodney, P. P. & Wong, S. L. (2019) Association of Hospital Employee Satisfaction with Patient Safety and Satisfaction within Veterans Affairs Medical Centers. *The American Journal of Medicine*, *132*(4), 530-534. Available from: https://doi.org/10.1016/j.amjmed.2018.11.031

*Kifor, C. V., Nicolaescu, S. S., Florea, A., Săvescu, R. F., Receu, I., Ţîrlea, A. V. & Dănuţ, R. E. (2021) Workforce analytics in teleworking. *IEEE Access*, *9*, 156451-156464. Available from: https://doi.org/10.1109/ACCESS.2021.3129248

Langley, A. (1999) Strategies for theorizing from process data. *Academy of Management Review*, *24*(4), 691–710. Available from: https://doi.org/10.2307/259349

Langley, A. & Truax, J. (1994) A process study of new technology adoption in smaller manufacturing firms. *Journal of Management Studies*, *31*(5), 619–652. Available from: https://doi.org/10.1111/j.1467-6486.1994.tb00632.x

*Larsson, A. S. & Edwards, M. R. (2022) Insider econometrics meets people analytics and strategic human resource management. *The International Journal of Human Resource Management*, *33*(12), 2373-2419. Available from: http://dx.doi.org/10.1080/09585192.2020.1847166

Lawler III, E. E., Levenson, A. & Boudreau, J. W. (2004) HR metrics and analytics–uses and impacts. *Human Resource Planning Journal, 27*(4), 27-35.

Lawler III, E. E. & Benson, G. S. (2022) The practitioner-academic gap: A view from the middle. *Human Resource Management Review*, *32*(1), 100748. Available from: https://doi.org/10.1016/j.hrmr.2020.100748

Lee, Y. & Cho, J. (2020) Knowledge representation for computational thinking using knowledge discovery computing. *Information Technology Management, 21*, 15–28 (2020). Available from: https://doi-org.ucd.idm.oclc.org/10.1007/s10799-019-00299-9

*Leonardi, P. & Contractor, N. (2018) Better people analytics. *Harvard Business Review*, *96*(6), 70-81. Available from: https://hbr.org/2018/11/better-people-analytics

*Levenson, A. & Fink, A. (2017) Human capital analytics: too much data and analysis, not enough models and business insights. *Journal of Organizational Effectiveness: People and Performance*, *4*(2), 145–156. Available from: https://doi.org/10.1108/JOEPP-03-2017-0029

Levenson, A. (2018) Using workforce analytics to improve strategy execution. *Human Resource Management*, *57*(3), 685-700. Available from: https://doi-org.ucd.idm.oclc.org/10.1002/hrm.21850

*Liu, X. & Raghuram, S. (2022) The effects of latent withdrawal profiles on employee turnover, destinations and job performance. *Human Resource Management Journal*, 32(2), 384-405. Available from: https://doi.org/10.1111/1748-8583.12362

*Liu, Y. and Pant, G. & Sheng, O. R. (2020) Predicting Labor Market Competition: Leveraging Interfirm Network

and Employee Skills. *Information Systems Research, 31*(4) 1443-1466. Available from:

https://doi.org/10.1287/isre.2020.0954

*Llort, N., Lusa, A., Martínez-Costa, C. & Mateo, M. (2019) A decision support system and a mathematical model

for strategic workforce planning in consultancies. *Flexible Services and Manufacturing Journal, 31*(2),

497–523. Available from: https://doi.org/10.1007/s10696-018-9321-2

*Luo, Z., Liu, L., Yin, J., Li, Y. & Wu, Z. (2019) Latent Ability Model: A Generative Probabilistic Learning Framework

for Workforce Analytics. *IEEE Transactions on Knowledge and Data Engineering, 31*(5), 923–937.

Available from: https://doi.org/10.1109/TKDE.2018.2848658

*Maghsoodi, A., Riahi, D., Herrera-Viedma, E. & Zavadskas, E. K. (2020) An integrated parallel big data decision

support tool using the W-CLUS-MCDA: A multi-scenario personnel assessment. *Knowledge-Based

Systems, 195*, 105749. Available from: https://doi.org/10.1016/j.knosys.2020.105749

*Margherita, A. (2021) Human resources analytics: A systematization of research topics and directions for future

research. *Human Resource Management Review,* 32(2), 100795. Available from:

https://doi.org/10.1016/j.hrmr.2020.100795

*Marler, J. H. & Boudreau, J. W. (2017) An evidence-based review of HR Analytics. *The International Journal of

Human Resource Management, 28*(1), 3-26. Available from:

https://doi.org/10.1080/09585192.2016.1244699

*McCartney, S. & Fu, N. (2022a) Bridging the gap: why, how and when HR analytics can impact organizational

performance. *Management Decision*, *60*(13), 25-47. Available from: http://dx.doi.org/10.1108/MD-12-

2020-1581

* McCartney, S. & Fu, N. (2022b) Promise versus reality: a systematic review of the ongoing debates in people

analytics. *Journal of Organizational Effectiveness: People and Performance*, *9*(2), 281-311. Available

from: https://doi.org/10.1108/JOEPP-01-2021-0013

*Michelotti, M., McColl, R., Puncheva-Michelotti, P., Clarke, R. & McNamara, T. (2021) The effects of medium

and sequence on personality trait assessments in face-to-face and videoconference selection interviews:

Implications for HR analytics. *Human Resource Management Journal*, 1-19. Available from:

https://doi.org/10.1111/1748-8583.12425

*Minbaeva, D. B. (2018) Building credible human capital analytics for organizational competitive advantage. *Human Resource Management*, *57*(3), 701-713. Available from: https://doi-org.ucd.idm.oclc.org/10.1002/hrm.21848

Mortenson, M.J., Doherty, N.F. & Robinson, S. (2015) Operational research from Taylorism to Terabytes: A research agenda for the analytics age. *European Journal of Operational Research*, *241* (3), 583-595. Available from: https://doi.org/10.1016/j.ejor.2014.08.029

*Necula, S.C. & Strîmbei, C. (2019) People Analytics of Semantic Web Human Resource Résumés for Sustainable Talent Acquisition. *Sustainability, 11*(13), 1–18. Available from: https://doi.org/10.3390/su11133520

*Newman, D. T., Fast, N. J. & Harmon, D. J. (2020) When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes*, *160*, 149-167. Available from: https://doi.org/10.1016/j.obhdp.2020.03.008

*Nicolaescu, S. S., Florea, A., Kifor, C. V., Fiore, U., Cocan, N., Receu, I. & Zanetti, P. (2020) Human capital evaluation in knowledge-based organizations based on big data analytics. *Future Generation Computer Systems*, *111*, 654–667. Available from: https://doi.org/10.1016/j.future.2019.09.048

*Ontrup, G., Schempp, P. S. & Kluge, A. (2021) Choosing the right (HR) metrics: digital data for capturing team proactivity and determinants of content validity. *Journal of Organizational Effectiveness: People and Performance*, *9*(2), 212-232. Available from: https://doi.org/10.1108/JOEPP-03-2021-0064

*Peeters, T., Paauwe, J. & Van De Voorde, K. (2020) People analytics effectiveness: developing a framework. *Journal of Organizational Effectiveness: People and Performance*, *7*(2), 203-219. Available from: https://doi.org/10.1108/JOEPP-04-2020-0071

*Pessach, D., Singer, G., Avrahami, D., Chalutz Ben-Gal, H., Shmueli, E. & Ben-Gal, I. (2020) Employees recruitment: A prescriptive analytics approach via machine learning and mathematical programming. *Decision Support Systems, 134*, 113290. Available from: https://doi.org/10.1016/j.dss.2020.113290

Pohler, D. & Willness, C. (2014) Balancing interests in the search for occupational legitimacy: The HR professionalization project in Canada. *Human Resource Management*, *53*(3), 467-488. Available from: https://doi.org/10.1002/hrm.21579

*Platanou, K., Mäkelä, K., Beletskiy, A. & Colicev, A. (2017) Using online data and network-based text analysis in HRM research. *Journal of Organizational Effectiveness: People and Performance*, *5*(1), 81–97. Available from: https://doi.org/10.1108/JOEPP-01-2017-0007

*Prathan, S. & Ow, S. H. (2020) Determining the best-fit programmers using Bayes' theorem and artificial neural

   network. *IET Software*, *14*(4), 433-442. Available from: https://doi.org/10.1049/iet-sen.2018.5440

*Raman, R., Bhattacharya, S. & Pramod, D. (2018) Predict employee attrition by using predictive

   analytics. *Benchmarking: An International Journal*, *26*(1), 2–18. Available from:

   https://doi.org/10.1108/BIJ-03-2018-0083

*Rasmussen, T. & Ulrich, D. (2015) Learning from practice: how HR analytics avoids being a management fad.

   Organizational Dynamics, *44*(3), 236-242. Available from:

   https://doi.org/10.1016/J.ORGDYN.2015.05.008

Rob, B.B. & Denyer, D. (2012) In book: *Oxford Handbook of Evidence-Based Management: Companies,*

   Classrooms and Research. Chapter: Systematic Review and Evidence Synthesis as a Practice and

   Scholarship. Oxford University Press.

*Rombaut, E. & Guerry, M.A. (2018) Predicting voluntary turnover through human resources database analysis.

   Management Research Review, 41(1), 96–112. Available from: https://doi.org/10.1108/MRR-04-2017-

   0098

*Rombaut, E. & Guerry, M.A. (2020) The effectiveness of employee retention through an uplift modeling

   approach*. International Journal of Manpower*, *41*(8), 1199–1220. Available from:

   https://doi.org/10.1108/IJM-04-2019-0184

*Safarishahrbijari, A. (2018) Workforce forecasting models: a systematic review. *Journal of Forecasting*, *37*(7),

   739-753. Available from: https://doi.org/10.1002/for.2541

Salk, J.E. (2005), Often called for but rarely chosen: alliance research that directly studies process. European

   Management Review, 2(2), 117-122. Available from: https://doi.org/10.1057/palgrave.emr.1500035

*Samson, K. & Bhanugopan, R. (2022) Strategic human capital analytics and organisation performance: The

   mediating effects of managerial decision-making*. Journal of Business Researc*h, *144*, 637-649. Available

   from: https://doi.org/10.1016/j.jbusres.2022.01.044

Saradhi, V.V. & Palshikar, G.K. (2011) Employee churn prediction. Expert systems with applications, 38*(3), 1999-

   2006. Available from: https://doi.org/10.1016/j.eswa.2010.07.134

*Schiemann, W. A., Seibert, J. H. & Blankenship, M. H. (2018) Putting human capital analytics to work: Predicting

   and driving business success. *Human Resource Management*, *57*(3), 795-807. Available from:

   https://doi.org/10.1002/hrm.21843

Schmidt, C. & Nan Sun, W. (2018) Synthesizing Agile and Knowledge Discovery: Case Study Results, *Journal of Computer Information Systems*, *58*(2), 142-150. Available from: http://dx.doi.org/10.1080/08874417.2016.1218308

*Schofield, J. A., Zens, C. L., Hill, R. R. & Robbins, M.J. (2018) Utilizing reliability modeling to analyze United States Air Force officer retention. *Computers & Industrial Engineering, 117*, 171–180. Available from: https://doi.org/10.1016/j.cie.2018.01.013

*Shah, N., Irani, Z. & Sharif, A. M. (2017) Big data in an HR context: Exploring organizational change readiness, employee attitudes and behaviors. *Journal of Business Research, 70*, 366–378. Available from: https://doi.org/10.1016/j.jbusres.2016.08.010

*Shahbazi, B., Akbarnezhad, A., Rey, D., Ahmadian Fard Fini, A. & Loosemore, M. (2019) Optimization of Job Allocation in Construction Organizations to Maximize Workers' Career Development Opportunities. *Journal of Construction Engineering and Management, 145*(6), 04019036. Available from: https://doi.org/10.1061/(ASCE)CO.1943-7862.0001652

Shearer, C. (2000) The CRISP-DM Model: The New Blueprint for Data Mining. *Journal of Data Warehousing*, *5* (4).

*Shet, S. V., Poddar, T., Samuel, F. W. & Dwivedi, Y. K. (2021) Examining the determinants of successful adoption of data analytics in human resource management–A framework for implications. *Journal of Business Research*, *131*, 311-326. Available from: https://doi.org/10.1016/j.jbusres.2021.03.054

Siddaway, A.P., Wood, A.M. & Hedges, L.V. (2019) How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology*, *70* (1), 747-770. Available from: https://doi.org/10.1146/annurev-psych-010418-102803

*Simeunovic, N., Kamenko, I., Bugarski, V., Jovanovic, M. & Lalic, B. (2017) Improving workforce scheduling using artificial neural networks model. *Advances in Production Engineering & Management*, *12*(4), 337–352. Available from: https://doi.org/10.14743/apem2017.4.26

*Simón, C. & Ferreiro, E. (2018) Workforce analytics: A case study of scholar–practitioner collaboration. *Human Resource Management, 57*(3), 781–793. Available from: https://doi.org/10.1002/hrm.21853

*Speer, A. B. (2021) Empirical attrition modelling and discrimination: Balancing validity and group differences. *Human Resource Management Journal*, 1-19. Available from: https://doi.org/10.1111/1748-8583.12355

*Strohmeier, S. & Piazza, F. (2013) Domain driven data mining in human resource management: A review of
current research. *Expert Systems with Applications*, *40*(7), 2410-2420. Available from:
https://doi.org/10.1016/j.eswa.2012.10.059

*Strohmeier, S., Collet, J. & Kabst, R. (2022) (How) do advanced data and analyses enable HR analytics success? A
neo-configurational analysis. *Baltic Journal of Management*, 17(3), 285-303. Available from:
http://dx.doi.org/10.1108/BJM-05-2021-0188

*Tambe, P. & Cappelli, P. V. Yakubovich (2019) Artificial Intelligence in Human Resources Management:
Challenges and a Path Forward. *California Management Review*, *61*(4), 15-42. Available from:
https://doi.org/10.1177/0008125619867910

*Tursunbayeva, A., Di Lauro, S. & Pagliari, C. (2018) People analytics—A scoping review of conceptual boundaries
and value propositions. *International Journal of Information Management*, *43*, 224-247. Available from:
https://doi.org/10.1016/j.ijinfomgt.2018.08.00

*Tursunbayeva, A., Pagliari, C., Di Lauro, S. & Antonelli, G. (2021) The ethics of people analytics: risks,
opportunities and recommendations. *Personnel Review*, *51*(3). Available from:
http://dx.doi.org/10.1108/PR-12-2019-0680

*Ulrich, D. & Dulebohn, J. H. (2015) Are we there yet? What's next for HR? *Human Resource Management
Review, 25*(2), 188-204. Available from: https://doi.org/10.1016/j.hrmr.2015.01.004

Valencia, E. (2018) *People Analytics. Analítica y datos para Recursos Humanos*. Editorial Aranzadi (Navarra,
España).

Van der Laken, P., Bakk, Z., Giagkoulas, V., Van Leeuwen, L. & Bongenaar, E. (2018) Expanding the
methodological toolbox of HRM researchers: The added value of latent bathtub models and optimal
matching analysis. *Human Resource Management, 57*, (3), 751-760. Available from: https://doi-
org.ucd.idm.oclc.org/10.1002/hrm.21847

*Van der Togt, J. & Rasmussen, T. H. (2017) Toward evidence-based HR. *Journal of Organizational Effectiveness:
People and Performance*, *4*(2), 127-132. Available from: https://doi.org/10.1108/JOEPP-02-2017-0013

*Vassilopoulou, J., Kyriakidou, O., Özbilgin, M. F. & Groutsis, D. (2022), Scientism as illusio in HR algorithms:
Towards a framework for algorithmic hygiene for bias proofing. *Human Resource Management Journal*,
1-15. Available from: https://doi.org/10.1111/1748-8583.12430

*Wang, L. & Cotton, R. (2018) Beyond Moneyball to social capital inside and out: The value of differentiated workforce experience ties to performance. *Human Resource Management, 57*(3), 761–780. Available from: https://doi.org/10.1002/hrm.21856

West, M. (2019) *People analytics for dummies*. John Wiley & Sons: New Jersey.

Williams, R.I., Clark, L.A., Clark, R. & Raffo, D.M. (2021) Re-examining systematic literature review in management research: Additional benefits and execution protocols. *European Management Journal, 39*(4), 521-533. Available from: https://doi.org/10.1016/j.emj.2020.09.007

*Xu, M. & Li, C. (2021) Data Mining Method of Enterprise Human Resource Management Based on Simulated Annealing Algorithm. *Security and Communication Networks*, 1–9. Available from: https://doi.org/10.1155/2021/6342970

*Yahia, N. B., Hlel, J. & Colomo-Palacios, R. (2021) From big data to deep data to support people analytics for employee attrition prediction. *IEEE Access*, *9*, 60447-60458. Available from: https://doi.org/10.1109/ACCESS.2021.3074559

*Yam, J. & Skorburg, J. A. (2021) From human resources to human rights: Impact assessments for hiring algorithms. *Ethics and Information Technology*, *23*(4), 611-623. Available from: https://doi.org/10.1007/s10676-021-09599-7

*Yuan, S., Kroon, B. & Kramer, A. (2021) Building prediction models with grouped data: A case study on the prediction of turnover intention. *Human Resource Management Journal*, 1-19. Available from: https://doi.org/10.1111/1748-8583.12396

*Zhang, Y., Xu, S., Zhang, L. & Yang, M. (2021) Big data and human resource management research: An integrative review and new directions for future research. *Journal of Business Research*, *133*, 34-50. Available from: http://dx.doi.org/10.1016/j.jbusres.2021.04.019