10

Sorting Teachers Out

Automated Performance Scoring and the Limit of Algorithmic Governance in the Education Sector

Ching-Fu Lin[*]

10.1 INTRODUCTION

Big data is increasingly mined to train ADM tools, with consequential reverberations. Governments are among the primary users of such tools to sort, rank, and rate their citizens, creating a data-driven infrastructure of preferences that condition people's behaviours and opinions. China's social credit system, Australia's robo-debt program,[1] and the United States' welfare distribution platform are prime examples of how governments resort to ADM to allocate resources and provide public services.[2] Some commentators point to the rule of law deficits in the automation of government functions;[3] others emphasize how such technologies systematically exacerbate inequalities;[4] and still others argue that a society constantly being scored, profiled, and predicted threatens due process and justice generally.[5] In contemporary workplaces, algorithmically powered tools have also been widely

[1] On Australia's robo-debt see Chapter 5 in this book.
[2] Han-Wei Liu et al, '"Rule of Trust": Powers and Perils of China's Social Credit Megaproject' (2018) 32(1) *Columbia Journal of Asian Law* 1–36.
[3] Monika Zalnieriute et al, 'The Rule of Law and Automation of Government Decision-Making' (2019) 82(3) *Modern Law Review* 425–55.
[4] Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (New York: St. Martin's Press, 2018).
[5] Danielle K Citron and Frank Pasquale, 'The Scored Society: Due Process for Automated Predictions' (2014) 89(1) *Washington Law Review* 1–33.

adopted in business practices for efficiency, productivity, and management pur-poses.[6] Camera surveillance, data analysis, and ranking and scoring systems are algorithmic tools that have given employers enormous power over the employed, yet their use also triggers serious controversies over privacy, ethical concerns, labour rights, and due process protection.[7]

Houston Federation of Teachers v Houston Independent School District presents yet another controversial example of government 'algorithmization' and the power and perils of automated ranking and rating, targeting at a specific profession – teachers. The case concerns the implementation of value-added models (VAMs) that algorithmically link a teacher's contributions to students' growth on standard-ized tests and hold teachers accountable through incentives such as termination, tenure, or contract nonrenewal. The Houston Independent School District refused to renew more than 200 teachers' contracts in 2011 based on low value-added scores. The VAM is proprietary and is not disclosed to those affected, precluding them from gaining an understanding of the internal logic and decision-making processes at work, thereby causing serious harm to due process rights. Similar practices prevail across the United States following the enactment of the 2002 No Child Left Behind Act and the 2011 Race to the Top Act, in conjunction with other federal policy actions. Interestingly, until the 2017 summary judgment rendered by the Court in *Houston Federation of Teachers v Houston Independent School District*, which ruled in favour of the affected teachers, federal constitutional challenges against the use of VAMs for termination or nonrenewal of teachers' contracts were generally rejected. Yet, the case has received little attention, as it was subsequently settled.

The growing algorithmization of worker performance evaluation and workplace surveillance in the name of efficiency and productivity is not limited to specific industry sectors or incomes, and it has been implemented so rapidly that regulators struggle to catch up and employees suffer in an ever-widening power asymmetry. Algorithmically powered workplace surveillance and worker performance evalu-ation effectively expand employers' capacity of control by shaping expectations and conditioning the behaviours of employees, which may further distort the nature of the relationship between the employer and the employed. Furthermore, such algorithmic tools have been widely criticized to be neither reliable nor transparent and also prone to bias and discrimination.[8] Hence, the prevalent use of algorithmic

---

[6] See e.g., Anne Fisher, 'An Algorithm May Decide Your Next Pay Raise' (14 July 2019) *Fortune*.

[7] Saul Levmore and Frank Fagan, 'Competing Algorithms for Law: Sentencing, Admissions, and Employment' (2021) 88(2) *University of Chicago Law Review* 367–412.

[8] See e.g., James A Allen, 'The Color of Algorithms: An Analysis and Proposed Research Agenda for Deterring Algorithmic Redlining' (2019) 46(2) *Fordham Urban Law Journal* 219–70; Estefania McCarroll, 'Weapons of Mass Deportation: Big Data and Automated Decision-Making Systems in Immigration Law' (2020) 34 *Georgetown Immigration Law Journal* 705–31; and Sarah Valentine, 'Impoverished Algorithms: Misguided Governments, Flawed Technologies, and Social Control' (2019) 46(2) *Fordham Urban Law Journal* 364–427.

worker productivity and performance evaluation systems poses serious economic, social, legal, and political ramifications.

This chapter therefore asks critical questions that remain unanswered. What are the normative ramifications of this case? How can due process protection – procedural or substantive – be ensured under the maze of crude algorithmic worker productivity and performance evaluation systems such as the VAM, especially in light of the black box problems?[9] Can judicial review provide a viable form of algorithmic governance? How are such ADM tools reshaping professions like education? Does the increasingly blurred line between public and private authority in designing and applying these algorithmic tools pose new threats? Premised upon these scholarly and practical inquiries, this article seeks to examine closely the case of *Houston Federation of Teachers v Houston Independent School District*, analyze its ramifications, and provide critical reflections on ways to harness the power of automated governments.

10.2 THE CONTESTED ALGORITHMIZATION OF WORKER PERFORMANCE EVALUATION

Recently, organizations have increased their use of algorithmically powered tools used for worker productivity monitoring and performance evaluation. With the help of camera surveillance, data analysis, and ranking and scoring systems,[10] such tools have given employers significant power over their employees. Growing power asymmetry thereby disrupts the labour market and redefines the way people work. Amazon notoriously uses a combination of AI tools to recruit, monitor, track, score, and even automatically fire its employees and contractors, and these second-by-second measurements have raised serious concerns regarding systematic bias, discrimination, and human rights abuse.[11] Specifically, Amazon uses AI automated tracking systems to monitor and evaluate its delivery drivers, who are categorized as 'lazy' if their movements are too slow and receive warning notifications if they fail to meet the required workloads.[12] The system can even generate an automated order to

---

[9] Han-Wei Liu et al, '*Beyond State v. Loomis*: Artificial Intelligence, Government Algorithmization, and Accountability' (2019) 27(2) *International Journal of Law and Information Technology* 122–41; Jenna Burrel, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms' (2016) 3(1) *Big Data & Society* 1–12.

[10] David Leonhardt, 'You're Being Watched' (15 August 2022) *The New York Times*.

[11] Jeffrey Dastin, 'Amazon Scraps Secret AI Recruiting Tool that Showed bias against Women' (11 October 2018) *Reuters*; Victor Tangermann, 'Amazon Used an AI to Automatically Fire Low-Productivity Workers' (26 April 2019) *Futurism*; Annabelle Williams, '5 Ways Amazon Monitors Its Employees, from AI Cameras to Hiring a Spy Agency' (6 April 2021) *Business Insider*.

[12] Yuanyu Bao et al, 'Ethical Disputes of AI Surveillance: Case Study of Amazon', in *Proceedings of the 7th International Conference on Financial Innovation and Economic Development* (2022), 1339.

lay off an employee without the intervention of a human supervisor.[13] Despite the associated physical and psychological suffering, if an employee does not agree to be algorithmically monitored and controlled, the individual will lose his or her job.[14]

Cashiers, truck drivers, nursing home workers, and many other lower-paying jobs across various sectors have followed suit in adopting Amazon's algorithmization of workers' performance evaluation, aimed at maximizing productivity per capita per second and automating constant micromanagement. Employees who are under such performance evaluation programs can feel pressured to skip interval breaks and bathroom or coffee breaks to avoid adverse consequences.[15] According to a recent in-depth study published in *The New York Times*, eight of the ten largest corporations in the United States have deployed systems to track, often in real time, individual workers' productivity metrics under varied frameworks of data-driven control.[16] The global COVID-19 pandemic has further prompted corporations under profit pressures to keep tighter tabs on employees by means of online and real-time AI evaluation, thus accelerating a paradigm shift of workplace power that was already well underway.[17] Many of the practices adopted during COVID-19 will likely continue and become normalized in the post-pandemic era.

White-collar jobs are not immune from the growing algorithmization of worker performance evaluation. Architects, financial advisors, lawyers, pharmaceutical assistants, academic administrators, and even doctors and chaplains can be placed under extensive monitoring software that constantly accumulates records, and they are paid 'only for the minutes when the system detected active work', or are subject to a 'productivity points' management system that calibrates pay based on individual scores.[18] For example, some law firms are increasingly subjecting their contract lawyers to smart surveillance systems that constantly monitor their performance during work days in the name of efficiency facilitation and quality control.[19] It appears evident that the growing automation of worker performance evaluation is not limited to specific industry sectors or incomes, and such practices are spreading at such a rapid rate that regulators struggle to catch up and employees suffer from widening power asymmetry.

As Ifeoma Ajunwa, Kate Crawford, and Jason Schultz observe, due to recent technological innovations, data-driven worker performance evaluation in the United States is on the rise through tools including employee ratings, productivity apps,

---

[13] Ibid.
[14] Ibid, 1340. *See also* Katie Schoolov, 'Pee Bottles, Constant Monitoring and Blowing through Stop Signs: Amazon DSP Drivers Describe the Job' (21 June 2021) *CNBC*.
[15] Jodi Kantor and Arya Sundaram, 'The Rise of the Worker Productivity Score' (14 August 2022) *The New York Times*.
[16] Ibid.
[17] Ibid.
[18] Ibid.
[19] Drew Harwell, 'Contract Lawyers Face a Growing Invasion of Surveillance Programs that Monitor Their Work' (11 November 2021) *The Washington Post*.

worker wellness programs, activity reports, and color-coded charts.[20] They further argue that such 'limitless worker surveillance' has left millions of employees at the mercy of minute-by-minute monitoring by their employers that undermines fair labour rights, yet the existing legal framework offers few meaningful constraints.[21]

Indeed, algorithmically powered workplace surveillance and worker performance evaluation are often adopted by enterprises to increase efficiency and improve productivity, expand corporate capacity by shaping expectations, and condition the behaviours of employees.[22] However, the adoption of such systems not only intrudes upon the privacy and labour rights of employees,[23] but also harms their physical and mental well-being under a lasting framework of suppression.[24] In a larger context, the dominance of ADM tools for workplace surveillance and worker performance evaluation may distort the nature of the relationship between the employer and the employed and weaken psychological contracts, job engagement, and employee trust.[25] The gap in power asymmetry is institutionally widened by the systematic use of ADM tools that are neither reliable nor transparent and are also prone to bias and discrimination.[26]

Automated worker productivity monitoring and performance evaluation represents a system of mechanical enforcement without empathy or moral responsibility, which potentially dehumanizes the inherently person-to-person process of work management, reward and punishment allocation, and contractual interactions. These tools, cloaked in the promise of technologically supported management and data-driven efficiency, focus not on process but on results, which are observed and calculated based on arbitrary parameters or existing unfair and discriminatory practices. Given the black box nature of these tools, human supervisors, if any, cannot easily detect and address the mistakes and biases that arise in the ADM process. As a result, the use of algorithmic worker productivity monitoring and performance evaluation systems is increasingly contested and criticized for its controversial economic, social, legal, and political ramifications.

---

[20] *See generally* Ifeoma Ajunwa et al, 'Limitless Worker Surveillance' (2017) 105 *California Law Review* 101–42.

[21] Ibid.

[22] Anna M Pluta and A Rudawska, 'Holistic Approach to Human Resources and Organizational Acceleration' (2016) 29(2) *Journal of Organizational Change Management* 293–309.

[23] Alfred Benedikt Brendel et al, 'Ethical Management of Artificial Intelligence' (2021) 13(4) *Sustainability* 1974–92.

[24] Brian Patrick Green, 'Ethical Reflections on Artificial Intelligence' (2018) 6(2) *Scientia et Fiedes* 9–31.

[25] *See* Ashley Braganza et al, 'Productive Employment and Decent Work: The Impact of AI Adoption on Psychological Contracts, Job Engagement and Employee Trust' (2021) 131 *Journal of Business Research* 485–94.

[26] *See generally* Citron and Pasquale, 'The Scored Society: Due Process for Automated Predictions'; Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge: Harvard University Press, 2016).

10.3 SORTING TEACHERS OUT? UNPACKING *HOUSTON FEDERATION OF TEACHERS V HOUSTON INDEPENDENT SCHOOL DISTRICT*

Concerns over algorithmic worker productivity monitoring and performance evaluation systems came to light in the recent lawsuit over the use of VAMs in the United States – *Houston Federation of Teachers v Houston Independent School District*.[27] This case presents yet another controversial dimension of algorithmic worker productivity monitoring and performance evaluation in the education sector. *Houston Federation of Teachers v Houston Independent School District* involves the implementation of VAMs by the Houston Independent School District that algorithmically link a teacher's contributions to students' growth on standardized tests, the results of which inform decisions on teachers' tenure or contract (non)renewal. In 2011, the Houston Independent School District, citing low value-added scores, refused to renew its contract with more than 200 teachers. The VAM is proprietary and is not disclosed to those affected, precluding them from gaining an understanding of the internal logic and decision-making processes at work and causing serious harm to due process rights. Similar practices prevail across the United States following the enactment of the 2002 No Child Left Behind Act and the 2011 Race to the Top Act, in conjunction with other federal policy actions. Before the 2017 summary judgment rendered by the Court in *Houston Federation of Teachers v Houston Independent School District*, which ruled in favour of the affected teachers, federal constitutional challenges against the use of VAMs for termination or nonrenewal of teachers' contracts were generally rejected. Nevertheless, the case was subsequently settled and has interestingly received little attention. This chapter unpacks the case and endeavours to offer a critical analysis of its legal and policy ramifications.

Since 2010, the Houston Independent School District has applied a data-driven approach to monitor and evaluate teachers' performance with the aim to enhance the effectiveness of teaching from an outcome-based perspective. The algorithmically powered evaluation system implemented by the Houston Independent School District has three appraisal criteria – instructional practice, professional expectations, and student performance.[28] To narrow down the parameters for discussion, it should be noted that the primary focus of the case, *Houston Federation of Teachers v Houston Independent School District*, resides in the third component – student performance. Under the algorithmic work performance evaluation system, it is assumed that student growth and improvement in standardized test scores could appropriately reflect a specific teacher's impact on (or added value to) individual

---

[27] *Hous. Fed'n of Teachers v Hous. Indep. Sch. Dist.*, 251 F. Supp. 3d 1168, at 1171 (S.D. Tex. 2017).
[28] Ibid.

student performance, which is known as the VAM for teaching evaluations.[29] By implementing this system, student growth is calculated using the Educational Value-Added Assessment System (EVAAS), a proprietary statistical model developed by a private software company, SAS, and licensed for use by the Houston Independent School District.[30] This automated teacher evaluation system works by comparing the average test score growth of students taught by the teacher being evaluated with the statewide average for students in the same grade or course. The score is then processed by SAS's proprietary algorithmic program and subsequently sorted into an effectiveness rating system.[31]

In essence, under the VAM model, a teacher's algorithmically generated score was based on comparing the average growth of student test scores of the specific teacher compared to the average number state-wide, and the score was then converted to a test statistic called the Teacher Gain Index.[32] This measure was used to classify teachers into five levels of performance, ranging from 'well above' to 'well below' average.[33] It should be noted that the automated teacher evaluation system was initially used to inform and determine teacher bonuses, but as later implemented by the Houston Independent School District, the algorithmic system was used to automate sanctions on employed teachers for low student performance on standardized tests.[34] The Houston Independent School District declared in 2012 its management goal of ensuring that 'no more than 15% of teachers with ratings of ineffective are retained', and around 25 per cent of the 'ineffective teachers' were 'exited'.[35]

The plaintiff in this case, Houston Federation of Teachers, argued that the use of EVAAS violated the following elements of the Fourteenth Amendment.[36] First, the use of EVAAS violates the procedural due process right of the plaintiff because of the lack of sufficient information needed to meaningfully challenge terminations of contracts based on low EVAAS scores. Second, the substantive due process right is also violated, as there is no rational relationship between EVAAS scores and the Houston Independent School District's goal of employing effective teachers. Furthermore, since the EVAAS system is too vague to provide notice to teachers regarding how to achieve higher ratings and avoid adverse employment consequences, the use of EVAAS again violates the plaintiff's substantive due process right. Third, the plaintiff's right to equal protection is harmed by the Houston

---

[29] Ibid, 1172.
[30] Ibid.
[31] Ibid.
[32] Ibid.
[33] Ibid.
[34] Ibid, 1174.
[35] Ibid, 1174–75.
[36] Ibid, 1172–73.

Independent School District's policy of aligning teachers' instructional performance ratings with their EVAAS scores.

The court began its analysis with the plaintiff's protected property interests.[37] Referring to past jurisprudence, the court notes that, regardless of their employment status under probationary, term, or continuing contract, teachers generally have a protected property interest under their respective employment contracts (either during the term of the contract or under continued employment, according to the type of contract).[38] In this sense, the teachers who were adversely impacted by the use of EVAAS in the present case have a constitutionally protected property interest derived from the contractual relationship. The court denied the Houston Independent School District's argument that 'a due process plaintiff must show actual deprivation of a constitutional right'.[39] Importantly, the plaintiff in the present case sought 'a declaratory judgment and permanent injunction' barring the use of EVAAS in determining the renewal or termination of teacher contracts rather than monetary compensation and seeking an institutional and systematic outcome. According to past jurisprudence relevant to this case, '[o]ne does not have to await the consummation of threatened injury to obtain preventive relief'. Such a statement recommends that a demonstration of 'realistic danger' be sufficient.[40] As the facts of the case demonstrate a relationship between EVAAS scores and teacher employment termination, the court found that the VAM evaluation system 'poses a realistic threat to protected property interests' for those teachers.[41]

The court then turned to the procedural due process issue, which consists of the core value of 'the opportunity to be heard at a meaningful time and in a meaningful manner' to ensure that governmental decisions are fair and accurate.[42] The Houston Federation of Teachers argued that the Houston Independent School District failed the minimum procedural due process standard to provide 'the cause for [the teacher's] termination in sufficient detail so as to enable [the teacher] to show any error that may exist'. The algorithms and data used for the EVAAS evaluation system were proprietary and remained unavailable and inaccessible to the teachers who were affected, and the accuracy of scores could not be verified.[43] To address this issue, the court first acknowledged that, as the Houston Independent School District

---

[37] Ibid, 1173 ('The Fourteenth Amendment prohibits a state from depriving any person of life, liberty, or property without due process of law ... To evaluate such a claim, a court must first consider whether there is sufficient evidence implicating a protected property right in plaintiff's employment').

[38] Ibid (Citing *Frazier v Garrison I.S.D.*, 980 F.2d 1514, 1529 (5th Cir. 1993)).

[39] Ibid, 1174 (HISD had cited *Villanueva v McInnis*, 723 F.2d 414, 418–19 (5th Cir. 1984)).

[40] Ibid (Citing *Pennsylvania v West Virginia*, 262 U.S. 553, 593, 43 S.Ct. 658, 67 L.Ed. 1117 (1923); *Pennell v City of San Jose*, 485 U.S. 1, 8, 108 S.Ct. 849, 99 L.Ed.2d 1 (1988)).

[41] Ibid, 1175.

[42] Ibid, 1175–76.

[43] Ibid, 1172, 1176–77 (Citing *Ferguson v Thomas*, 430 F.2d 852 (5th Cir. 1970), the court has deemed that in the case of public school teacher termination, the minimum standards of procedural due process include the rights to

had admitted, the algorithms were retained by SAS as a trade secret, prohibiting access by the teachers as well as the Houston Independent School District, and any efforts to replicate the scores would fail. Furthermore, the calculation of EVAAS scores may be erroneous due to mistakes in the data or the algorithm code itself. Such mistakes could not be promptly corrected, and any reanalysis would potentially affect all other teachers' scores.[44]

The court then agreed to the plaintiff's application of the following standard from *Banks v. Federal Aviation Admin.*, 687 F.2d 92 (5th Cir. 1982), that 'due process required an opportunity by the controllers to test on their own behalf to evaluate the accuracy of the government-sponsored tests'.[45] When a potential violation of constitutional rights arises from a policy that concerns trade secrets, 'the proper remedy is to overturn the policy, while leaving the trade secrets intact'.[46] Even if the Houston Independent School District had provided the teachers some basic information (e.g., a general explanation of the EVAAS test methods) under the standard adopted in *Banks v Federal Aviation Admin.*, the measure still falls short of due process, since it does not change the fact that the teachers are unable to verify or replicate the EVAAS scores.[47] Since it is nearly impossible for the teachers to obtain or ensure accurate EVAAS scores and they are therefore 'unfairly subject to mistaken deprivation of constitutionally protected property interests in their jobs', the Houston Independent School District was denied summary judgment on this procedural due process claim.[48]

The issues involved in the substantive due process are twofold. The first issue relates to whether the challenged measure had a rational basis.[49] The Houston Federation of Teachers argued that EVAAS went against the protection of substantive due process, since there was no rational relationship between EVAAS scores and the Houston Independent School District's goal of 'having an effective teacher in every [Houston Independent School District] classroom so that every [Houston

---

(1) be advised of the cause for his termination in sufficient detail so as to enable him to show any error that may exist;

(2) be advised of the names and testimony of the witnesses against him;

(3) a meaningful opportunity to be heard in his own defense within a reasonable time;

(4) a hearing before a tribunal that possesses some academic expertise and an apparent impartiality towards the charges).

[44] Ibid, 1177.

[45] Ibid, 1178 (In *Banks v Federal Aviation Admin.*, 687 F.2d 92 (5th Cir. 1982), two air traffic controllers were dismissed on the grounds of drug usage. However, their urine samples were subsequently destroyed and were unavailable for independent testing. The lab tests that showed traces of cocaine became the only evidence of drug use in the record. The Fifth Circuit found that the controllers had been denied due process).

[46] Ibid, 1179.

[47] Ibid.

[48] Ibid, 1180.

[49] Ibid. (Citing *Finch v Fort Bend Independent School Dist.*, 333 F.3d 555, 563 (5th Cir. 2003), the challenged law or practice should have 'a rational means of advancing a legitimate governmental purpose').

Independent School District] student is set up for success'.[50] However, the court cited several examples of case law which supported the argument that a rational relationship existed in the present case and that 'the loose constitutional standard of rationality allows governments to use blunt tools which may produce only marginal results'.[51] The second issue surrounding substantive due process concerned vagueness. The general standard for unconstitutional vagueness is whether a measure 'fail[s] to provide the kind of notice that will enable ordinary people to understand what conduct it prohibits' or 'authorize[s] and even encourage[s] arbitrary and discriminatory enforcement'.[52] On the other hand, the court also acknowledged that a lesser degree of specificity is required in civil cases and that 'broad and general regulations are not necessarily vague'.[53] The court determined that the disputed measure in the present case was not vague, as the teachers who were impacted had been noticed or advised of the general information and possible effect of the use of the EVAAS evaluation system by their institutions.[54]

Finally, the court reviewed the plaintiff's equal protection claim. If a measure lacks a rational basis for the difference in treatment, that is, if the classification system used to justify the different treatment fails to rationally relate to a legitimate governmental objective, it may violate the Equal Protection Clause.[55] However, in this present case, the court denied the plaintiff's claim that the EVAAS rating scores represented a classification system. Even if they had, the court deemed that a rational basis existed, as explored with regard to the substantive due process claims.[56] In summary, the Houston Independent School District's motion for summary judgment on the procedural due process claim was denied, but summary judgment on all other claims was granted.[57]

10.4 JUDICIAL REVIEW AS ALGORITHMIC GOVERNANCE? CONTROVERSIES, RAMIFICATIONS, AND CRITICAL REFLECTIONS

It should be noted that, before the summary judgment ruling was reached in *Houston Federation of Teachers v Houston Independent School District*, some existing literature mentioned the issue of policy failures within the Houston

---

[50] Ibid.
[51] Ibid, 1180–82 (Citing *Cook v Bennett*, 792 F.3d 1294 (11th Cir. 2015); *Wagner v Haslam*, 112 F.Supp.3d 673 (M.D.Tenn. 2015); *Trout v Knox* Cty. Brd. of Educ., 163 F.Supp.3d 492 (E.D. Tenn. 2016)).
[52] Ibid, 1182 (Citing *City of Chicago v Morales*, 527 U.S. 41, 56, 119 S.Ct. 1849, 144 L.Ed.2d 67 (1999)).
[53] Ibid.
[54] Ibid, 1182–83.
[55] Ibid, 1183.
[56] Ibid.
[57] Ibid.

Independent School District's algorithmic work performance evaluation systems and the subsequent measures implemented on the teachers who were adversely affected. Some policies have noted that, while high-quality teachers can greatly benefit students, the 'effectiveness' of teachers may be difficult to assess because it correlates with non-observable characteristics.[58] To address the challenges of teacher evaluation and management, better information on real-world quality contributes to the productiveness of personnel policies and management decisions, but the accuracy of such information and its correlation with student performance cannot be easily observed.[59]

Julie Cullen and others conducted an empirical study that compared the patterns of attrition before and after the implementation of the Houston Independent School District's automated work performance evaluation system as well as the relationship between these patterns and student achievement. These researchers found that, although the algorithmic work performance evaluation system seemingly improves the quality teacher workforce, as it increases the exit rate of low-performing teachers, the statistics that imply this relationship are exclusively more obvious in low-achieving schools, as opposed to middle- and high-achieving schools.[60] More importantly, Cullen et al. also found that the exits resulting from the automated work performance evaluation system were too poorly targeted to induce any meaningful gains in student achievement and net policy effects.[61] They further suggested that the Houston Independent School District's algorithmic work performance measures were ineffective and proposed other substitutive measures via recruitment of new teachers or improvements in existing teaching employees.[62]

Bruce Baker and colleagues discussed legal controversies over unfair treatment and inadequate due process mechanisms since such automated teacher evaluation models are embedded with problematic features and parameters, such as non-negotiable final decisions, inaccessible information, and the use of imprecise data.[63] Algorithmic teacher evaluation models like EVAAS systems are prone to structural problems. First, such systems require that all 'objective measures of student achievement growth' be considered, which may lead to inaccurate outcomes, since the model disregards the fact that the validity and reliability of these measures can vary and that random errors or biases may occur, with no opportunity to question and reassess the validity of any measure.[64] Second, the standards for placing teachers into

---

[58] Cullen et al, 'The Compositional Effect of Rigorous Teacher Evaluation on Workforce Quality' (2021) 16(1) *Education Finance and Policy* 7–41.

[59] Ibid.

[60] Ibid, 21.

[61] Ibid, 21–26.

[62] Ibid, 26.

[63] Bruce D Baker et al, 'The Legal Consequences of Mandating High Stakes Decisions Based on Low Quality Information: Teacher Evaluation in the Race-to-the-Top Era' (2013) 21(5) *Education Policy Analysis Archives* 1–65 at 5.

[64] Ibid, 5–6.

effectiveness score bands and categories are unjustifiable, as the numerical cutoffs are rigid and temporally static. A difference in one point or percentile does not necessarily indicate any actual differences in the performance of the evaluated teachers. However, it can lead to a distinctly different effectiveness category and consequentially endanger a teacher's employment rights.[65] While models that are based on VAMs theoretically attempt to reflect student achievement growth that can be attributed (directly) to a specific instructor's teaching quality and performance, they can hardly succeed in making a fair connection in reality, since it is nearly impossible to discern whether the evaluation estimates have been contaminated by uncontrollable or biased factors, and the variation in ratings is quite broad.[66] By dismissing teachers under such an arbitrary evaluation system, possible violations of due process rights under the Fourteenth Amendment in the form of harm to liberty interests by adversely affecting teachers' employment or harm to property interests in continued employment may likely occur, as shown in *Houston Federation of Teachers v Houston Independent School District*. Likewise, VAMs may be challenged against procedural or substantive due process claims surrounding the technical flaws of value-added testing policies, including the instability of the reliability of those measures along with their questionable interpretations, the doubtful validity of the measure and the extent to which it proves a specific teacher's influence over student achievement, and the accessibility and understandability of the measures to an evaluated teacher as well as the teacher's ability to control relevant factors.[67] VAMs are limited measures in terms of properly assessing teacher 'effectiveness', and 'it would be foolish to impose on these measures, rigid, overly precise high stakes decision frameworks'.[68]

In *Houston Federation of Teachers v Houston Independent School District*, the court found a procedural due process violation mainly because those teachers had no way to replicate and challenge their scores. In addition, the court also indicated concern over the accuracy issue of the algorithmic tool, which has never been verified or audited whatsoever.[69] In a way, the case marks 'an unprecedented development in VAM litigation', and as a result, VAMs used in other states and elsewhere in education management policies should garner greater interest and concern.[70] As per the judge in *Houston Federation of Teachers v Houston Independent School District*, when a government agency adopts a management policy of making highly consequential decisions with regard to employment renewal

---

[65] Ibid, 6.
[66] Ibid, 9.
[67] Ibid, 10–11.
[68] Ibid, 18.
[69] *Hous. Fed'n of Teachers v Hous. Indep. Sch. Dist.*, 251 F. Supp. 3d 1168, at 1177–80 (S.D. Tex. 2017).
[70] Audrey Amrein-Beardsley, 'The Education Value-Added Assessment System (EVAAS) on Trial: A Precedent-Setting Lawsuit with Implications for Policy and Practice' (2019) *eJournal of Education Policy* 1–11 at 7.

and termination based on opaque algorithms incompatible with minimum due process, the court is poised to offer a proper remedy to overturn the use of this algorithmic tool.[71] After *Houston Federation of Teachers v Houston Independent School District*, other states and districts in similar situations have been strongly incentivized to reconsider their use of the EVAAS algorithmic teacher evaluation system or other VAMs by separating consequential personnel decisions from evaluation estimates to avoid potential claims of due process violations.[72] On the other hand, the use of EVAAS (or other VAMs) for low-stakes purposes should also be reconsidered, as the court in *Houston Federation of Teachers v Houston Independent School District* expressed its concern over the actual extent to which 'teachers might understand their EVAAS estimates so as to use them to improve upon their practice'.[73]

As a number of states have adopted automated teacher performance evaluation systems that allow VAM data to be the sole or primary consideration in the decision-making process with regard to review, renewal, or termination of employment contracts, the outcome of *Houston Federation of Teachers v Houston Independent School District* and its legal and policy ramifications might demonstrate a broad reach.[74] Indeed, the lawsuit itself has opened up the possibility for teachers (at least those employed in public schools) to seek remedies for the controversial use of VAMs and other algorithmic teacher performance evaluation systems, especially when the teachers who had challenged such systems had been generally unsuccessful. *Houston Federation of Teachers v Houston Independent School District*, despite being ultimately settled, paves a viable litigation path to challenge the increasingly automated worker performance evaluation in the education sector.

Now it seems possible that due process challenges (at least procedural due process) will persist, as the court drew attention to 'the fact that procedural due process requires a hearing to determine if a district's decision to terminate employment is both fair and accurate'.[75] As noted by Mark Paige and Audrey Amrein-Beardsley, *Houston Federation of Teachers v Houston Independent School District* raised awareness about concerns over government transparency and 'control of private, for-profit corporations engaged in providing a public good',[76] especially with regard to the use of black box algorithmic decision-making tools in the

---

71 *Hous. Fed'n of Teachers v Hous. Indep. Sch. Dist.*, 251 F. Supp. 3d 1168, at 1179 (S.D. Tex. 2017).

72 Amrein-Beardsley, 'The Education Value-Added Assessment System (EVAAS) on Trial: A Precedent-Setting Lawsuit with Implications for Policy and Practice', 8.

73 Ibid; *Hous. Fed'n of Teachers v Hous. Indep. Sch. Dist.*, 251 F. Supp. 3d 1168, at 1171 (S.D. Tex. 2017).

74 Mark A Paige and Audrey Amrein-Beardsley, '"Houston, We Have a Lawsuit": A Cautionary Tale for the Implementation of Value-Added Models for High-Stakes Employment Decisions' (2020) 49(5) *Educational Researcher* 350–59.

75 Ibid, 355.

76 Ibid.

education sector. The case strongly questions the reliability of the EVAAS system in assessing and improving teacher quality, especially since undetectable errors can lead to significant consequences, including calls for public scrutiny, and seems to offer the potential to compel policymakers and practitioners to both re-examine and reflect on the level of importance (if any) VAM estimations should play in personnel decisions. An independent study on automated decision-making on the basis of personal data in the context of comparison between European Union and United States, which has been submitted to the European Commission's Directorate-General for Justice and Consumers, also underlines that the court's decision in *Houston Federation of Teachers v Houston Independent School District* 'demonstrates that the Due Process Clause can serve as an important safeguard when automated decisions have a legal effect'.[77]

Nevertheless, regrettably, the controversial characteristics of such worker performance evaluation algorithms – the proprietary, black box, inaccessible, and unexplainable decision-making routes[78] – have not occupied a critical spot of concern for legal challenges. The lawsuit in no way means that VAMs and other algorithmic worker evaluation systems should be systematically examined, fixed, or abandoned. As noted, the dominance of automated tools for workplace surveillance and worker performance evaluation may distort the nature of the relationship between the employer and the employed and weaken psychological contracts, job engagement, and employee trust. The gap in power asymmetry has been institutionally widened by the systematic use of algorithmic tools that are neither reliable nor transparent and are also prone to bias and discrimination. All of these issues remain out of the scope of examination in terms of judicial review. In line with this argument, Ryan Calo and Danielle Citron point out the problems of this growingly Automated State, noting a number of controversial cases, including *Houston Federation of Teachers v Houston Independent School District*. The researchers cite the 'looming legitimacy crisis' and call for a reconceptualization and new vision of the modern administrative state in the algorithmic society.[79] They argue that, while scholarly have been asking how we might ensure that these automated tools can align with the existing legal contours such as due process, broader and structural questions on the legitimacy of automating public power remain unanswered.[80] Indeed, without proper gatekeeping or accountability mechanisms, the growing algorithmization of worker performance evaluation can go unharnessed, especially when such practices are

---

[77] Gabriela Bodea et al, *Automated Decision-Making on the Basis of Personal Data that Has Been Transferred from the EU to Companies Certified under the EU-U.S. Privacy Shield Fact-Finding and Assessment of Safeguards Provided by U.S. Law* (Final Report submitted to European Commission Directorate-General for Justice and Consumers Directorate C: Fundamental Rights and Rule of Law Unit C.4 International Data Flows and Protection, 2018) 92.

[78] *See* Hannah Bloch-Wehba, 'Access to Algorithms' (2020) 88 *Fordham Law Review* 1265–314.

[79] *See generally* Ryan Calo and Danielle Keats Citron, 'The Automated Administrative State: A Crisis of Legitimacy' (2021) 70(4) *Emory Law Journal* 797–845.

[80] Ibid.

spreading at such a rapid rate that regulators struggle to catch up and employees face widening power asymmetry.

10.5 CONCLUSION

Automated worker productivity monitoring and performance evaluation indicate a system of mechanical enforcement, if not suppression, which practically dehumanizes the inherent person-to-person process of work management without empathy[81] or moral responsibility. The algorithmic tool, as implemented widely in *Houston Federation of Teachers v Houston Independent School District*, focuses not on process but on results, which are observed and calculated based on arbitrary parameters or the existing unfair and discriminatory practices. Cloaked in technologically supported management and data-driven efficiency, algorithmic worker productivity monitoring and performance evaluation systems create and likely perpetuate a way to rationalize automatic layoffs without meaningful human supervision. Given the black box characteristics of these automated systems, human supervisors cannot easily detect and address mistakes and biases in practice.

The court in *Houston Federation of Teachers v Houston Independent School District* provides a baseline for future challenges in the use of these algorithmic worker productivity monitoring and performance evaluation systems by public authority (not the private sector). Here, judicial review appears necessary and to some extent effective to ensure a basic level of due process protection. However, the ruling arguably only scratches the surface of the growing automation of workplace management and control and the resulting power asymmetry. Indeed, it merely touches on procedural due process and leaves intact critical questions such as algorithmic transparency, explainability, and accountability. In this sense, judicial review, with the conventional understanding of due process and rule of law, cannot readily serve as an adequate form of algorithmic governance that can harness data-driven worker evaluation systems.

Again, salient in *Houston Federation of Teachers v Houston Independent School District*, the affected teachers encountered formidable challenges to examine proprietary algorithms developed by a private company to assess public school teacher performance and make consequential employment decisions. The teachers who were 'exited' had no access to the algorithmic systems and received little explanation or context for their termination. Experts who were offered limited access to the source codes of the EVAAS also concluded that the teachers had no way to meaningfully verify their scores assigned by the system. The algorithmization of worker performance evaluation and surveillance is not and will not be limited to specific industry sectors or incomes. Individuals in other professions may not enjoy comparable social and economic support systems as the teachers in *Houston*

---

[81] See also Chapter 9 in this book.

Federation of Teachers v Houston Independent School District to pursue judicial review and remedies, and the algorithmic injustice they face may never be addressed.

Finally, the increasingly blurred line between public and private authorities and their intertwined collaboration in designing and applying these algorithmic tools pose new threats to the already weak effectiveness of rule of law and due process protection under the existing legal framework.[82] Any due process examination falls short at the interface of public and private collaboration, since the proprietary algorithms held by the private company constitute a black box barrier. The court in *Houston Federation of Teachers v Houston Independent School District* expressed significant concerns over the accuracy of the algorithmic system, noting that the entire algorithmic system was flawed with inaccuracies and was like a house of cards – the 'wrong score of a single teacher could alter the scores of every other teacher in the district' and 'the accuracy of one score hinges upon the accuracy of all'.[83] However, the black box process and automation itself were not considered problematic at all. Due process is needed in the context of the growing algorithmization of worker monitoring and evaluation so that affected employees may be able to partially ascertain the rationale behind data-driven decisions and control programs,[84] but it must be reconceptualized and retooled to protect against the above-mentioned threats to the new power dynamics.

---

[82] The court dismissed the substantive due process claim because the 'loose constitutional standard of rationality allows government to use blunt tools which may produce marginal results'. The court hinted that the algorithmic evaluation system would pass the rationality test even if the system and scores were accurate only a little over half of the time. *Hous. Fed'n of Teachers v Hous. Indep. Sch. Dist.*, 251 F. Supp. 3d 1168, at 1178 (S.D. Tex. 2017).

[83] Ibid.

[84] Sonia K Katyal, 'Democracy & Distrust in an Era of Artificial Intelligence' (2022) 151(2) *Daedalus* 322–34 at 331; *see also* Aziz Z Huq, 'Constitutional Rights in the Machine-Learning State' (2020) 105 *Cornell Law Review* 1875–954.