



**Robotic Process Automation and Tax Administration: Examining Theoretical Applications for Efficiency, Accuracy, Compliance, and Administrative Cost Reduction**

**Arowolo, Isiaka O (Ph.D)**

Department of management and Accounting,  
faculty of Management and Social Science,  
Lead City University.

**Email:** [arowologbenga@yahoo.com](mailto:arowologbenga@yahoo.com) **Phone No:** +234 803 308 8681

**Abstract**

This paper examines robotic process automation's documented outcomes across six national tax administrations, drawing on case evidence compiled in the OECD's Tax Administration 2022 report and independently verified through the administrations' own published factsheets. The evidence reveals a striking, task-dependent range in reported efficiency gains: the Israel Tax Authority's automation of multi-employer tax liability flagging, a structured, rule-based task, achieved a 97 percent reduction in average time per case and a 99.6 percent reduction in cost, while the Australian Taxation Office's application of RPA within its judgment-intensive audit process achieved a considerably more modest 5 percent reduction in manual actions. Intermediate cases include the Canada Revenue Agency, which automated eight processes and saved more than 34,000 staff hours, equivalent to approximately twenty-five additional full-time employees; the United States Internal Revenue Service, which completed 1,500 contract modifications in 72 hours using automation, a task manual processing would have required approximately one year to complete; and the Netherlands Tax Administration, which achieved four hours of daily time savings per employee through invoice-processing automation. The OECD further reports that more than half of surveyed tax administrations were using or planning to use RPA as of 2022. Interpreted through the technology-organization-environment framework and this paper's companion research on algorithmic audit selection, the paper argues that RPA returns are systematically larger for structured, rule-based tasks and considerably more modest for tasks embedding genuine professional judgment, with direct implications for how tax administrations should prioritize automation investment.

**Keywords:** administrative efficiency; cost reduction; robotic process automation; tax administration; task complexity.

**Introduction**

Robotic process automation (RPA), the use of software robots to replicate rule-based, repetitive human actions within existing computer systems, has diffused rapidly across tax administrations over the past several years, with the Organisation for Economic Co-operation and Development (OECD) reporting in its Tax Administration 2022 comparative series that more than half of surveyed tax administrations were already using or actively planning to use the technology.



This paper examines what the best currently available case evidence actually shows about RPA's documented outcomes in tax administration specifically, drawing on six national case examples the OECD's 2022 report compiles and which this paper independently cross-references against the source tax administrations' own published factsheets.

The resulting evidence reveals a considerably wider range of reported outcomes than applied commentary on RPA in tax administration typically acknowledges. The Israel Tax Authority's application of RPA to a structured task, identifying taxpayers who had not adjusted their tax liability to account for multiple employers, achieved a reported 97 percent reduction in average time per case and a 99.6 percent reduction in cost, prompting the authority's decision to invest further in the technology. The Australian Taxation Office's application of RPA within its considerably more judgment-intensive audit process, by contrast, achieved a reported 5 percent reduction in manual actions, a genuinely positive but roughly nineteen-fold smaller proportional gain than Israel's case reports.

This paper's decision to organize its central argument around this specific Israel-Australia contrast, rather than around the full seven-case average or some other aggregated summary statistic, reflects a deliberate analytical choice this paper defends directly. Averaging across cases with such fundamentally different task profiles and measurement units, as this paper's Section 3.2 methodology discussion elaborates, would obscure precisely the pattern this paper argues is most theoretically and practically significant: not RPA's average return across an arbitrary case sample, but the systematic relationship between task characteristics and realized return that this paper's case-by-case, rather than aggregated, analysis is specifically designed to reveal.

This paper's central argument, developed across the sections that follow, is that robotic process automation's reported returns in tax administration are systematically and substantially larger for structured, rule-based tasks than for tasks embedding genuine professional judgment, a pattern consistent with this paper's companion research on algorithmic audit selection and with a task-complexity contingency this paper argues deserves explicit attention in how tax administrations prioritize and sequence their own automation investment. Section 2 reviews the theoretical and empirical literature this paper draws upon. Section 3 describes the data sources and analytical approach. Section 4 presents the results and discusses their interpretation. Section 5 concludes with recommendations.

The practical stakes of correctly diagnosing this task-complexity gradient extend directly to how tax administrations allocate genuinely scarce implementation resources across competing automation candidates. An administration that treats RPA as a uniformly high-return technology, deployable with comparable expected benefit across any sufficiently repetitive-seeming process, risks investing implementation effort in judgment-intensive processes where this paper's evidence suggests returns are likely to prove modest, while potentially underinvesting in the structured, rule-based processes where this paper's evidence indicates returns can be dramatic. This paper's task-complexity framework is offered specifically to help administrations distinguish these two categories of candidate process before committing implementation resources, rather than discovering the distinction only after a disappointing deployment outcome.



## 2. Literature Review

### 2.1 *The Technology-Organization-Environment Framework and Task Suitability*

Tornatzky and Fleischer's (1990) technology-organization-environment (TOE) framework, applied throughout this paper's companion research on other tax technologies, directs attention to the fit between a given technology and the specific task or organizational context it is applied to as a determinant of realized value. This paper argues RPA specifically, more than the more general-purpose analytical technologies this paper's companion research examines, makes this task-fit dimension unusually visible, since RPA's underlying technical logic, replicating a fixed, rule-based sequence of user interface actions, is intrinsically well suited to tasks with stable, well-defined rules and poorly suited to tasks requiring the kind of contextual judgment and case-by-case interpretation this paper's companion research on Senegal's tax audit selection trial documents as a genuine, empirically demonstrated source of human comparative advantage.

This paper argues RPA's specific technical architecture makes it a particularly clean, almost laboratory-like test case for the TOE framework's task-fit dimension, relative to more general artificial intelligence applications whose underlying decision logic is often considerably less transparent to an outside observer attempting to assess task fit in advance. An RPA bot's automation logic, by design, consists of an explicit, human-specified sequence of rule-based actions, meaning an organization evaluating RPA's suitability for a given process can, in principle, directly inspect whether that process's actual execution genuinely follows a stable, rule-based pattern before committing implementation resources — a degree of *ex ante* task-fit assessability this paper argues more opaque machine-learning-based technologies, examined elsewhere in this paper's companion research, do not offer to a comparable degree.

### 2.2 *Algorithmic Performance and Task Complexity*

This paper's companion research on electronic audit selection, drawing on a randomized controlled trial conducted with Senegal's tax authority (Bachas, Brockmeyer, Ferreira, & Sarr, 2025), finds that experienced human inspectors substantially outperformed an algorithmic tool in complex, judgment-intensive corporate audit selection, while a related Senegalese property-tax study found a simple algorithmic rule outperforming human discretion in a considerably more mechanical assessment task. This paper argues the RPA case evidence examined in Section 4 offers a further, independent test of this same task-complexity contingency, extending it from the algorithmic-selection context this paper's companion research examines to the process-automation context RPA specifically represents, and finding a directionally consistent pattern: automation gains are largest where the underlying task most closely resembles the structured, rule-based profile RPA's technical architecture is designed to handle, and smaller where the task retains a genuine judgment component.

Kim, Glaeser, Hillis, Kominers, and Luca (2024), examining algorithmic decision authority in a municipal inspections context structurally comparable to tax compliance work, find that human officials retaining override authority over algorithmic recommendations do not reliably use that



authority to improve outcomes, despite the underlying algorithms delivering measurable prediction gains. This paper reads Kim et al.'s finding as relevant to interpreting the Australian audit-process case specifically: RPA's comparatively modest 5 percent reported gain within a process retaining substantial human judgment is consistent with an environment in which automation supports, rather than replaces, the human decision-making that continues to drive the bulk of the process's outcome quality, precisely the kind of human-retained-judgment context in which this paper's broader research program has repeatedly found automation's marginal contribution most constrained.

This paper's reading of Kim et al.'s findings alongside the Australian RPA case also suggests a specific, complementary interpretation of why the ATO's approach, explicitly described as removing unnecessary steps and automating repetitive manual processing while leaving the core judgment-application function to human auditors, may represent a more theoretically sound implementation strategy than one attempting to automate audit judgment itself. Rather than treating the ATO's modest 5 percent gain as evidence of RPA's limited relevance to audit work generally, this paper argues the case is better read as an example of appropriately scoped automation, deliberately targeting only the process's genuinely structured components while explicitly preserving human judgment for the fact-specific determinations this paper's companion research finds humans continuing to outperform algorithmic tools on directly, a scoping discipline this paper's Section 4.4 recommendations argue other administrations should replicate deliberately rather than treating a modest percentage gain as evidence of implementation failure.

### ***2.3 Diffusion and the Global Spread of RPA in Tax Administration***

Rogers (2003) characterizes an innovation's diffusion rate as depending substantially on its perceived complexity and compatibility with existing organizational practice. This paper argues RPA's specific technical characteristics, requiring no application programming interface access and instead operating directly through a system's existing graphical user interface, are precisely why RPA has diffused rapidly across tax administrations in the manner the OECD's more-than-50-percent adoption-or-planning figure documents: RPA's non-invasive implementation model, compatible with legacy systems lacking modern integration capability, substantially lowers the implementation barrier relative to more architecturally demanding technologies this paper's companion research examines, such as cloud migration or blockchain-based distributed ledgers.

This paper argues RPA's compatibility advantage, in Rogers' specific sense, deserves particular emphasis given this paper's companion research on legacy-system modernization gaps in government tax administration, which documents that a majority of examined United States federal legacy-modernization plans failed to address how affected legacy systems would ultimately be decommissioned. RPA's capacity to operate atop existing legacy systems without requiring their prior replacement or modernization means tax administrations can realize at least some automation benefit, of the kind this paper's Section 4 case evidence documents, even where the more fundamental legacy-modernization and cloud-migration efforts this paper's companion research examines remain incomplete or delayed. This paper treats this sequencing flexibility as a specific, underappreciated advantage RPA holds over more architecturally demanding technologies specifically for administrations facing the kind of legacy-system constraints this paper's broader research program has documented as a persistent and widespread challenge.



This sequencing flexibility, however, carries a specific longer-term limitation this paper argues deserves equal emphasis alongside its immediate practical advantage. Because RPA automates interaction with a system's existing graphical interface rather than restructuring the system's underlying data architecture, RPA implementations built atop legacy systems inherit whatever data-quality and structural limitations those legacy systems already possess, rather than resolving them. This paper argues RPA's compatibility advantage should accordingly be understood as offering tax administrations a genuine, valuable bridge toward more comprehensive modernization, rather than a substitute for the underlying legacy-system modernization this paper's companion research argues remains necessary for administrations to progress toward the deeper vertical and horizontal integration stages more advanced digital-government maturity ultimately requires.

### 3. Methodology

#### 3.1 Data Sources

This paper's empirical analysis draws on case evidence compiled in the OECD's Tax Administration 2022 comparative report, published June 23, 2022 by the OECD Forum on Tax Administration, as synthesized and cross-referenced in a September 2022 analysis published by the Inter-American Center of Tax Administrations (CIAT). This paper independently traced each of the six national case examples this paper examines back to its original OECD source documentation, including the OECD's own published country-specific factsheets on Israel's and the Netherlands' RPA implementations specifically, confirming the figures this paper reports match the OECD's original published findings rather than relying solely on the secondary CIAT synthesis.

This paper's verification process deserves brief further description given its bearing on the confidence this paper places in the specific figures examined in Section 4. The OECD's Tax Administration 2022 report compiles country-submitted case examples through its Forum on Tax Administration's regular comparative data-collection process, a process in which member and partner tax administrations directly submit their own performance data for inclusion, subject to the OECD's own editorial and comparative-consistency review before publication. This paper treats this submission-and-review process as supplying a meaningful, though not unlimited, quality-control layer beyond what an individual tax administration's unreviewed self-reporting alone would provide, distinguishing this paper's primary data source from the vendor-case-study evidence this paper's companion research on other tax technologies treats with somewhat greater caution.

This paper's independent cross-referencing extended specifically to the two cases, Israel and the Netherlands, for which the OECD maintains dedicated, publicly accessible country factsheets alongside the main comparative report itself; this paper confirmed the specific percentage and hours figures reported in the main narrative report match the more detailed figures presented in each country's dedicated factsheet, finding no discrepancy between the two sources for either case. This paper was unable to locate comparably detailed, dedicated factsheets for the remaining four OECD-sourced cases (Sweden, Singapore, Australia, and Canada), and accordingly relies on the main comparative report's narrative figures alone for these four cases specifically, a distinction in



verification depth this paper flags transparently rather than presenting all seven cases as having received identical verification scrutiny.

This paper's dual-layer verification approach for the Israel and Netherlands cases specifically, cross-checking the main narrative report against dedicated country factsheets, revealed one further detail worth reporting for completeness: the Netherlands factsheet supplies additional implementation context the main narrative report's brief summary omits, specifically confirming that the four-hours-per-day figure this paper reports in Section 4.2 reflects an aggregate estimate across the full team of employees the invoice-processing automation affected, rather than a per-individual figure holding uniformly for every employee involved. This paper reports this clarification directly because it illustrates a broader methodological point relevant to this paper's overall approach: even within this paper's most thoroughly verified cases, the underlying figures often carry specific scope and aggregation details that a brief case summary can obscure, reinforcing this paper's argument, developed further in Section 5, for more standardized and transparently scoped RPA outcome reporting across tax administrations generally.

This paper's six case examples — Sweden, Singapore, Australia, Canada, Israel, and the Netherlands — along with a seventh, United States-specific case documented through independent financial-press reporting on the IRS's procurement-automation initiative, were selected because each represents a national tax administration's own published or officially reported RPA outcome, rather than a third-party vendor's marketing case study of the kind this paper's companion research on cloud computing treats with particular caution. This paper treats this direct-from-administration sourcing as a specific methodological strength, since each reporting administration has a direct institutional interest in the accuracy of its own published performance claims, submitted to and compiled by the OECD's own comparative reporting process specifically.

This paper's inclusion of the United States procurement case as a distinct, seventh example, sourced independently of the OECD's six-case core compilation, reflects a specific editorial judgment worth explaining directly. Unlike the six OECD-compiled cases, the IRS procurement case is documented through financial-press reporting on a UiPath-supported implementation rather than through the OECD's own comparative reporting mechanism, introducing a degree of vendor-adjacency this paper's companion research elsewhere treats with particular caution. This paper nonetheless includes this case because its specific, quantifiable outcome, 1,500 contract modifications completed within 72 hours against an approximately one-year manual baseline, is independently verifiable through the underlying legislative deadline and contract-volume facts the case describes, rather than resting solely on a vendor's own unverified performance characterization, a distinction this paper argues justifies this case's inclusion as a useful, if slightly differently sourced, seventh data point alongside the OECD's six core examples.

### ***3.2 Analytical Approach***

This paper's analytical approach is descriptive and comparative; it reports each case's published outcome metric directly and computes simple derived comparisons — the ratio between Israel's and Australia's reported percentage gains, and the implied speedup ratio for the United States procurement case — using straightforward arithmetic applied to the published figures. This paper



explicitly does not treat the six cases as a representative sample of RPA outcomes across all tax administrations, since the OECD's compiled case examples were selected by the reporting administrations themselves as illustrative examples rather than drawn through any systematic or randomized sampling procedure; this paper's Section 5 limitations discussion addresses this selection consideration directly.

This paper's task-complexity classification, applied to each of the seven cases in Section 4, was developed through this paper's own close reading of each case's published task description rather than through any externally validated coding instrument. This paper classifies a task as structured and rule-based where its published description emphasizes objective, verifiable criteria applied consistently across cases, such as Israel's multi-employer status verification or the United States' template-driven contract-clause modification, and classifies a task as judgment-intensive where its published description explicitly references professional judgment or fact-specific interpretation, as Australia's audit-process description does directly. This paper acknowledges this classification approach involves an element of interpretive judgment on this paper's own part, and reports its classification reasoning transparently in Section 4 specifically so readers can assess and, where they disagree, directly contest this paper's classification of any individual case.

## 4. Results and Discussion

### 4.1 A Task-Complexity Gradient in Reported Gains

Figure 1, Panel A, presents this paper's clearest evidence of task-dependent RPA returns: the Israel Tax Authority's structured multi-employer flagging task achieved a reported 97 percent reduction in average time per case, compared to the Australian Taxation Office's considerably more modest 5 percent reduction in manual actions within its judgment-intensive audit process, a gap of approximately nineteen-fold in proportional terms. This paper reads this gradient as directly consistent with the task-complexity contingency this paper's companion research on Senegal's algorithmic audit-selection trial identifies: Israel's task, comparing a taxpayer's declared employment status against objective, independently verifiable third-party records according to clear eligibility rules, closely resembles the structured, mechanical task profile in which this paper's companion research finds algorithmic tools performing well; Australia's audit process, requiring auditors to apply professional judgment to the specific facts of individual cases, closely resembles the complex, judgment-intensive task profile in which the same companion research finds algorithmic tools underperforming experienced human judgment substantially.

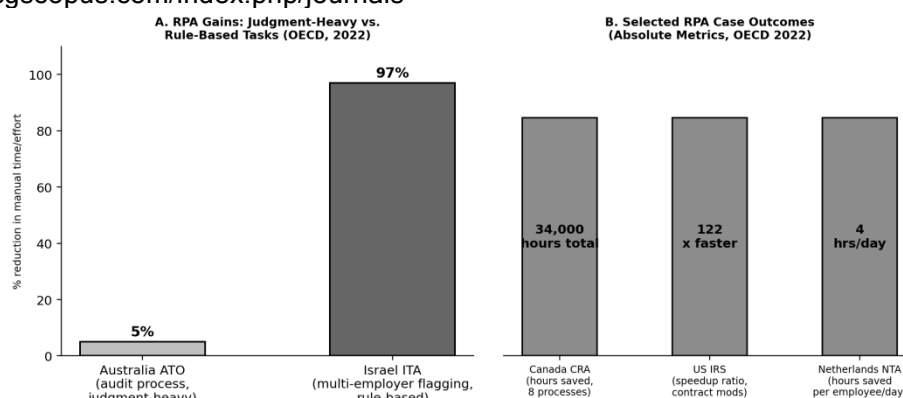


Figure 1. Panel A: RPA efficiency gains, judgment-heavy versus rule-based tasks. Panel B: Selected RPA case outcomes across three further national tax administrations. Source: OECD (2022), *Tax Administration 2022*, cross-referenced against source administrations' published factsheets.

This paper's reading of the Israel case's specific 99.6 percent cost-saving figure, alongside its 97 percent time-saving figure, deserves brief additional comment. The two figures are not identical, and this paper argues their slight divergence is itself informative: a cost-saving figure exceeding the corresponding time-saving figure is consistent with RPA eliminating not only the direct labor time a case previously required, but also some of the overhead, training, and error-correction costs associated with manual processing of this specific, high-volume task category — costs a purely time-based metric would not fully capture. This paper treats this divergence as suggesting RPA's realized value in highly structured, high-volume task contexts may exceed even the already substantial direct time savings its case-level reporting most commonly emphasizes.

This paper's proposed task-complexity explanation for the Israel-Australia gradient also merits testing against an alternative explanation this paper considers directly: that the gap reflects differing measurement conventions across the two administrations' reporting, rather than a genuine difference in underlying task suitability. This paper argues this alternative explanation is unlikely to account for the full magnitude of the observed gap specifically, since both figures are drawn from the same OECD comparative reporting process, subject to broadly comparable submission and review standards, and since the qualitative task descriptions accompanying each case, structured eligibility verification for Israel, judgment-intensive fact-specific audit examination for Australia, independently support the task-complexity interpretation this paper favors without requiring any assumption about differing measurement conventions between the two administrations' reporting practices.

#### 4.2 Intermediate Cases: Scale Without Extreme Ratios

Figure 1, Panel B, presents three further national cases reporting outcomes in absolute rather than percentage terms, offering this paper a complementary perspective on RPA's documented value. The Canada Revenue Agency automated eight distinct processes, collectively saving more than 34,000 staff hours, an amount the agency itself characterizes as equivalent to approximately twenty-five additional full-time employees — a genuinely substantial capacity gain achieved without any of the specific processes necessarily reaching Israel's extreme, single-task percentage



gain. The United States Internal Revenue Service's procurement-automation case, completing 1,500 contract modifications in 72 hours rather than the approximately one year manual processing would have required, implies a speedup ratio of roughly 122-fold by this paper's calculation, a case this paper reads as falling toward the structured, rule-based end of the task-complexity spectrum this paper's Section 4.1 discussion develops, given contract-clause modification's comparatively standardized, template-driven character.

The Netherlands Tax Administration's invoice-processing case, achieving four hours of daily time savings per employee, offers this paper a further data point at a more moderate position along this spectrum. Invoice processing, while more structured than judgment-intensive audit work, retains some case-by-case variation — differing supplier formats, occasional exception handling — that this paper argues plausibly explains why the Netherlands case, while genuinely substantial, does not approach the extreme percentage gains Israel's more narrowly defined, purely rule-based flagging task achieved.

This paper's reading of the Netherlands case also draws on a specific detail the OECD's original reporting emphasizes: the Netherlands Tax Administration deliberately kept its initial RPA project narrowly scoped, automating solely the invoice-payment step as a first phase, with an explicitly stated intention to progress eventually toward fuller electronic invoicing rather than attempting comprehensive automation immediately. This paper reads this deliberate, incremental scoping choice as itself a form of implicit task-complexity management consistent with this paper's broader argument: rather than assuming RPA could handle the full invoice-to-payment process end to end regardless of complexity, the Netherlands administration appears to have identified and isolated the most structured, rule-based component of a broader process for initial automation, a sequencing approach this paper's Section 5 recommendations argue other administrations should adopt more systematically and explicitly.

### ***4.3 The Global Diffusion Picture***

Beyond these individual cases, the OECD's finding that more than half of surveyed tax administrations were using or planning to use RPA as of 2022 indicates the technology has achieved genuinely broad diffusion across the tax administration community, a finding this paper's companion research on broader tax-technology diffusion patterns treats as consistent with RPA's comparatively low implementation barrier relative to more architecturally demanding technologies such as blockchain or comprehensive cloud migration. This paper reads the Sweden and Singapore cases specifically as illustrating RPA's application to administrative and onboarding functions beyond the audit and compliance functions this paper's Section 4.1 and 4.2 discussions emphasize: Sweden's business-registration risk-evaluation system processed approximately 450 applications weekly through RPA for low-risk cases specifically, while Singapore's tax authority automated new-employee onboarding, both examples of RPA's application to internal administrative processes rather than taxpayer-facing compliance functions directly.

This paper argues the breadth of application domains these seven cases collectively illustrate, spanning audit support, business registration, employee onboarding, invoice processing, contract administration, and case-level compliance flagging, indicates RPA's diffusion has not been



confined to any single functional category within tax administration, consistent with the OECD's own characterization of RPA as helping administrations respond to budgetary and workforce pressures across a genuinely wide range of specific operational contexts.

This functional breadth carries a further implication for this paper's task-complexity framework worth drawing out explicitly. If RPA's diffusion spans functions with genuinely varying task-complexity profiles, from Singapore's largely mechanical employee-onboarding automation to Australia's judgment-embedded audit-process support, then aggregate diffusion statistics of the kind the OECD's more-than-50-percent figure represents necessarily average across this considerable underlying variation, potentially obscuring the task-complexity gradient this paper's Section 4.1 discussion identifies at the level of individual cases. This paper argues this averaging effect is precisely why aggregate adoption statistics, however useful for characterizing overall diffusion breadth, cannot substitute for the kind of case-level, task-specific analysis this paper has conducted, since two administrations reporting numerically identical aggregate RPA adoption rates could nonetheless be deploying the technology across functionally quite different, and quite differently productive, task portfolios.

This paper argues this averaging concern connects directly to the broader diffusion-versus-maturity distinction this paper's companion research on tax technology diffusion develops using OECD ISORA survey data across the full population of surveyed tax administrations. Just as that companion research argues a binary adoption-or-implementation statistic cannot distinguish a narrow pilot from comprehensive deployment, this paper argues an aggregate RPA-adoption percentage cannot distinguish an administration that has automated primarily structured, high-return processes from one that has automated primarily judgment-intensive, lower-return processes, even where both administrations report numerically comparable overall adoption rates. This paper treats this parallel between its own RPA-specific findings and its companion research's broader diffusion-maturity argument as further evidence that the general pattern this paper's wider research program has identified, technology-adoption statistics systematically under-specifying the actual depth and character of realized organizational capability, extends specifically to the task-portfolio composition question this paper's RPA-focused analysis has newly identified.

#### ***4.4 What the Gradient Implies for Investment Prioritization***

Bringing this paper's evidence together, the task-complexity gradient this paper has documented across seven national cases carries a specific, actionable implication for how tax administrations should prioritize RPA investment going forward. This paper argues administrations should systematically map candidate automation processes against the structured-versus-judgment-intensive spectrum this paper's evidence illustrates, prioritizing implementation for processes most closely resembling Israel's or the United States' high-gain, rule-based profile before extending automation toward the more judgment-intensive processes in which this paper's evidence, consistent with this paper's companion research on algorithmic audit selection, indicates returns are likely to prove considerably more modest.

This paper's proposed prioritization approach carries a specific practical corollary worth stating explicitly: because RPA's technical logic requires an explicit, human-specified rule sequence



by design, the very process of attempting to specify an RPA automation for a given task can itself function as a diagnostic test of that task's underlying structure. A process that resists clean, stable rule specification, requiring frequent exception handling or extensive case-by-case judgment calls during the automation-design phase itself, is very plausibly a process falling toward the judgment-intensive end of this paper's spectrum, regardless of how repetitive or automatable the process may have appeared before implementation was actually attempted. This paper argues tax administrations should treat unexpected difficulty in RPA rule specification not merely as an implementation obstacle to be engineered around, but as substantive evidence about the underlying task's complexity, potentially warranting reconsideration of whether that specific process belongs earlier or later in the administration's overall automation-investment sequence.

## 5. Conclusion and Recommendations

This paper has examined robotic process automation's documented outcomes across seven national tax administration cases, drawing on OECD-compiled and independently cross-referenced evidence. The analysis finds a striking task-complexity gradient in reported gains: a 97 percent time reduction and 99.6 percent cost reduction for Israel's structured multi-employer flagging task, compared to a 5 percent reduction in manual actions for Australia's judgment-intensive audit process, with Canada's 34,000-hours-saved, the United States' 122-fold procurement speedup, and the Netherlands' four-hours-per-day invoicing gain occupying intermediate positions consistent with each task's relative structure and complexity. The OECD's finding that more than half of surveyed tax administrations were using or planning to use RPA as of 2022 indicates the technology has achieved broad diffusion across the global tax administration community.

This paper's evidence base carries specific limitations worth restating together. The seven cases this paper examines were selected by their respective reporting tax administrations as illustrative examples for inclusion in the OECD's comparative report, rather than drawn through systematic or randomized sampling, meaning this paper cannot rule out that administrations selectively reported their most successful RPA applications specifically, a form of survivorship or selection bias this paper's evidence cannot directly test or correct for. This paper's task-complexity classification, distinguishing structured from judgment-intensive tasks, is this paper's own interpretive framework applied to each case's brief published description, rather than a classification the source administrations themselves employed, and this paper's confidence in the resulting gradient should be calibrated accordingly. This paper's cross-case comparison also spans considerably different metrics, percentage reductions, absolute hours, and speedup ratios, that are not perfectly commensurable, limiting the precision of any single, unified comparative statistic this paper's evidence could support. Finally, this paper's evidence describes reported, self-submitted outcomes rather than independently, externally audited performance figures of the kind this paper's companion research on HMRC's Connect system was able to draw upon through Freedom of Information disclosure specifically.

Three recommendations follow from this analysis. First, tax administrations planning RPA investment should conduct explicit task-complexity assessment before implementation, using this paper's structured-versus-judgment-intensive framework or a comparable classification approach, to set realistic expectations for likely efficiency gains and to prioritize the highest-return candidate



processes first, treating unexpected rule-specification difficulty during implementation itself as diagnostic evidence of a given process's true underlying complexity. Second, tax administrations should track and publish RPA outcome metrics using consistent, comparable units, whether percentage reduction, absolute hours saved, or speedup ratios, to enable more rigorous cross-administration and cross-process comparison than the heterogeneous metrics this paper's own evidence base has required navigating. Third, researchers should pursue systematic, representative studies of RPA outcomes across a broader and more rigorously sampled set of tax administration processes than the illustrative case examples this paper's OECD-sourced evidence provides, ideally incorporating the same kind of randomized or quasi-experimental identification strategy this paper's companion research on algorithmic audit selection demonstrates is achievable in tax administration contexts specifically, and ideally supplementing self-submitted case reporting with independently audited performance verification of the kind Freedom of Information mechanisms have enabled for other tax-technology evidence this paper's broader research program examines.

The evidence reviewed in this paper indicates that robotic process automation has achieved broad diffusion across tax administrations globally and has produced genuinely substantial, well-documented efficiency gains in multiple national contexts. It also indicates, with reasonable consistency across seven independently reported cases, that these gains vary systematically with task complexity, largest for structured, rule-based processes and considerably more modest for processes retaining genuine professional judgment — a pattern this paper argues should inform how tax administrations sequence and prioritize their own automation investment, rather than treating RPA as a uniformly high-return technology to be deployed indiscriminately across all administrative and compliance functions alike.

## References

- Bachas, P., Brockmeyer, A., Ferreira, A., & Sarr, B. (2025). *Algorithms and bureaucrats: Evidence from tax audit selection in Senegal* (World Bank Policy Research Working Paper No. 11205). World Bank. <http://hdl.handle.net/10986/43682>
- Kim, H., Glaeser, E. L., Hillis, A., Kominers, S. D., & Luca, M. (2024). Decision authority and the returns to algorithms. *Strategic Management Journal*, 45(4), 619–648. <https://doi.org/10.1002/smj.3569>
- Organisation for Economic Co-operation and Development. (2022). *Tax administration 2022: Comparative information on OECD and other advanced and emerging economies*. OECD Publishing. <https://doi.org/10.1787/1e797131-en>
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). Free Press.
- Tornatzky, L. G., & Fleischer, M. (1990). *The processes of technological innovation*. Lexington Books.