

DUE PROCESS AND MASS ADJUDICATION: CRISIS AND REFORM*

72 STAN. L. REV. (forthcoming, 2019)

David Ames[‡]

Cassandra Handan-Nader[‡]

Daniel E. Ho[†]

David Marcus[‡]

Abstract

Goldberg v. Kelly and its progeny imposed a judicial model for decision-making on much of the administrative state. Procedural due process's lynchpin was accuracy: *Goldberg*'s premise was that agencies could improve the accuracy of their decision-making by giving individuals the sort of procedural rights enjoyed in court. In the wake of the due process revolution, federal agencies now adjudicate more cases than all Article III courts combined, and state adjudicators handle millions of cases with court-like procedures in their administrative systems. Yet despite *Goldberg*'s premise, mass adjudication has struggled to achieve an adequate threshold of accuracy. In much of the administrative state, this struggle has deepened into an urgent crisis. The leading academic response argues for a turn to "internal administrative law" and management techniques, not external law, to

* We thank Anne McDonough, Oluchi Mbonu, and Reid Whitaker for terrific research assistance, our many anonymous interview subjects across agencies for insightful conversations, and Michael Asimow, Jacob Goldin, David Hausman, Mark Krass, Daryl Levinson, Rob MacCoun, Anne Joseph O'Connell, Nick Parrillo, James Ridgway, Sam Sherman, Bill Simon, Angela Teuscher, Zac Townsend, and participants at the Munro Distinguished Lecture in Stanford's political science department, the faculty workshop at Northwestern Pritzker School of Law, the faculty workshop at SMU Dedman School of Law, the faculty workshop at Stanford Law School, and the C. Boyden Gray Center's conference on "The Veterans Appeals Process: A Case of Administrative Crisis and Possible Reforms" for helpful comments and conversations.

[‡] Director, Appellate Operations at Bergmann and Moore, LLC and Former-Chief of Office of Quality Review, Board of Veterans' Appeals, U.S. Department of Veterans Affairs.

[‡] Ph.D. student, Graduate Student Fellow, Regulation, Evaluation, and Governance Lab (RegLab), Stanford University.

[†] William Benjamin Scott and Luna M. Scott Professor of Law, Professor (by courtesy) of Political Science, Senior Fellow at Stanford Institute for Economic Policy Research, and Faculty Director of Regulation, Evaluation, and Governance Lab (RegLab), Stanford University.

[‡] Professor of Law, UCLA School of Law.

improve the quality of agency adjudication. Many agencies in turn have responded with such quality assurance programs, but we know next to nothing about how such programs have evolved, how they function, and whether they work.

Our Article is the first to rigorously investigate the promise and pitfalls of quality assurance as a guarantor of accuracy in agency adjudication. We make three contributions. First, we use in-depth interviews with senior agency officials and a wide array of internal agency materials to document the evolving use of quality assurance at three federal agencies whose mass adjudication epitomizes *Goldberg's* domain. This history documents years of fits-and-starts, as agencies tried to manage what is commonly referred to as a "quantity-quality" tradeoff. It also reveals deep tensions and ambiguities in what the agencies intend as the purpose of quality assurance.

Second, we provide the first rigorous test of quality assurance, the leading academic response to *Goldberg's* limitations. We use a rich dataset, never before available to outside academics, of over 500,000 cases decided by the Board of Veterans' Appeals (BVA) to craft a rigorous evaluation of a natural experiment of its "Quality Review" program. Cases were *randomly* selected for review of draft decisions by an elite squadron of attorneys to correct substantive legal errors. BVA used this program ostensibly to reduce appeals to and remands from the court reviewing its decisions. We show that the program failed on its own terms. Cases selected for Quality Review fared no better than cases that were not. The BVA used the program not to vindicate *Goldberg's* promise, but to mollify external oversight bodies, most notably Congress, with the appearance of accuracy.

Third, our historical and empirical evidence has substantial implications for major theoretical debates about "internal administrative law" and the emerging crisis in mass adjudication. We show that conventional scholarly accounts are in need of much refinement. Deficiencies in mass adjudication will not be fixed solely through external constitutional law, with courts imposing remedies from outside. Nor will they be fixed solely by internal administrative law. *Goldberg's* original premise of decisional accuracy requires a hybrid of external intervention, stakeholder oversight, and internal agency management. We offer concrete policy prescriptions, based on a pilot one of us designed as the BVA's Chief of the Office for Quality Review, for how quality assurance might be re-envisioned to solve the looming crisis of decisional accuracy.

TABLE OF CONTENTS

INTRODUCTION 1

I. THE LOOMING CRISIS OF DECISIONAL QUALITY 3

 A. *Three Pivotal Agencies* 4

 1. Social Security Disability Benefits Adjudication 5

 2. Immigration Adjudication 6

 3. Veterans’ Benefits Adjudication 7

 B. *The Caseload Crisis and the Quantity/Quality Tradeoff* 8

II. SYSTEMIC QUALITY, CONSTITUTIONAL OBLIGATION, AND INTERNAL ADMINISTRATIVE LAW 11

 A. *The Limits of External Law* 11

 1. Accuracy, Constitutional Obligation, and Remedial Mismatch 11

 2. Jurisdiction Channeling 13

 B. *Internal Administrative Law and Quality Assurance* 15

III. QUALITY ASSURANCE IN HISTORICAL PERSPECTIVE 17

 A. *SSA’s Decades of Experiments* 17

 1. Initial Efforts 18

 2. Bellmon Review 19

 3. Failed Efforts 21

 4. SSA’s Modern Quality Assurance System 22

 B. *EOIR’s Inattention* 24

 C. *BVA and Quality Review* 25

IV. EMPIRICAL CASE STUDY: BVA’S QUALITY REVIEW PROGRAM 29

 A. *A Critical Test Case* 29

 B. *Results and Analysis* 30

V. IMPLICATIONS 35

 A. *Theoretical Implications* 35

 1. Internal Administrative Law and the Threat from Outside 35

 2. Internal Pressures and Distortions 36

 B. *Improving Quality Assurance: Externally Influenced Internal Administrative Law* 42

 1. Improving Oversight 43

 2. Improving Incentives and Blunting Resistance 44

 3. A Role for Structural Reform Litigation 45

CONCLUSION 47

INTRODUCTION

Federal administrative adjudication is in distress. When a veteran applies for benefits and loses, more than five years will pass before his appeal gets decided.¹ Judges handling these appeals must decide 25-30 cases per week lest the backlog worsen. This pace is crushing. A judge often has no more than an hour to review thousands of pages in the record.² In 2017, about 100 agency attorneys issued a public “loss of confidence” statement, declaring that increased caseloads prevented them from reviewing appeals as closely as applicable law requires (*de novo* review).³ “I could have integrity here or I could stay employed,” one judge confided, describing the pressure to produce and the impact it has on the quality of his decision-making.⁴

Immigration judges likewise labor under immense backlogs, with caseloads quadrupling over the past fifteen years.⁵ The quality of decisions prompted scathing judicial critique, including Richard Posner’s takedown of “the Immigration Court” as “the least competent federal agency.”⁶ The social security disability benefits adjudication system similarly struggles, with long waits for hearings and rushed, poor decision-making excoriated by federal judges.⁷ In 2018, one federal judge lamented, “the Social Security system is broken.”⁸

An onslaught of cases threatens the capacity of these agencies – the Social Security Administration, the Veterans Administration, and the Department of Justice – to render accurate decisions. This crisis of decisional quality has major policy implications for the rights of immigrants to asylum, the rights of veterans to just compensation for their service, and disabled workers’ access to the social safety net. But its significance extends across the administrative state. *Goldberg v. Kelly* demands quasi-judicial procedures for a myriad of areas involving mass administrative decision-making.⁹ For any of these systems, whether they involve welfare benefits, access to Medicaid, or applications for patent rights – i.e., the vast terrain of quasi-adjudication in the administrative state – a crush of caseloads can undermine access to accurate decision-making.

The crisis of decisional quality in administrative adjudication seems to beg for a constitutional response. Under *Mathews v. Eldridge*’s famed approach to due process, an individual’s entitlement to enhanced procedural rights increases as a system of adjudication’s error

¹ 2017 U.S. DEP’T OF VET. AFFAIRS, BD. VET. APP. ANN. REP. 25.

² James D. Ridgway, Opening Remarks at the C. Boyden Gray Center for the Study of the Administrative State Conference: The Veterans Appeals Process: A Case of Administrative Crisis and Possible Reforms (Oct. 17, 2018) (noting in PowerPoint Slide 13 the number of electronic documents per appeal) (PowerPoint slides available with authors) [hereinafter “Ridgway Presentation”].

³ Letter from Douglas E. Massey, Counsel, Board of Veterans’ Appeals and President AFGE Local 17 to David Shulkin, Secretary of Veterans Affairs, Sep. 18, 2017 (including “Loss of Confidence Statement”).

⁴ Telephone Interview with Interview Subject #4, Transcript at 17 (Sept. 11, 2018). We interviewed current and former Board of Veterans’ Affairs personnel and others with inside knowledge of the veterans’ benefits adjudication system. We promised interview subjects anonymity, and thus throughout we use generic identifiers.

⁵ Nat’l Ass’n of Immigr. Judges, *NAIJ Has Grave Concerns Regarding Implementation of Quotas on Immigration Judge Performance Reviews* 3 (Oct. 18, 2017), https://www.naij-usa.org/images/uploads/publications/NAIJ_-_Concerns_Regarding_Implementation_of_Quotas_10-17-17.pdf.

⁶ *Chavarria-Reyes v. Lynch*, 845 F.3d 275, 280 (7th Cir. 2016) (Posner, J., dissenting).

⁷ E.g., *Wallace v. Colvin*, 193 F. Supp. 3d 939, 941 (N.D. Ill. 2016); *Batista v. Colvin*, Civ. No. 13-4185, 2014 U.S. Dist. LEXIS 80576, at *6 (E.D.N.Y. June 11, 2014); *Freismuth v. Astrue*, 920 F. Supp. 2d 943, 954 (E.D. Wis. 2013).

⁸ *Williams v. Berryhill*, Civ. No. 17-1203, 2018 WL 3569076, at *2 (D.S.C. July 25, 2018).

⁹ 397 U.S. 254 (1970).

rate rises.¹⁰ One might therefore expect procedural changes as decisional quality worsens. But the Supreme Court’s willingness to recognize new procedural rights for individuals caught in the Bleak House of agency adjudication quickly dissipated after *Mathews*.¹¹ Even if the Due Process Revolution of the 1970s had not long since crested,¹² there is reason to doubt the remedial efficacy of enhanced procedural rights.¹³

Another remedy may hold more promise. In 1974, the luminary Yale Professor Jerry Mashaw famously criticized *Goldberg* and the Due Process Revolution precisely for their incapacity to ensure systemic accuracy in agency adjudication. He suggested an alternative to individual procedural rights. Agencies ought to have an obligation to develop standards for decisional quality, to evaluate decision-making against these standards, and to develop responsive managerial interventions to improve adjudicator performance agency-wide. Mashaw made a “quality assurance system” a key plank in the “internal law of administration” he described and defended in *Bureaucratic Justice*, a book that remains the canonical account of justice in agency adjudication.¹⁴

In recent years, “internal administrative law” – the corpus of rules, evaluative strategies, and management techniques agencies use to govern themselves – has risen to the top of scholars’ research agendas. Despite this attention, and despite the dire policy need that crises of decisional quality pose, administrative law scholarship has ignored quality assurance initiatives for mass decision-making by agency adjudicators. Our Article fills this void. It is the first academic study in nearly forty years of the quality assurance initiatives that agencies have used to evaluate and improve decisional quality for systems of mass adjudication.

We make three chief contributions. First, we use material generated by Freedom of Information Act (“FOIA”) requests, in-depth interviews with senior agency officials, historical records, and other sources to provide a rich history and account of the various initiatives that the major systems of federal administrative adjudication have used to evaluate and try to improve their judges’ decisional quality. Promoted by institutions like the Administrative Conference of the United States and the U.S. Government Accountability Office, these initiatives are “internal administrative law” in that they can exist distinct from external legal oversight.

Second, we draw upon a rich internal database of over 500,000 decisions by Veterans’ Administration judges to evaluate a prototypical initiative of internal administrative law. The Board of Veterans’ Appeals (“BVA”) operated its “Quality Review Program” without formal change for fifteen years. It randomly selected cases for review, creating a unique natural experiment. The dataset, which the agency itself uses to manage its adjudicatory process, has never before been publicly studied. We use this data to examine whether the program, which an elite detail of 4-6 full-time attorneys implemented for over 15 years, improved decisional quality. Through in-depth investigation, we show that the program generated an all-but-meaningless measure of decisional quality. The program failed to identify errors in decision-making in any rigorous way, but it did generate an “accuracy rate” the agency used to defend its work. Quality review, in essence, became public relations. Our findings reveal the extent of the crisis threatening

¹⁰ 424 U.S. 319, 335 (1976).

¹¹ Thomas W. Merrill, *Jerry L. Mashaw, the Due Process Revolution, and the Limits of Judicial Power*, in ADMINISTRATIVE LAW FROM THE INSIDE OUT: ESSAYS ON THEMES IN THE WORK OF JERRY L. MASHAW 39, 43-44 (Nicholas R. Parrillo ed., 2017); Richard J. Pierce, Jr., *The Due Process Counterrevolution of the 1990s?*, 96 COLUM. L. REV. 1973, 1981-1983 (1996).

¹² Pierce, *supra* note 11, at 1989.

¹³ See *infra* pp. 18-21.

¹⁴ JERRY L. MASHAW, BUREAUCRATIC JUSTICE: MANAGING SOCIAL SECURITY 15 (1983); *id.* at 149.

the federal administrative state: even an agency that claims success may in fact be hiding dramatic declines in decisional quality.

Third, we derive broad theoretical and policy implications from these findings. The important example of quality assurance challenges leading accounts of how internal administrative law can best develop and function. These accounts argue that agencies should be given more room to govern themselves than reviewing courts tend to allow, faulting judicial review for its disruptive effect on internal administrative law. Left unarticulated are the pressures that can distort this law's evolution, ones that BVA's Quality Review Program and other experiments with quality assurance highlight. We argue that a more nuanced relationship between external and internal sources of administrative governance can counteract these pressures. This relationship would allow internal administrative law to develop in ameliorative ways but keep it from drifting in pathological directions. We derive concrete policy prescriptions for how internal and external sources of agency governance can productively interact. We explain how legislators and courts can prompt agencies to design and administer successful quality assurance initiatives. Even agencies laboring under huge caseloads can take steps to protect decisional quality and thereby meet their constitutional obligations.

Our Article proceeds as follows. In Part I, we introduce three agencies that are, in many ways, the most prominent examples of mass adjudication in the federal administrative state: VA's Board of Veterans' Appeals, the Social Security Administration's Office of Hearings Operations, and the Department of Justice's Executive Office for Immigration Review. Part II explains that, while the caseload crises faced by these agencies may be matters of constitutional significance, traditional constitutional and administrative law do not offer promising remedies. The leading alternatives are quality assurance initiatives – the internal training, guidance, monitoring, and continuous improvement programs agencies use to manage and improve accuracy. In Part III, we survey four mostly unexamined decades of agency experiences with such programs. We show that they occur in fits and starts, are heavily contested, and expose deep ambiguities as to the purposes of adjudication. In Part IV, we provide the results of our rich empirical study that leverages random case selection to examine the impact of BVA's Quality Review program. We show that on its own terms, the program failed and was co-opted to generate the *appearance* of effectiveness. Part V identifies the theoretical and policy implications of our findings, for internal administrative law and for agencies' efforts to meet an adequate threshold of decisional quality.

I. THE LOOMING CRISIS OF DECISIONAL QUALITY

Agencies handling large volumes of cases or claims have long weathered crises that have tested their capacities to render accurate decisions. In the early 1980s, for example, changes in disability policy and a vigorous effort to terminate benefits for those no longer disabled caused disability benefits dockets to explode and prompted a significant backlash against the Social Security Administration (“SSA”) in the federal courts.¹⁵ The struggles state adjudicators face to give welfare beneficiaries meaningful “fair hearings” are well-documented.¹⁶

But the present moment is unprecedented. Caseload crises simultaneously challenge the capacity of SSA, BVA, and the Executive Office of Immigration Review (“EOIR”) to decide cases

¹⁵ MARTHA DERTHICK, AGENCY UNDER STRESS: THE SOCIAL SECURITY ADMINISTRATION IN AMERICAN GOVERNMENT 5 (1990).

¹⁶ *E.g.*, VICKI LENS, POOR JUSTICE: HOW THE POOR FARE IN THE COURTS (2015).

adequately. No single reason explains why otherwise unrelated dockets have expanded so rapidly at approximately the same time. But the three agencies, which we introduce here, have responded in similar ways, including by asking their judges to increase the pace of decision-making. The quotas and case completion goals adjudicators face pose a systemic threat to decisional quality.

A. *Three Pivotal Agencies*

Hundreds of agencies across the United States adjudicate millions of cases, claims, applications, requests, petitions, and the like. We focus here on three of the most prominent sets of adjudicators in the federal administrative state: BVA’s veterans’ law judges (“VLJs”), SSA’s administrative law judges (“ALJs”), and EOIR’s immigration judges (“IJs”). Often bundled together for the analysis of administrative adjudication and its difficulties,¹⁷ these three agencies usefully represent the challenges that *Goldberg*-style mass adjudication faces for several reasons. First, these systems are unquestionably three of the most important adjudicatory agencies in the federal government. Together, their judges adjudicate more cases annually than do the nation’s federal district and court of appeals judges.¹⁸ SSA alone employs the vast majority of federal ALJs. Second, VLJs, IJs, and ALJs enjoy a quasi-judicial status that poses particular difficulties for quality assurance.¹⁹ Some agency adjudicators are civil servants subject to control by bureaucratic superiors. Supervisors can tell them how to decide matters, and the agency can discipline them based on the quality of their work.²⁰ ALJs, by contrast, enjoy guarantees of independence that strictly limit quality-based evaluation, performance management, and discipline.²¹ Although VLJs and IJs have less robust protections, their judicial roles insulate them from strict hierarchical control, and custom, if not formal regulation, limits decisional quality as a factor in performance evaluation.²² *Goldberg* requires a degree of judicial independence for adjudication.²³ But it can be a double-edged sword for quality assurance and create a significant management complexity.

¹⁷ See, e.g., ADMINISTRATIVE LAW FROM THE INSIDE OUT: ESSAYS ON THEMES IN THE WORK OF JERRY L. MASHAW 205, 244-45 (Nicholas R. Parrillo ed., 2017).

¹⁸ During the 12-month period ending September 30, 2017, the U.S. Courts of Appeals and the U.S. District Courts terminated 54,347 and 365,238 cases, respectively. *Judicial Caseload Indicators – Judicial Business 2017*, <https://www.uscourts.gov/statistics-reports/judicial-caseload-indicators-judicial-business-2017> (last visited Jan. 7, 2019). In FY 2017, BVA decided 52,661 appeals. 2017 Ann. Rep., *supra* note 1, at 30. IJs rendered 149,581 “initial case completions” in FY 2017. 2017 U.S. DEP’T OF JUSTICE, EXEC. OFFICE FOR IMMIGR. REV. STATISTICS Y.B. 10. In FY 2017, SSA ALJs completed 685,657 hearings. SOC. SEC. ADMIN, ANN. PERFORMANCE REP. FISCAL YEARS 2017-2019, 35.

¹⁹ See *infra* pp. 9-12.

²⁰ E.g., OFFICE OF THE INSPECTOR GEN., U.S. DEP’T OF COM., U.S. PATENT AND TRADEMARK OFFICE: U.S. PTO NEEDS TO STRENGTHEN PATENT QUALITY ASSURANCE PRACTICES 5 (2015) (describing quality as a factor in performance evaluation for patent examiners); *id.* at 8 (describing quality as a factor in examiner discipline).

²¹ Kent Barnett, *Against Administrative Judges*, 49 U.C. DAVIS L. REV. 1643, 1655-66 (2016).

²² Results generated by BVA’s Quality Review program, for instance, do not factor into VLJ performance evaluation. Telephone Interview with Interview Subject #3, Transcript at 10 (Aug. 31, 2018) (“[I]t always bugged me . . . that [quality review] didn’t seem to factor into” a VLJ’s performance evaluation). Before 2007 IJs were exempt from performance evaluation. *NAIJ Has Grave Concerns*, *supra* note 5, at 4. Presently, IJs may be subject to “performance appraisals” based, in part, on “poor decisional quality,” but these appraisals must “fully respect[] their roles as adjudicators.” 8 C.F.R. § 1003.0(b)(v) (2016).

²³ 397 U.S. 254, 271 (1970).

Third, VLJs, IJs, and ALJs have similar tasks that justify an inter-agency comparison of quality assurance initiatives.²⁴ They often review decisions made in non-judicial settings by non-attorney adjudicators;²⁵ they receive written evidence, develop cases in an inquisitorial manner, and can hold hearings;²⁶ they decide cases subject to review by both specialized and generalist appellate tribunals;²⁷ and each works for an agency that administers a single national policy through the one-by-one adjudication of huge numbers of cases.²⁸ To be sure, other agencies administer policy through ALJs or their equivalents.²⁹ For some, modest caseloads make the obligation to meet an adequate threshold of decisional quality an entirely different challenge.³⁰ Others have large dockets, but they handle cases with much lower stakes than what BVA, EOIR, and SSA adjudicate.³¹

1. Social Security Disability Benefits Adjudication

SSA administers two principal disability programs that in FY 2019 will pay \$208 billion in benefits to 18 million Americans.³² An applicant to the more significant of these programs must demonstrate that his physical and/or mental impairments render him unable to work. If successful, a typical claimant will receive about \$270,000 in lifetime benefits, plus Medicare coverage.³³

A claimant first files for disability benefits with a state-run office SSA supervises. A non-attorney “claims representative” decides whether the person meets the requirements for disability based on the written application and supporting records. After requesting reconsideration, an unsuccessful claimant can further appeal to an ALJ for a *de novo* review. The ALJ first “develops the record,” ensuring that the claimant’s file includes complete medical records and other supporting evidence, before holding a hearing. Afterwards, the ALJ prepares instructions and sends them to a “decision writer,” typically an attorney, to draft the decision. If the ALJ denies the claim, the claimant can appeal to SSA’s Appeals Council, an appellate body within the agency. Its administrative appeals judges, assisted by paralegal or attorney analysts, apply a “substantial evidence” standard of review to decide appeals. If the claimant loses again, she can seek judicial review in a federal district court.³⁴

²⁴ Paul Verkuil, *Meeting the Mashaw Test for Consistency in Administrative Adjudication*, in ADMINISTRATIVE LAW FROM THE INSIDE OUT, *supra* note 17, at 239, 244.

²⁵ *See infra* pp. 8-12.

²⁶ *Id.*

²⁷ *Id.*

²⁸ *Cf.* Jerry L. Mashaw, *Organizing Adjudication: Reflections on the Prospects for Artisans in the Age of Robots*, 39 UCLA L. REV. 1055, 1059 (1992) (describing this challenge for SSA).

²⁹ On ALJs and “administrative judges” generally, see Barnett, *supra* note 21, at 1652-62.

³⁰ ALJs working for the U.S. Securities and Exchange Commission rendered 114 “initial decisions” in 2018. *ALJ Initial Decisions: Administrative Law Judges*, U.S. SEC. & EXCH. COMM’N, <https://www.sec.gov/alj/aljdec.shtml> (last visited Jan. 7, 2019).

³¹ ALJs at the Office of Medicare Hearings and Appeals decide large numbers of cases, but the value of each is very modest. Jonah B. Gelbach & David Marcus, *Rethinking Judicial Review of High Volume Agency Adjudication*, 96 TEX. L. REV. 1097, 1104-05 & n. 33 (2018).

³² *See* FY 2019 CONGRESSIONAL JUSTIFICATION 3, <https://www.ssa.gov/budget/FY19Files/2019CJ.pdf> (last visited Jan. 7, 2019) (describing Disability Insurance and Supplemental Security Income payments).

³³ JONAH B. GELBACH & DAVID MARCUS, A STUDY OF SOCIAL SECURITY DISABILITY LITIGATION IN THE FEDERAL COURTS 47-48 (2016).

³⁴ For these details, see *id.* at 16-30.

SSA is the largest employer of federal ALJs, with about 1,600 working in about 170 hearing offices around the country.³⁵ Until very recently, ALJs secured jobs through a non-political process administered by the Office of Personnel Management. In 2018, however, the Supreme Court cast doubt on this system's constitutionality,³⁶ and in July 2018 President Trump asserted his power to appoint ALJs independent of the OPM process.³⁷ ALJs enjoy qualified decisional independence, as guaranteed by the Administrative Procedure Act ("APA"), and they are exempt from performance evaluation.³⁸ They can only be removed from office for cause,³⁹ through a protracted process that involves a right to be heard before another agency adjudicator and ultimately judicial review.⁴⁰

2. Immigration Adjudication

IJs work for EOIR in the Department of Justice. They decide several types of cases, all of which involve defensive efforts by immigrants to stay in the United States.⁴¹ An immigrant gets before an IJ one of two ways. The Department of Homeland Security ("DHS") can issue the immigrant a notice to appear if DHS believes the immigrant must leave. The immigrant then must attempt to convince the IJ that she has a right to remain. Alternatively, an asylum seeker can request an IJ hearing if an asylum officer denies the application. The IJ is supposed to help the immigrant develop the record before her hearing on the merits of her claim for relief. After the hearing, at which a DHS attorney represents the United States, the IJ prepares a decision, sometimes with staff attorney assistance. Either DHS or the immigrant can appeal, first to EOIR's Board of Immigration Appeals, its internal appellate body, and then to one of the regional courts of appeals.

The national ranks of IJs have expanded quickly. As of September 2018, more than 390 IJs work in 66 immigration courts around the country, up from 273 in June 2016.⁴² The Attorney General hires IJs, with only minimal eligibility criteria limiting his choices.⁴³ IJs are subject to performance appraisals every two years, and a rating of "less than satisfactory" can result in the IJ's removal.⁴⁴ An IJ can challenge a disciplinary action through an arbitration process established

³⁵ FY 2019 CONGRESSIONAL JUSTIFICATION, *supra* note 32, at 200; Hearing Office Locator, https://www.ssa.gov/appeals/ho_locator.html (last visited Jan. 7, 2019).

³⁶ *Lucia v. S.E.C.*, 138 S. Ct. 2044 (2018).

³⁷ Exec. Order No. 13,843, 83 Fed. Reg. 32755 (July 10, 2018).

³⁸ 5 U.S.C. §4301(2)(D).

³⁹ 5 U.S.C. § 7521(a).

⁴⁰ 5 U.S.C. §§ 7521(a), 7703(a)(1).

⁴¹ The five types of cases are deportation, removal, withholding-only, asylum-only, and exclusion. See Caseload Statistics Press Release, https://www.justice.gov/opa/press-release/file/1060936/download?utm_medium=email&utm_source=govdelivery (last visited Jan. 7, 2019).

⁴² *EOIR Announces Largest Ever Immigration Judge Investiture* (Sept. 28, 2018), <https://www.justice.gov/opa/pr/eoir-announces-largest-ever-immigration-judge-investiture>; *EOIR Swears in 15 Immigration Judges* (June 27, 2016), <https://www.justice.gov/eoir/pr/eoir-swears-15-immigration-judges>. For a list of immigration courts, see <https://www.justice.gov/eoir/eoir-immigration-court-listing> (last visited Jan. 7, 2019).

⁴³ For requirements, see <https://www.justice.gov/legal-careers/job/immigration-judge> (last visited Jan. 29, 2019).

⁴⁴ Labor Agreement Between the National Association of Immigration Judges and USDOJ, Executive Office for Immigration Review Art. 22.6.1, https://www.americanimmigrationcouncil.org/sites/default/files/foia_documents/immigration_judge_performance_metrics_foia_request_labor_agreement.pdf; *id.* Ar. 22.2.

in a collective bargaining agreement between the IJ's union and EOIR.⁴⁵ IJs' pay approximates that of ALJs and VLJs.⁴⁶

3. Veterans' Benefits Adjudication

The VA benefits system provides approximately \$90 billion per year to over 6.5 million veterans and their dependents.⁴⁷ Although claims for VA benefits involve education, home loans, burial expenses, and other matters, most seek compensation for injuries and diseases incurred during military service.⁴⁸ Whereas a typical social security claimant is either disabled or not disabled, a veteran's disability can range from 0% to 100%, in ten percentage point increments for different impairments, ranging from "hearing loss" to "posttraumatic stress disorder." Monthly disability payments to veterans with no dependents can vary from \$140 to \$8,750.⁴⁹

A veteran seeking benefits first files a claim with one of the Veterans' Benefits Administration's 56 regional offices, which together received 1.4 million claims in FY 2017.⁵⁰ Non-attorney staff help the veteran assemble necessary records and then make a decision based on the documentary record. If dissatisfied, the veteran can file a "notice of disagreement,"⁵¹ which obliges the regional office to take a second look and issue a "statement of the case."⁵² If the veteran continues to disagree with the decision, he can file an appeal to BVA in Washington D.C. for *de novo* review.⁵³ There, VLJs, working with staff attorneys, develop the record, ensuring that the veteran's claim includes all relevant documentary evidence.⁵⁴ Unless the veteran requests a hearing, a staff attorney will draft a decision before the VLJ considers a case's merits. The VLJ will review the record and the draft, then sign the decision once it is ready. A veteran can appeal

⁴⁵ *Id.* Art. 10.

⁴⁶ Compare Exec. Office for Immigr. Rev., *2018 Immigration Judge Pay Rates* (Jan. 7, 2018), <https://www.justice.gov/eoir/page/file/991386/download>, with <https://www.federalpay.org/employees/occupations/administrative-law-judge> (last visited Jan. 29, 2019).

⁴⁷ 3 U.S. DEP'T OF VET. AFFAIRS, CONG. SUBMISSION, FY 2019, at VBA-51-VBA-52.

⁴⁸ *Id.* at VBA-52; U.S. DEP'T OF VET. AFFAIRS, VET. BENEFITS ADMIN., ANNUAL BENEFITS REPORT FY 2016, 7-11.

⁴⁹ U.S. Dep't of Vet. Affairs, *2018 COLA Rates*, https://www.knowva.ebenefits.va.gov/system/templates/selfservice/va_ssnew/help/customer/locale/en-US/portal/55440000001018/content/55440000098322/2018-COLA-Rates (last visited Jan. 7, 2019).

⁵⁰ VA FY2019 CONG. SUBMISSION, *supra* note 47, at 53.

⁵¹ 38 U.S.C. § 7105.

⁵² See U.S. Dep't of Vet. Affairs, *File a VA Disability Appeal*, <https://www.va.gov/disability/file-an-appeal/> (last visited Jan. 7, 2019).

⁵³ 38 U.S.C. § 7104. The Veterans Appeals Improvement and Modernization Act of 2017 made significant changes to the VA appeals process, effective February 19, 2019. Pub. L. No. 115-55, 131 Stat. 1105 (2017); 84 Fed. Reg. 138 (Jan. 18, 2019). However, no substantive changes were made to BVA's standard of review and the workload remains significant. Under the new system, BVA projects the number of cases it receives will more than double in 2019 and 2020, as compared to 2018. See 2018 U.S. DEP'T OF VET. AFFAIRS, BD. VET. APP. ANN. REP. 24, 27.

⁵⁴ *E.g.*, *McGee v. Peake*, 511 F.3d 1352, 1357 (Fed. Cir. 2008).

an unfavorable decision to the Court of Appeals for Veterans Claims (“CAVC”), an Article I court,⁵⁵ and from there to the Federal Circuit.⁵⁶

BVA has grown rapidly, with its VLJ corps expanding from 56 in 2006 to 92 by 2018.⁵⁷ Founded in 1933, BVA predates the APA,⁵⁸ and thus VLJs are not formally ALJs. The VA Secretary appoints them with the President’s approval and based upon the BVA Chairman’s recommendations.⁵⁹ VLJs go through performance reviews every one to three years to obtain recertification.⁶⁰ In other respects, though, they resemble ALJs. They operate on the same pay scale as ALJs, they enjoy “for cause” removal protection, and they have a right to an agency hearing and ultimately judicial review to challenge any discipline.⁶¹

B. *The Caseload Crisis and the Quantity/Quality Tradeoff*

At the end of 1998, the average SSA ALJ had 326 hearings pending. That figure more than doubled, to 723, by the end of 2016.⁶² On average, a disability benefits claimant in 2012 waited 353 days for an ALJ to decide her case. In 2017, the average wait time rose to 605 days. Over 10,000 claimants died awaiting a hearing.⁶³

Caseloads for immigration judges have likewise exploded. From 1998 to 2018 the number of deportation cases pending in immigration courts grew from slightly over 100,000 to slightly over 800,000, while the number of IJs increased by only a third.⁶⁴ At their present pace, IJs would take 2.6 years to clear the existing backlog, even if the government initiated no cases in the interim.⁶⁵ In fact, filings have increased at staggering rates. In 2017, a record 295,294 new cases were filed in immigration courts; 2018 topped that figure by 10,685 cases.⁶⁶ Although times vary

⁵⁵ 38 U.S.C. § 7252. Congress created the CAVC and provided for judicial review of BVA decisions after Article III courts, prompted by *Goldberg v. Kelly* and other developments, increasingly disregarded statutory language purporting to limit their review authority. See H.R. REP. NO. 100-693, at 17-22 (1988). CAVC’s creation, in turn, prompted a judicialization of review at BVA. See James D. Ridgway, *The Veterans’ Judicial Review Act Twenty Years Later: Confronting the New Complexities of the Veterans Benefits System*, 66 N.Y.U. ANN. SURV. AM. L. 251, 272-276 (2010).

⁵⁶ 38 U.S.C. § 7292.

⁵⁷ 2006 U.S. DEP’T OF VET. AFFAIRS, BD. OF VET. APPEALS, REPORT OF THE CHAIRMAN 2; 2018 ANN. REP., *supra* note 1, at 36.

⁵⁸ 1992 U.S. DEP’T OF VET. AFFAIRS, BD. OF VET. APPEALS, REPORT OF THE CHAIRMAN 3.

⁵⁹ 38 U.S.C. § 7101A(a)(1).

⁶⁰ 38 U.S.C. § 7101A(c), (d).

⁶¹ 1995 U.S. DEP’T OF VET. AFFAIRS, BD. OF VET. APPEALS, REPORT OF THE CHAIRMAN 7.

⁶² 2017 SOC. SEC. ADMIN., ANNUAL STATISTICAL SUPPLEMENT, Tbl. 2.8; 2000 SOC. SEC. ADMIN., ANNUAL STATISTICAL SUPPLEMENT, Tbl. 2.F8.

⁶³ Jackie Walorski, *Reforms Are Needed to Fix Social Security Disability Insurance Backlog*, THE HILL, Dec. 18, 2017.

⁶⁴ TRAC Immigration, *Backlog of Pending Cases in Immigration Courts as of November 2018*, http://www.trac.syr.edu/phptools/immigration/court_backlog/apprep_backlog.php; *NAIJ Has Grave Concerns*, *supra* note 5, at 3.

⁶⁵ Miguel de Figueiredo et al., *Swift Injustice*, SLATE (Apr. 18, 2018), <https://slate.com/news-and-politics/2018/04/its-a-terrible-idea-to-push-immigration-judges-to-close-cases-more-quickly.html>.

⁶⁶ Exec. Office for Immigr. Rev., *Adjudication Statistics*, <https://www.justice.gov/eoir/page/file/1060841/download> (last visited Jan. 7, 2019).

considerably by facility, immigrants in many courts now wait more than 1,000 days for a hearing,⁶⁷ and some courts schedule hearings for up to six years out.⁶⁸

BVA's caseload has exploded over the past decade, doubling over the past three years alone.⁶⁹ At the end of FY 2018, BVA had 1,821 cases pending per VLJ.⁷⁰ If its caseload continues to grow as projected, each VLJ will have a docket of 4,384 cases by FY 2024, assuming no increase in staffing.⁷¹ In FY 2017, veterans waited more than five years from the day they filed their appeals for a BVA decision.⁷² Despite the recent legislation to reform the appeals system, the VA projects BVA's backlog will more than double between 2018 and 2024.⁷³ Thousands of veterans die while their appeals languish.⁷⁴ Because BVA remands half of its cases for further development, veterans commonly "find themselves trapped for years in a bureaucratic labyrinth, plagued by delays and inaction," as the Federal Circuit recently lamented. Recent backlogs have only worsened this "churn."⁷⁵

While the backlogs have some distinct causes across the agencies,⁷⁶ each agency has responded with similar measures. Each agency has added officials at rapid clips. Beyond that, each has ratcheted up expectations for how quickly judges decide cases. When it first set a case completion goal in 1975, SSA asked ALJs to decide 26 cases per month. In March 2017, SSA set this figure at 50 decisions per month.⁷⁷ The Attorney General in April 2018 threatened to issue subpar performance ratings to any IJ deciding fewer than 700 cases per year, or about 20 more than the average IJ decided from 2011-16.⁷⁸ BVA expects VLJs to decide 25-30 cases per week, a threshold that in 2018 helped the agency render 85,288 decisions, or 32,627 more than in 2017.⁷⁹ Before "the line judges were doing 750 [to] 800 [cases]," one official told us, but "it's now up to

⁶⁷ TRAC Immigration, *Despite Hiring, Immigration Court Backlog and Wait Times Climb* (May 15, 2017), <http://trac.syr.edu/immigration/reports/468/> (last visited Jan. 7, 2019).

⁶⁸ *NAIJ Has Grave Concerns*, *supra* note 5, at 3.

⁶⁹ Ridgway Presentation, *supra* note 2, at Slide 3.

⁷⁰ U.S. DEP'T OF VET. AFFAIRS, COMPREHENSIVE PLAN FOR PROCESSING LEGACY APPEALS AND IMPLEMENTING THE MODERNIZED APPEALS SYSTEM 20 (Aug. Update 2018) (listing number of VLJs); *id.* at 24 (describing Board inventory).

⁷¹ *Id.* at 20-24.

⁷² 2017 U.S. DEP'T OF VET. AFFAIRS, BD. VET. APP. ANN. REP. 25.

⁷³ COMPREHENSIVE PLAN, *supra* note 70, at 24.

⁷⁴ OIG, U.S. DEP'T OF VET. AFFAIRS, 16-01750-79, VETERANS BENEFITS ADMINISTRATION: REVIEW OF TIMELINESS OF THE APPEALS PROCESS iv-v (2018) (reporting that, during the first quarter of FY 2016, 1,600 veterans died while their appeals were pending).

⁷⁵ *Martin v. O'Rourke*, 891 F.3d 1338, 1349 (Fed. Cir. 2018) (Moore, J., concurring).

⁷⁶ On possible causes, see, e.g., T.J. Sutcliffe, *Social Security's Disability Hearings Backlog: A National Crisis*, 41 HEALTH & SOCIAL WORK 215, 216 (2016) (SSA); Lorelei Laird, *Legal Logjam*, ABA J., Apr. 2017, at 52, 55-56 (EOIR); Mark Duggan, *The Labor Market Effects of the VA's Disability Compensation System*, SIEPR POLICY BRIEF, Nov. 2014, at 2-3 (VA).

⁷⁷ 2017 Updated Compassionate And Responsive Service (CARES) and Anomaly Plan, at 14, https://www.ssa.gov/appeals/documents/2017_Updated_CARES_Anomaly_Plan.pdf (last visited Jan. 19, 2019); DONNA PRICE COFER, JUDGES, BUREAUCRATS, AND THE QUESTION OF INDEPENDENCE: A STUDY OF THE SOCIAL SECURITY ADMINISTRATION HEARING PROCESS 81 (1985).

⁷⁸ Nick Miroff, *Trump Administration, seeking to speed deportations, to impose quotas on immigration judges*, WASH. POST, Apr. 2, 2018.

⁷⁹ 2018 U.S. DEP'T OF VET. AFFAIRS, BD. VET. APPEALS ANN. REP. 8.

1,000 cases a year.”⁸⁰ Another official described the case completion goal as “far higher now than it was five years ago.”⁸¹

These caseload pressures pose a systemic threat to accuracy. Conventional wisdom holds that an uptick in the pace of decision-making degrades decisional quality.⁸² This quantity/quality tradeoff has a straightforward logic: the faster a decision-maker has to work, the more she is likely to err.⁸³

Many of those involved with high volume agency adjudication, going back decades, accept this conventional wisdom.⁸⁴ Seventy percent of the several hundred ALJs surveyed in 1978, for example, reported that production pressures caused decisional quality to suffer.⁸⁵ SSA studied the quantity-quality relationship in 2012 and concluded that, “[a]s ALJ production increases, the general trend for decisional quality is to go down.”⁸⁶ Among the 395 BVA attorneys surveyed in 2018, most agreed that their production standard “results in a decrease in the quality of decisions for veterans, and substantially increases the likelihood of errors”⁸⁷

* * *

Much rides on decisional quality. An erroneous denial of benefits can subject a veteran or disabled worker to extreme hardship while they slog through a frustratingly slow appeals process. “Delay, deny, hope they die” has become a slogan of sorts for cynical veterans stuck in the VA morass.⁸⁸ Errors coupled with long delay times can be a “death sentence” for social security disability benefits claimants, as a district judge recently noted.⁸⁹ It takes little to imagine the worst for an asylum applicant, facing persecution at home, who is wrongly denied relief.⁹⁰ Poor quality adjudication also has systemic consequences, for the federal budget, the federal courts, and for the just and accurate implementation of immigration policy.

⁸⁰ Telephone Interview with Interview Subject #2, Transcript at 4 (Aug. 29, 2018).

⁸¹ Interview Subject #4, *supra* note 4, at 10.

⁸² *E.g.*, Heckler v. Day, 467 U.S. 104, 111-115 (1984); Stuart L. Lustig, *Inside the Judges’ Chambers: Narrative Responses from the National Association of Immigration Judges Stress and Burnout Survey*, 23 GEO. IMMIG. L.J. 57, 65-66 (2008) (reporting perceptions about this tradeoff among IJs); THE LEWIN GROUP, INC. ET AL., EVALUATION OF SSA’S DISABILITY QUALITY ASSURANCE (QA) PROCESSES AND DEVELOPMENT OF QA OPTIONS THAT WILL SUPPORT LONG-TERM MANAGEMENT OF THE DISABILITY PROGRAM 7 (2001) (discussing this tradeoff for ALJs).

⁸³ *Association of Administrative Law Judges v. Colvin*, 777 F.3d 402, 404-405 (7th Cir. 2015).

⁸⁴ Agency officials on occasion have denied that an increased pace of decision-making comes at the expense of decisional accuracy. *E.g.*, Gerald K. Ray & Jeffrey S. Lubbers, *A Government Success Story: How Data Analysis by the Social Security Appeals Council (with a Push from the Administrative Conference of the United States) is Transforming Social Security Disability Adjudication*, 83 GEO. WASH. L. REV. 1575, 1606 (2015).

⁸⁵ STAFF OF H. SUBCOMMITTEE ON SOCIAL SECURITY, COMM. ON WAYS AND MEANS, 96TH CONG., ADMINISTRATIVE LAW JUDGES: SURVEY AND ISSUE PAPER 32 (Comm. Print 1979).

⁸⁶ STAFF OF H. COMM. ON OVERSIGHT AND GOVERNMENT REFORM, 113TH CONG., MISPLACED PRIORITIES: HOW THE SOCIAL SECURITY ADMINISTRATION SACRIFICED QUALITY FOR QUANTITY IN THE DISABILITY DETERMINATION PROCESS 43 (Comm. Print 2014).

⁸⁷ Letter from Douglas E. Massey, President, AFGE Local 17, to Chairman Mason, at 4 (April 26, 2018) (on file with authors).

⁸⁸ *E.g.*, WALTER C. PINKNEY SR., DELAY, DENY, HOPE THEY DIE: WINNING YOUR BATTLE WITH THE VA APPEAL PROCESS (2015); Craig M. Wax, *VA to Vets: ‘Delay, Deny, Wait til They Die’*, WASH. TIMES, Dec. 1, 2015.

⁸⁹ *Boatner v. Berryhill*, Civ. No. 16-243, 2018 WL 2191804, at *10 (S.D. Miss. May 11, 2018).

⁹⁰ *Cf.* Kevin Sieff, *When Death Awaits Deported Asylum Seekers*, WASH. POST, Dec. 26, 2018.

II. SYSTEMIC QUALITY, CONSTITUTIONAL OBLIGATION, AND INTERNAL ADMINISTRATIVE LAW

If caseload pressures have indeed placed high volume agency adjudication on a precipice, a plunge could have enormous policy and political implications. A plunge also will have legal ramifications, because systemic accuracy in high volume agency adjudication is a constitutional obligation. A failure to achieve an adequate threshold of decisional quality can trigger a due process challenge. One response to the looming crisis of decisional quality could thus take the form of an order, imposed by a federal judge, instructing the agency to change how it does business. Under presently existing doctrine, however, this sort of externally imposed redress is unlikely to prove availing.

If decisional quality risks becoming an “underenforced constitutional norm” in the federal courts,⁹¹ internal administrative law may fill the void. After examining the limits of external sources of agency governance, this Part introduces quality assurance as a key component in an internally administered effort to improve adjudicator performance.

A. *The Limits of External Law*

1. Accuracy, Constitutional Obligation, and Remedial Mismatch

An agency’s obligation to meet an adequate threshold of decisional quality comes out of the Due Process Revolution. In *Goldberg*, the revolution’s trigger,⁹² the Supreme Court declared that due process requires a welfare agency to afford beneficiaries certain procedural rights, including a hearing, before it could terminate benefits. This “pre-termination hearing has one function only,” the Court explained: “to produce an initial determination of the validity of the . . . grounds for discontinuance [and thereby] to protect a recipient against an *erroneous termination*”⁹³ The Court in *Mathews v. Eldridge* recast *Goldberg* as its famous balancing test for the adjudication of due process challenges to agency decision-making. Whether the agency offers constitutionally adequate procedures when it threatens liberty or property interests depends in important part on “the risk of an erroneous deprivation of such interest through the procedures used”⁹⁴ If the private interest in the enjoyment of the benefit at stake is strong enough, and if the procedures the agency offers to decide entitlements to that benefit err too frequently, then a court can order the agency to offer enhanced procedural rights.⁹⁵

Goldberg and *Mathews* have three germane legacies. First, they cast the quality of agency adjudication in terms of accuracy. This understanding of quality may be unduly narrow,⁹⁶ but it

⁹¹ Lawrence Gene Sager, *Fair Measure: The Legal Status of Underenforced Constitutional Norms*, 91 HARV. L. REV. 1212, 1226 (1978).

⁹² 397 U.S. 254 (1970); see also Jerry L. Mashaw, *The Supreme Court’s Due Process Calculus for Administrative Adjudication in Mathews v. Eldridge: Three Factors in Search of a Theory of Value*, 44 U. CHI. L. REV. 28, 29 (1976).

⁹³ 397 U.S. 254, 267 (1970) (emphasis added).

⁹⁴ 424 U.S. 319, 335 (1976).

⁹⁵ Cf. *Hamdi v. Rumsfeld*, 542 U.S. 507, 532-533 (2004) (applying the *Mathews* balancing test to require additional procedural protections).

⁹⁶ E.g., THE LEWIN GROUP, *supra* note 82, at iv-xiii.

continues to be the lynchpin of due process in agency adjudication.⁹⁷ Second, *Goldberg* and *Mathews* emphasize that a system's overall accuracy, not the accuracy of an individual decision, matters to the due process calculus. In most instances, a particular adjudicative process does not offend due process if it errs in one case or another.⁹⁸ The procedures courts can order as remedies for due process violations therefore have as their "sole goal" systemic accuracy, or accuracy for the category of decision-making at issue.⁹⁹ *Goldberg's* and *Mathews's* third legacy are the additional procedural protections that courts can order if a system errs too often. The logic is that these enhanced procedural rights will reduce errors in adjudication.

Critics of the Due Process Revolution have long questioned whether individual procedural rights can ensure that an agency meets its constitutional obligation to systemic accuracy. To the extent that such rights can improve the overall accuracy of an agency's decision-making, as a practical matter they do so only if unrelated individuals collectively exercise their rights in significant enough numbers and with some degree of unintended coordination. This happy result is unlikely.¹⁰⁰ Also, the population that ostensibly benefits from enhanced procedural rights may lack the resources or wherewithal to exercise them in large enough numbers.¹⁰¹ Appeals, to name one such procedure, can be "highly and mysteriously selective," and may depend on factors wholly unrelated to accuracy.¹⁰² IJs' decisions to deny continuances to immigrants seeking counsel and the government's decision to detain immigrants, for instance, both depress appeals.¹⁰³ Appellate dockets thus may not reflect actual error patterns, nor lead an agency to understand systematic sources of error.

Second, an individual's use of an accuracy-promoting procedure, without more, does nothing to promote systemic accuracy.¹⁰⁴ A single appeal may correct a single error, for example, but it does not oblige the agency or a reviewing court to determine whether the appeal represents

⁹⁷ Cf. Jerry L. Mashaw, *Conflict and Compromise Among Models of Administrative Justice*, 1981 DUKE L.J. 181, 188 ("From the perspective of bureaucratic rationality, administrative justice is accurate decision-making carried on through processes that take account of costs."); *id.* at 207 (identifying *Mathews* with a bureaucratic rationality model of administrative justice); THE LEWIN GROUP, *supra* note 82, at iv-xiii (linking accuracy to quality).

⁹⁸ *Mathews*, 424 U.S. at 344 (emphasis added); see also *Parham v. J.R.*, 442 U.S. 584, 615 (1979).

⁹⁹ Mashaw, *supra* note 92, at 48; see also Richard H. Fallon, Jr., *Some Confusions About Due Process, Judicial Review, and Constitutional Remedies*, 93 COLUM. L. REV. 309, 337 (1993); Robert L. Rabin, *Federal Regulation in Historical Perspective*, 38 STAN. L. REV. 1189, 1314 (1986); Gillian E. Metzger, *The Constitutional Duty to Supervise*, 124 YALE L.J. 1836, 1866 (2016).

¹⁰⁰ E.g., JOEL F. HANDLER, *THE CONDITIONS OF DISCRETION: AUTONOMY, COMMUNITY, BUREAUCRACY* 32-40 (1986); David Hausman, *The Failure of Immigration Appeals*, 164 U. PA. L. REV. 1177, 1214 (2016); Jason Parkin, *Adaptable Due Process*, 160 U. PA. L. REV. 1309, 1331 (2012); Kathleen G. Noonan et al., *Legal Accountability in the Service-Based Welfare State: Lessons from Child Welfare Reform*, 34 LAW & SOC. INQ. 523, 553 (2009); William H. Simon, *Legality, Bureaucracy, and Class in the Welfare System*, 92 YALE L.J. 1198, 1248 (1984); Jerry L. Mashaw, *The Management Side of Due Process: Some Theoretical and Litigation Notes on the Assurance of Accuracy, Fairness, and Timeliness in the Adjudication of Social Welfare Claims*, 59 CORNELL L. REV. 772, 775 (1974).

¹⁰¹ HANDLER, *CONDITIONS OF DISCRETION*, *supra* note 100, at 23; Mashaw, *supra* note 100, at 775; William H. Simon, *The Organizational Premises of Administrative Law*, 78 LAW & CONTEMP. PROB. 61, 82-83 (2015).

¹⁰² Mashaw, *supra* note 100, at 789. Even when litigants have imperfect information about errors, their selection of cases for appeal is likely to perform better than a random selection. But the social costs of appeals under such conditions are likely to be high, in part because poorly chosen appeals will make it more difficult for appellate courts to recognize error. Steven Shavell, *The Appeals Process as a Means of Error Correction*, 24 J. LEGAL STUD. 379, 413 (1995).

¹⁰³ BANKS MILLER ET AL., *IMMIGRATION JUDGES AND U.S. ASYLUM POLICY* 131-32 (2015); Hausman, *supra* note 100, at 1214.

¹⁰⁴ Mashaw, *supra* note 100, at 785; see also Simon, *supra* note 101, at 82.

“an isolated problem or the tip of an iceberg of similar but unappealed cases.”¹⁰⁵ A critical mass of appeals over time could, in principle, reveal entrenched pathologies in agency decision-making, even if the selection of cases for appellate review is imperfect.¹⁰⁶ But in practice, the agency must make some coordinated, managerial effort to gather and analyze data on appeals patterns and then act on what this information reveals. On its own, the mere provision of a right to appeal does not require or enable the agency to do so.

Finally, the provision of individual procedural rights may have systemic costs, including to systemic accuracy.¹⁰⁷ More errors and more appeals should produce more remands, and thus further increase caseload pressures impacting quality. A reviewing court, exposed to only a sliver of the agency’s caseload, may impose a burdensome requirement on the agency that increases costs and delay dramatically. Requiring VA to provide veterans seeking benefits with expert medical opinions, for instance, means a government-funded examination “probably better than every three seconds,” as one official, skeptical of the Federal Circuit’s supervision of VA decision-making, told us: “Right now, each time you breathe out, VA just produced an expert medical opinion on a claim.”¹⁰⁸ The result is that one claimant’s, immigrant’s, or veteran’s decision to avail himself of procedural protections to avoid or correct an error may lead to more errors in other cases, with questionable benefits and potentially significant costs for supposed due process beneficiaries.¹⁰⁹

To be sure, strengthened procedural rights for individuals to exercise may also improve systemic accuracy. A due process right to court-appointed and government-funded counsel, for example, would probably strengthen the overall quality of immigration adjudication.¹¹⁰ But, even if as-of-yet undeclared rights could improve systemic accuracy for different systems of adjudication, the likelihood that courts will find them lurking in the Due Process Clause is exceedingly low.¹¹¹ Claims seeking additional procedural protections for social welfare beneficiaries have uniformly failed in the Supreme Court this century,¹¹² and to date immigrants have failed to obtain additional rights under the most sympathetic of circumstances.¹¹³

2. Jurisdiction Channeling

¹⁰⁵ Mashaw, *supra* note 100, at 78

¹⁰⁶ Shavell, *supra* note 102, at 413.

¹⁰⁷ JERRY L. MASHAW, *DUE PROCESS IN THE ADMINISTRATIVE STATE* 34 (1985); William H. Simon, *The Rule of Law and the Two Realms of Welfare Administration*, 56 *BROOK. L. REV.* 777, 786 (1990).

¹⁰⁸ Interview Subject #2, *supra* note 80, at 9.

¹⁰⁹ *E.g.*, James D. Ridgway & David S. Ames, *Misunderstanding Chenery and the Problem of Reasons-Or-Bases Review*, 68 *SYRACUSE L. REV.* 303, 316-24 (2018) (discussing how “reasons-or-bases” CAVC remands rarely result in the award of benefits).

¹¹⁰ *Cf.* Ingrid V. Eagly & Steven Shafter, *A National Study of Access to Counsel in Immigration Court*, 164 *U. PA. L. REV.* 1, 49-50 (2015) (describing the impact of access to counsel for immigrants).

¹¹¹ ADRIAN VERMEULE, *LAW’S ABNEGATION: FROM LAW’S EMPIRE TO THE ADMINISTRATIVE STATE* 123 (2016).

¹¹² Merrill, *supra* note 11, at 44-45.

¹¹³ *CJLG v. Sessions*, 880 F.3d 1122 (9th Cir. 2018) (denying that Due Process requires government-provided counsel for indigent juveniles in removal proceedings).

Procedural due process is not the only external source of legal obligation concerned with decisional quality. The APA is another, of course,¹¹⁴ as are agency-specific statutes.¹¹⁵ In theory, nothing in the substance of these laws would prevent them from evolving to support claims plaintiffs can bring for remedies more systemic in reach than individual procedural rights.¹¹⁶ Such an evolution would fit neatly with the “duty to supervise” that Professor Gillian Metzger argues the Constitution imposes on federal agencies.¹¹⁷ But statutes governing disability benefits, immigration, and veterans’ benefits litigation mostly limit the federal courts to an appellate role concerned with the individual remediation of errors.¹¹⁸ These jurisdiction-channeling provisions would steer claims alleging systemic problems with decisional quality to fora ill-situated to receive them favorably.¹¹⁹

In theory, a claimant, immigrant, or veteran could challenge some systemic inadequacy in agency decision-making as part of an individual appeal or petition for review. But an appellant or petitioner will rarely pursue, much less win, these remedies. Presumably a disability benefits claimant who appeals to a federal district court does so because she believes SSA denied her benefits erroneously. The agency may have inadequately supervised ALJs in a particular hearing office, and perhaps this inadequate supervision made the error she alleges more likely. A systemic remedy, such as an order requiring better supervision, might indirectly benefit her. But she is more likely to benefit directly if she simply asks for the district court to correct the error in her individual case.

Legal and institutional hurdles stand in the way of an altruistic appellant or petitioner who might eschew a direct individualized remedy for a more indirect systemic one. Although its contours are contested, scope-of-remedy doctrine tends to limit a litigant to an injunction or injunction-like remedy no broader than what is necessary to resolve the harm the litigant herself has experienced.¹²⁰ A petitioner or appellant may proceed on behalf of a class of similarly situated people to broaden the remedy’s scope.¹²¹ But jurisdiction-channeling provisions steer litigants to appellate fora that do not typically handle a class action-type challenge of this sort. Rule 23 of the Federal Rules of Civil Procedure does not apply in the courts of appeals, and whatever other power they might have to certify classes of immigrant petitioners is uncertain and largely untested.¹²² The Federal Circuit recently recognized the CAVC’s power to certify classes of veterans.¹²³ But neither the CAVC nor the federal courts of appeals can supervise a discovery process of the sort necessary to establish that agency practices produce systemic inaccuracies; such evidence would

¹¹⁴ *E.g.*, 5 U.S.C. § 706(2)(A) (authorizing a court to set aside an “agency action” that is “arbitrary and capricious”).

¹¹⁵ *E.g.*, 42 U.S.C. § 405(g) (authorizing a court to set aside a disability benefits determination not supported by “substantial evidence”).

¹¹⁶ *But see* JERRY L. MASHAW, BUREAUCRATIC JUSTICE 187 (1983) (questioning whether anyone would have standing to bring a systemic challenge).

¹¹⁷ Metzger, *supra* note 99, at 1842.

¹¹⁸ 38 U.S.C. § 511(a); *id.* §§ 7252, 7292; 42 U.S.C. § 405(h); 8 U.S.C. §§ 1252(a)(5), (b)(9).

¹¹⁹ On jurisdiction channeling, see *Shalala v. Illinois Council on Long Term Care, Inc.*, 529 U.S. 1, 13 (2000). For a discussion of how a jurisdiction-channeling provision effectively prohibits a generic APA challenge to administrative adjudication, see *Martinez v. Napolitano*, 704 F.3d 620, 621-623 (9th Cir. 2012).

¹²⁰ *Trump v. Hawaii*, 138 S. Ct. 2392, 2424-29 (2018) (Thomas, J., concurring); DOUGLAS LAYCOCK, MODERN AMERICAN REMEDIES: CASES AND MATERIALS 246 (3d ed. 2002).

¹²¹ Samuel L. Bray, *Multiple Chancellors: Reforming the National Injunction*, 131 HARV. L. REV. 417 (2017).

¹²² E-mail from Adam Zimmerman, Professor of Law, Loyola Law School-Los Angeles, to David Marcus, Professor of Law, UCLA School of Law (Jan. 11, 2019, 14:55 PST) (on file with authors).

¹²³ *Ebanks v. Shulkin*, 877 F.3d 1037, 1040 (Fed. Cir. 2017); *Monk v. Shulkin*, 855 F.3d 1312 (Fed. Cir. 2017).

have to be compiled in the agency before the appeal.¹²⁴ District courts, the appellate tribunals for SSA appeals, do have these procedural powers. But legal restrictions may limit their review to evidence in the record,¹²⁵ and exhaustion requirements may limit class membership to those who have pursued all internal agency appeals.¹²⁶

To be sure, jurisdiction-channeling provisions do permit some systemic challenges to aspects of administrative adjudication pursued outside the designated appeals or petitions process. For a plaintiff to file such a case directly in a district court, the systemic challenge must involve matters that are “collateral” to the merits of individual claims.¹²⁷ A challenge to an SSA policy treating in-kind loans as a form of income, for instance, was collateral because it was “not essentially a claim for benefits.”¹²⁸ But the various tests courts use to determine when claims are indeed “collateral”¹²⁹ tend to exclude those alleging systemic inaccuracy in adjudicator decision-making, on grounds that these allegations necessarily require examination of cases’ merits.¹³⁰

The doctrine governing judicial review of administrative adjudication includes more nuance than we can discuss here, and we return to the possibility of systemic remedies imposed through structural reform litigation in Part V. For now, the point is simply that jurisdiction channeling reinforces the remedial limitations of procedural due process. It steers claims about decisional quality into venues that privilege individual error avoidance or correction over systemic reform. The result is a body of external law for administrative governance that presently offers little promise as a bulwark against a systemic decline in decisional quality.¹³¹

B. *Internal Administrative Law and Quality Assurance*

In 1974, Mashaw argued for a different due process doctrine, emphasizing its “management side.”¹³² Rather than give individuals procedures they can use to avoid or correct particular errors, due process should require agencies to use “systematic management techniques [to] discover errors, identify their causes, and implement corrective action.”¹³³ Mashaw quickly rethought a recommendation that courts enforce this version of Due Process with structural injunctions,¹³⁴ insisting in *Bureaucratic Justice* that such interventions would lie beyond judicial competence.¹³⁵ But Mashaw did not give up on a “right” to good administration. Rather, he turned

¹²⁴ *E.g.*, *Small v. Barnhart*, 329 F. Supp. 2d 1272 (N.D. Ala. 2004).

¹²⁵ *E.g.*, *Pronti v. Barnhart*, 339 F. Supp. 2d 480, 497 (W.D.N.Y. 2004)

¹²⁶ *Bowen v. City of New York*, 476 U.S. 467, 482-83 (1986).

¹²⁷ *Id.* at 483; *J.E.F.M. v. Lynch*, 837 F.3d 1026, 1032 (9th Cir. 2016).

¹²⁸ *Johnson v. Shalala*, 2 F.3d 918, 921 (9th Cir. 1993).

¹²⁹ *E.g.*, *Broudy v. Mather*, 460 F.3d 106, 114 (D.C. Cir. 2006); *Kildare v. Saenz*, 325 F.3d 1078, 1083 (9th Cir. 2003); *Veterans for Common Sense v. Shinseki*, 678 F.3d 1013, 1025 (9th Cir. 2012); *J.E.F.M.*, 837 F.3d at 1032.

¹³⁰ *Kildare*, 325 F.3d at 1083; *J.E.F.M.*, 837 F.3d at 1031; *Veterans for Common Sense*, 678 F.3d at 1025.

¹³¹ Metzger, *supra* note 99, at 1870 (“Systemic administrative functioning [has been] denied constitutional relevance, and no constitutional claim can be made simply because a government agency or institution is inadequately managed or supervised.”).

¹³² *See generally* Mashaw, *supra* note 100.

¹³³ *Id.* at 816.

¹³⁴ Jerry L. Mashaw, *How Much of What Quality? A Comment on Conscientious Procedural Design*, 65 CORNELL L. REV. 823, 835 (1980).

¹³⁵ MASHAW, *supra* note 116, at 226.

away from Due Process and reforms imposed from outside the agency and toward an “internal law of administration” as the better means for achieving improvement in agency performance.¹³⁶

In recent years, more and more administrative law scholars have come to agree with Mashaw’s view that externally devised and imposed law, such as procedural due process and the APA, only controls decision-making at the margins.¹³⁷ “Internal administrative law” – “internal rules and procedures, bureaucratic systems, and internal techniques of instruction, oversight, and control of agency personnel”¹³⁸ – matters far more to the day-to-day work of agency officials, judges, and employees.¹³⁹

Administrative law theorists have reinvigorated debates about internal administrative law as a promising vein to tap to “encourage consistency, predictability, and reasoned argument” in agency decision-making.¹⁴⁰ Deficits of institutional competence limit judicial capacity to devise doctrine that will succeed in ensuring that agencies balance all of the competing goals they have to serve.¹⁴¹ Judicial review and external oversight are too infrequent and too abstract to ensure that granular realities of day-to-day decision-making align with legal requirements.¹⁴² “Law’s abnegation” for the governance of agency adjudication reflects this reality, Professor Adrian Vermeule argues: “the federal judicial system is not set up, not equipped, to engage in a sustained course of synoptic institutional engineering.”¹⁴³ Agencies in contrast have the practical expertise to assess what they can achieve in light of institutional and budgetary constraints; the experience with political oversight to know what Congress might and might not allow, and what Congress does and does not want them to prioritize; and the best sense of what the universe of regulatory beneficiaries need from their decision-making, not just the few who seek judicial review.

Quality assurance initiatives were central to the internal administrative law Mashaw described for SSA in 1983 and epitomize it today.¹⁴⁴ Following recommendations Mashaw spearheaded for the Administrative Conference of the United States and subsequently touted by the now-Government Accountability Office, agencies have used quality assurance programs for decades. Indeed, SSA’s first effort in this regard responded to Mashaw’s argument for internal management strategies as a substitute for individual procedural rights.¹⁴⁵ These experiments offer a tantalizing basis to determine whether internal governance strategies can ensure decisional

¹³⁶ *Id.*; Merrill, *supra* note 11, at 53; *id.* at 57.

¹³⁷ VERMEULE, *supra* note 111, at 217.

¹³⁸ Jerry L. Mashaw, *The Inside Out Perspective: A First-Person Account*, in ADMINISTRATIVE LAW FROM THE INSIDE OUT: ESSAYS ON THEMES IN THE WORK OF JERRY L. MASHAW 501, 502 (Nicholas R. Parrillo ed., 2017); *see also* Gillian Metzger & Kevin M. Stack, *Internal Administrative Law*, 115 MICH. L. REV. 1239, 1243-1244 (2017); *id.* at 1253.

¹³⁹ *E.g.*, MASHAW, *supra* note 116, at 19.

¹⁴⁰ Christopher J. Walker, *Administrative Law Without Courts*, 65 UCLA L. REV. 1620, 1624 (2018) (“Internal administrative law . . . has become a hot subfield in administrative law.”); *see also* Metzger & Stack, *supra* note 138, at 1244; Kristen A. Collins, *Bureaucracy as the Border: Administrative Law and the Citizen Family*, 66 DUKE L.J. 1727, 1731-1732 (2017); Jennifer Nou, *Intra-Agency Coordination*, 129 HARV. L. REV. 421, 426-27 (2015); Elizabeth Magill, Foreword, *Agency Self-Regulation*, 77 GEO. WASH. L. REV. 859 (2009).

¹⁴¹ Metzger & Stack, *supra* note 138, at 1296; Merrill, *supra* note 11, at 57.

¹⁴² Metzger & Stack, *supra* note 138, at 1263-1264.

¹⁴³ VERMEULE, *supra* note 111, at 115.

¹⁴⁴ Metzger & Stack, *supra* note 138, at 1253.

¹⁴⁵ Deborah A. Chassman & Howard Rolston, *Social Security Disability Hearings: A Case Study in Quality Assurance and Due Process*, 65 CORNELL L. REV. 801, 801-803 (1980).

quality system-wide and thereby succeed where external law fails short.¹⁴⁶ They also provide an example of internal administrative law-in-action, a case study that can help test scholars' claims about internal administrative law's potential and limitations.

III. QUALITY ASSURANCE IN HISTORICAL PERSPECTIVE

We now provide an historical overview of the quality assurance initiatives the three federal agencies have used (or have failed to use, in EOIR's case), based on extensive materials secured through FOIA, in-depth interviews, and administrative records.¹⁴⁷ This untold story helps illuminate the factors that have influenced how agencies measure and try to improve decisional quality. The implications of this history come in Part V, but for the moment three observations merit mention. SSA's decades of fitful experiences with quality assurance illuminate how concerns not necessarily coterminous with accuracy in decision-making, including sensitivity to ALJ independence and overall priorities for policy implementation, determine a program's contours. Second, EOIR's near-total neglect of quality assurance suggests that an agency's commitment to accuracy may depend on the politics of error distribution. Third, BVA's refusal to upgrade its Quality Review program hints at a tension that agencies must resolve as they design their initiatives: a program can paint a rosy accuracy picture without improving accuracy.¹⁴⁸

A. SSA's Decades of Experiments

¹⁴⁶ A skeptic might challenge the usefulness of quality assurance initiatives for ALJ-type decision-making as a case study for internal administrative law. To Mashaw, hearing-based disability benefits adjudication resulted from a recreation of agency decision-making in judicial guise and imported into SSA a model of justice ill-suited to the consistent adjudication of a large number of claims efficiently. He proposed his "internal administrative law" as an alternative. It's no surprise that we have found pathologies in VLJ, ALJ, and IJ governance, one sympathetic to Mashaw's position might maintain, since these sorts of quasi-judicial figures are poorly suited to internal direction. But quasi-judicial decision-making for disability benefits, immigration, and veterans' benefits is here to stay. A conception of internal administrative law that declares ALJ-type decision-making inapposite risks irrelevance. It also leaves *Mathews v. Eldridge* and individual procedural rights as the sole legal response to crises of decisional quality, an outcome we reject as tantamount to surrender. Finally, BVA and SSA have tried to govern their judges with strategies that fit the internal administrative law mold, suggesting that internal administrative law-in-action has attempted to accommodate this decision-making.

¹⁴⁷ Agencies draw from a menu of techniques, including training, peer review, auditing, and internal appeals, among others, to manage the discretion they give frontline adjudicators. See Daniel E. Ho & Sam Sherman, *Managing Street-Level Arbitrariness: The Evidence Base for Public Sector Quality Improvement*, 13 ANN. REV. LAW & SOC. SCI. 251, 253-265 (2017). Agencies have deployed these techniques for programs we do not include within the rubric of quality assurance, even as they bear upon the accuracy of decision-making. The "quality control system" Mashaw had in mind included "the development of standards, the evaluation of performance against these standards, and action to upgrade substandard performance." Mashaw, *supra* note 100, at 791. We use these characteristics to identify initiatives that attempt to assess the systemic accuracy of agency decision-making and to improve decisional quality.

¹⁴⁸ While we discuss perceptions of the efficacy of various programs that agencies have used throughout this Part, to our knowledge no outside observer has ever subjected any of the initiatives to rigorous and reliable empirical examination, which we turn to in Part IV. While working at BVA, one of us did a superficial internal analysis of its quality review program in 2016, reaching similar conclusions about the overall effectiveness of the program to those we make here. This analysis has not been made public, although we quote documents obtained through FOIA that indicate its overall conclusions.

SSA has by far the most experience with quality assurance initiatives. Its use of them waxed and waned between the 1970s and mid-2000s. In recent years, the agency has institutionalized its commitment to these programs.

1. Initial Efforts

Disability benefits adjudication started in 1956. During the early years, the agency used “own motion review” to examine every allowance decision (a decision by an ALJ granting benefits) and many denials,¹⁴⁹ and it captured systematic data on aspects of ALJ decision-making.¹⁵⁰ But SSA lacked any sort of systematic initiative to assess the overall quality of ALJ decision-making and recommend improvements.¹⁵¹ This absence grew glaring as several legal and cultural changes fueled a huge increase in ALJ caseloads in the early 1970s.¹⁵² Despite worrying patterns of ALJ inconsistencies,¹⁵³ SSA jettisoned its modest quality efforts in 1975 to shift the staff tasked with the work to case production.¹⁵⁴ By 1975, an agency official lamented, “quality control” was “virtually nonexistent,” with “no feedback . . . to the ALJ corps in terms of improving the ultimate quality of the decision.”¹⁵⁵

Amidst this first major “crisis” for the disability benefits adjudication system,¹⁵⁶ the agency announced a management initiative that coupled new case completion goals with a three-pronged quality assurance program.¹⁵⁷ Designed to focus on ALJs with aberrational decision patterns, the first two prongs generated instructions tailored to specific ALJs for how they should hold hearings and write decisions.¹⁵⁸ The third, dubbed the “Quality Review System,” has become a prototype

¹⁴⁹ Charles H. Koch, Jr. & David A. Koplow, *The Fourth Bite at the Apple: A Study of the Operation and Utility of the Social Security Administration Appeals Council*, 17 FLA. ST. U. L. REV. 199 (1990).

¹⁵⁰ JERRY L. MASHAW, QUALITY ASSURANCE SYSTEMS IN THE ADJUDICATION OF CLAIMS OF ENTITLEMENT TO BENEFITS OR COMPENSATION 179-180 (1973). Before 1972 SSA ALJs were “hearing examiners.” See Ray & Lubbers, *supra* note 84, at 1583.

¹⁵¹ *Id.*

¹⁵² *Delays in Social Security Appeals: Hearings Before the Subcomm. on Social Security of the H. Comm. on Ways and Means*, 94th Cong. 17-18 (1975); DERTHICK, *supra* note 15 at 33-34; David E. Rosenbaum, *Huge Federal Disability Program Faces Inequities, Fund Woes, Suits*, N.Y. TIMES, July 27, 1977, at A1; Richard E. Levy, *Social Security Disability Determinations: Recommendations for Reform*, 1990 B.Y.U. L. REV. 461, 478.

¹⁵³ Robert G. Dixon, Jr., *The Welfare State and Mass Justice: A Warning from the Social Security Disability Program*, 1972 DUKE L.J. 681, 716.

¹⁵⁴ SURVEY AND ISSUE PAPER, *supra* note 85, at 25; STAFF OF H. COMM. ON WAYS AND MEANS, 93RD CONG., REP. ON THE DISABILITY INSURANCE PROGRAM 136 (1974); *Delays in Social Security Appeals*, *supra* note 152, at 75 (Statement of James B. Cardwell, Commissioner of Social Security).

¹⁵⁵ *Delays in Social Security Appeals*, *supra* note 152, at 538 (Statement of Robert L. Trachtenberg, Director, Bureau of Hearings and Appeals).

¹⁵⁶ Chassman & Rolston, *supra* note 145, at 807; see also *Delays in Social Security Appeals*, *supra* note 152, at 17-18; JERRY L. MASHAW ET AL., SOCIAL SECURITY HEARINGS AND APPEALS 116-120 (1978).

¹⁵⁷ Jeffrey S. Lubbers, *Federal Administrative Law Judges: A Focus on Our Invisible Judiciary*, 33 ADMIN. L. REV. 109, 125-126 (1981); Paul W. Parmele, *Preserving the Judicial Independence of Federal Administrative Law Judges: Are Existing Protections Sufficient?*, 4 J. LAW & POLITICS 207, 219 (1987); SURVEY AND ISSUE PAPER, *supra* note 85, at 18; *id.* at 28; COFER, *supra* note 77, at 84.

¹⁵⁸ SURVEY AND ISSUE PAPER, *supra* note 85, at 19-20; *id.* at 28; COFER, *supra* note 77, at 84-85; *id.* at 87; Nash v. Califano, 613 F.2d 10, 13 (2d Cir. 1980); Memorandum from Acting Associate Commissioner for Hearings and Appeals to All RCALJs et al. Regarding OHA’s Quality Assurance Program - INFORMATION 3-4 (July 1979) (on file with authors).

of sorts and resembles the Quality Review program we examine in-depth in Part IV.¹⁵⁹ Beginning in 1976, the agency selected 5% of all ALJ decisions for “post-effectuation” review (i.e., after the agency finalized the decision and notified the claimant).¹⁶⁰ Appeals Council analysts looked for errors in these decisions, defined as “any defect so serious that it required either corrective action by the Appeals Council, or would have required such action” had the case been appealed.¹⁶¹ A more searching standard of review, SSA feared, would rankle ALJs.¹⁶² A division within SSA used information about “serious” errors analysts observed, along with data culled from other sources, to prepare “quality assurance reports” documenting various trends that the agency distributed periodically to central and regional supervisory personnel.¹⁶³ The division also used the data to prepare instructional material on specific error areas.¹⁶⁴

Convinced that SSA cared mostly about case production, ALJs resisted this effort at quality assurance.¹⁶⁵ They sued to challenge the case completion goals and the prongs that targeted specific ALJs as unlawful interference with their judicial independence.¹⁶⁶ The lawsuit settled for an agreement by the agency not to “issue directives or memoranda which set any specific number of dispositions by ALJs as quotas or goals,” and to limit the feedback individual ALJs received from the Appeals Council on deficient decisions.¹⁶⁷ SSA officials concluded that the Quality Review System, which the lawsuit did not challenge, achieved little for decisional accuracy.¹⁶⁸ The initiative did not yield individualized feedback for ALJs, in part due to a concern for a “storm of protest” that might otherwise erupt among them.¹⁶⁹

2. Bellmon Review

SSA’s second major effort at quality assurance triggered deep and lasting bitterness among ALJs. Concerned by decisional inconsistencies and the high number of claims ALJs allowed, Congress in 1980 passed the “Bellmon Amendment” to require SSA to restart “own motion” review.¹⁷⁰ The duty of implementing Bellmon Review fell to the Reagan Administration, which was determined to cut social security costs.¹⁷¹ The program focused on high allowance ALJs, an emphasis justified by Congress’ supposed concern with excessive grants and a preliminary study indicating that high allowance ALJs err more than others.¹⁷² As eventually constructed, the sampling protocol selected 15% of ALJ allowance decisions for review, with most from ALJs who

¹⁵⁹ See generally *id.*

¹⁶⁰ Chassman & Rolston, *supra* note 145, at 813.

¹⁶¹ *Id.* at 811.

¹⁶² *Id.* at 811.

¹⁶³ *Id.* at 813; *id.* at 816.

¹⁶⁴ *Id.*

¹⁶⁵ SURVEY AND ISSUE PAPER, *supra* note 85, at 24; *id.* at 32.

¹⁶⁶ DERTHICK, *supra* note 15 at 41; COFER, *supra* note 77, at 111.

¹⁶⁷ Memorandum from Acting Associate Commissioner for Hearings and Appeals to All RCALJs et al. Regarding Resolution of Issues Raised in *Bono* Lawsuit – INFORMATION 1-2 (July 1979) (on file with authors).

¹⁶⁸ Chassman & Rolston, *supra* note 145, at 817-20.

¹⁶⁹ *Id.* at 816; see also Mashaw, *supra* note 134, at 823.

¹⁷⁰ Pub. L. No. 96-265, 94 Stat. 441, § 304(g) (1980); H.R. Rep. No. 96-944, at 57 (1980) (Conf. Rep.); Stieberger v. Heckler, 615 F. Supp. 1315, 1388 (S.D.N.Y. 1985).

¹⁷¹ DERTHICK, *supra* note 15 at 53.

¹⁷² S. SUBCOMM. ON OVERSIGHT OF GOVERNMENT MANAGEMENT OF THE COMM. ON GOVERNMENTAL AFFAIRS, 98TH CONG., REP. ON THE ROLE OF THE ADMINISTRATIVE LAW JUDGE IN THE TITLE II SOCIAL SECURITY DISABILITY INSURANCE PROGRAM 15 (Comm. Print 1983).

allowed claims most liberally.¹⁷³ An ALJ's decisions would remain subject to this targeted sampling until his or her allowance rate declined.¹⁷⁴ SSA claimed to use information gleaned from Bellmon Review to determine training needs and to identify areas where the agency needed to clarify disability policy.¹⁷⁵ It also proposed to create a feedback system to advise targeted ALJs of "decisional weaknesses"¹⁷⁶ and to put ALJs on improvement plans.¹⁷⁷

Other controversial changes to disability policy caused hearing requests to skyrocket in the early 1980s, creating another caseload crisis for the agency.¹⁷⁸ SSA thus coupled Bellmon Review with a new and much higher case completion goal, increasing each ALJ's target to 45 decisions per month in 1982.¹⁷⁹ The agency also began to take more muscular action to push ALJs to conform to its expectations for decision-making. The agency disclaimed any sort of performance evaluation system based on productivity or allowance rates.¹⁸⁰ But it exerted informal pressure by, for example, allocating support staff and office space based on "how the ALJ behaved"¹⁸¹

To some, Bellmon Review was not a quality assurance initiative per se, but rather a tool to achieve a lower national disability benefits tab.¹⁸² SSA's fixation on case production also crowded out concerns about decisional quality.¹⁸³ With most ALJs reporting pressure to disallow claims,¹⁸⁴ Bellmon Review prompted another lawsuit on behalf of ALJs complaining of lost decisional independence.¹⁸⁵ SSA gutted the program, then just starting, before the lawsuit went to trial.¹⁸⁶ The agency replaced its targeted sampling with a protocol that selected cases for review from all ALJ decisions, including denials,¹⁸⁷ and it used the information gleaned from the program to compile bland reports with little guidance to improve decisional quality.

¹⁷³ Memorandum from Assoc. Comm'r, Office of Hearings and Appeals, to All Administrative Law Judges (Sept. 24, 1982), reprinted in *Social Security Disability Reviews: The Human Costs: Joint Hearing Before the Subcomm. on Social Security of the H. Comm. on Ways and Means and the S. Special Comm. on Aging*, 98th Cong. 93, 94 (1982).
¹⁷⁴ *Id.*

¹⁷⁵ THE ROLE OF THE ADMINISTRATIVE LAW JUDGE, *supra* note 172, at 16.

¹⁷⁶ Memorandum, *Social Security Disability Reviews: The Human Costs*, *supra* note 173, at 95.

¹⁷⁷ *Id.* at 96.

¹⁷⁸ On these changes and their effects, see DERTHICK, *supra* note 15 at 33-37; Levy, *supra* note 152, at 479; THE ROLE OF THE ADMINISTRATIVE LAW JUDGE, *supra* note 172, at 2.

¹⁷⁹ *Id.* at 236; *id.* at 343. The average ALJ rendered 33 decisions in March 1981, then a record for the agency.

¹⁸⁰ THE ROLE OF THE ADMINISTRATIVE LAW JUDGE, *supra* note 172, at 16.

¹⁸¹ *Social Security Restoration Act of 1990: Hearing Before the Subcomm. on Social Security and Family Policy of the S. Comm. on Finance*, 101st Cong. 133, 136 (Prepared Statement of Eileen P. Sweeney, National Senior Citizen Law Center).

¹⁸² *Assoc. of Admin. Law Judges, Inc. v. Heckler*, 594 F. Supp. 1132, 1142 (D.D.C. 1984).

¹⁸³ *Cf.* Memorandum from Regional Chief Administrative Law Judge, OHA Regional Office, Region X, to All Administrative Law Judges, OHA, Region X, 344 (July 1, 1982), reprinted in *Social Security Disability Reviews: The Role of the Administrative Law Judge: Hearing Before the Subcomm. on Oversight of Government Management of the S. Comm. on Governmental Affairs*, 98th Cong. 343, 344 (1982) ("We should not be required to choose between [quality and quantity] but rather obtain both. However, it is stressed that quantity is necessary to survive.")

¹⁸⁴ *Social Security Disability Reviews: The Human Costs*, *supra* note 173, at 82 (Statement of Jerry Thomasson, Francis Mayhue, and David T. Hubbard, Administrative Law Judges).

¹⁸⁵ *Assoc. of Admin. Law Judges*, 594 F. Supp. at 1133.

¹⁸⁶ *Stieberger v. Heckler*, 615 F. Supp. 1315, 1378 n.45 (S.D.N.Y. 1985).

¹⁸⁷ *Id.* at 1379; Office of Hearings and Appeals, Report on Quality Review, June 1986, reprinted in *Administrative Law Judge Corps Act: Hearing Before the Subcomm. on Courts and Administrative Justice of the S. Comm. on the Judiciary*, 100th Cong. 73, 74 (1987).

Bellmon Review is now viewed as a colossal failure. Whatever its effect on the accuracy of ALJ decision-making,¹⁸⁸ the program cast a long shadow over quality assurance experiments going forward.¹⁸⁹ Its legacy includes, for instance, regulations amended in 1998 to disclaim SSA's power to select decisions by individual ALJs or particular hearing offices for pre-effectuation review.¹⁹⁰

3. Failed Efforts

SSA used stopgap quality assurance measures from the mid-1980s to 2007, trying but never quite implementing a more expansive program. The agency ended its quality assurance efforts altogether shortly after the Bellmon Review debacle.¹⁹¹ It began again in 1992, prompted in part by an explosive report documenting racial disparities in ALJ allowance rates.¹⁹² SSA rebooted the random sampling program it had started in 1976, first subjecting allowance and denial decisions for pre-effectuation review,¹⁹³ then selecting only allowance decisions.¹⁹⁴ The agency also began a peer review program.¹⁹⁵ Teams of ALJs scrutinized a random sample of post-effectuation decisions, including allowances and denials.¹⁹⁶ This initiative generated information “to address broad program issues,” such as reasons for high ALJ award rates; resulted in the issuance of several reports that indicated the accuracy of ALJ decision-making; and ostensibly informed ALJ training.¹⁹⁷ But the program never grappled with the racial bias issue that prompted the agency's renewed commitment to quality assurance.¹⁹⁸

SSA did not envision either of these programs as permanent responses to concerns about decisional quality. Several times in the 1990s and early 2000s it planned extensive changes to the

¹⁸⁸ *Stieberger*, 615 F. Supp. at 1379; *id.* at 1393-1394; *see also* U.S. GENERAL ACCOUNTING OFFICE, GAO/HEHS-97-102, SOCIAL SECURITY DISABILITY: SSA MUST HOLD ITSELF ACCOUNTABLE FOR CONTINUED IMPROVEMENT IN DECISION-MAKING 19 (1997).

¹⁸⁹ *E.g.*, *A Proposal to Restructure the Social Security Administration's Disability Determination Process: Hearing Before the Subcomm. on Social Security of the H. Comm. on Ways and Means*, 103rd Cong. 92, 92-93 (1993) (Statement of David P. Tennant, Treasurer, Association of Administrative Law Judges); THE LEWIN GROUP, *supra* note 82, at 24; Statement of Patricia Jonas, Executive Director, Office of Appellate Operations, Office of Disability Adjudication and Review, Social Security Administration, Before Permanent Subcomm. on Investigations of the S. Comm. on Homeland Security, Sept. 13, 2012, https://www.ssa.gov/legislation/testimony_091312.html (last visited Jan. 8, 2019). All congressional statements of SSA officials from 1995 to the present are available at this website.

¹⁹⁰ 20 C.F.R. §§ 404.969, 416.1469; *see also* 63 Fed. Reg. 36570 (July 7, 1998).

¹⁹¹ *Oversight of the Disability Hearings Process: Hearing Before the Subcomm. on Social Security of the H. Comm. on Ways and Means*, 105th Cong. 7, 11 (Statement of Carolyn W. Colvin, Deputy Commissioner, Social Security Administration) (noting that “[p]rior to 1993, there was no ongoing quality review of hearing decisions per se” and mentioning Appeals Council review of individual cases as the only “quasi-quality review component” in SSA's process).

¹⁹² GAO, *supra* note 188, at 20 n. 11.

¹⁹³ Colvin Statement, *supra* note 191, at 11.

¹⁹⁴ THE LEWIN GROUP, *supra* note 82, at 26-27; GAO, *supra* note 188, at 45 & n.b; Colvin Statement, *supra* note 191, at 11 (claiming that this program generated “valuable” management information “which has resulted in both ALJ training” and policy reformulation).

¹⁹⁵ GAO, *supra* note 188, at 20 n.11.

¹⁹⁶ THE LEWIN GROUP, *supra* note 82, at 27.

¹⁹⁷ GAO, *supra* note 188, at 20; *id.* at 64; *First in Series on Social Security Disability Programs' Challenges and Opportunities: Hearing Before the Subcomm. on Social Security of the H. Comm. on Ways and Means*, 107th Cong. 86 (2001).

¹⁹⁸ LINDA G. MILLS, A PENCHANT FOR PREJUDICE: UNRAVELING BIAS IN JUDICIAL DECISION MAKING 59 (1999).

disability adjudication process, and in each of these proposals it included revised quality review initiatives.¹⁹⁹ But almost nothing materialized.²⁰⁰ Inadequate funding and rising backlogs forced SSA to deprioritize these efforts.²⁰¹ In the early 2000s, SSA halted its random sampling initiative.²⁰² The peer review program limped on, albeit with no apparent impact after the late 1990s, until 2009, when the agency returned ALJs assigned to the program to the bench to hold hearings and decide cases.²⁰³

4. SSA's Modern Quality Assurance System

In 2006, SSA created an Office of Quality Performance as part of a planned large-scale reconstruction of the disability adjudication process.²⁰⁴ This effort ended in failure a year later, as rising backlogs and Congressional pressure forced the agency to shift its priorities.²⁰⁵ It adopted an aggressive backlog reduction plan that included, among other planks, an ALJ hiring spree, a new case completion goal of 500-700 decisions per ALJ annually,²⁰⁶ and a revived quality assurance program that remains in place today.²⁰⁷

The first aspect of this program, which a new Division of Quality began in 2010, is the latest iteration of pre-effectuation review, this time of a sample of allowance decisions by a team of Appeals Council analysts.²⁰⁸ As their predecessors did in the 1970s, staff attorneys gather structured data from each reviewed case on a large array of issues.²⁰⁹ These data help SSA track variances in decision patterns among the country's various regions and determine which sorts of

¹⁹⁹ Statement of Michael J. Astrue, Commissioner, Social Security Administration, Before the Subcomm. on Social Security of the H. Comm. on Ways and Means, June 27, 2012; U.S. GEN. ACCOUNTING OFFICE, GAO-02-322, SOCIAL SECURITY DISABILITY: DISAPPOINTING RESULTS FROM SSA'S EFFORTS TO IMPROVE THE DISABILITY CLAIMS PROCESS WARRANT IMMEDIATE ATTENTION 7 (2002).

²⁰⁰ GAO, *supra* note 199, at 25.

²⁰¹ Astrue Statement, *supra* note 199; Statement of Debra Bice, Chief Administrative Law Judge, Social Security Administration, Before the S. Comm. on Homeland Security and Governmental Affairs, Oct. 7, 2013 (blaming a lack of quality assurance on the fact that "the Appeals Council was not adequately funded to perform its oversight responsibilities").

²⁰² *Id.*

²⁰³ U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-18-37, SOCIAL SECURITY DISABILITY: ADDITIONAL MEASURES AND EVALUATION NEEDED TO ENHANCE ACCURACY AND CONSISTENCY OF HEARING DECISIONS 29 (2017).

²⁰⁴ Statement by Jo Anne B. Barnhart Before the Subcomm. on Social Security of the H. Comm. on Ways and Means, June 15, 2006; *see also* Frank S. Bloch et al., *The Social Security Administration's New Disability Adjudication Rules: A Significant and Promising Reform*, 92 CORNELL L. REV. 235 (2007).

²⁰⁵ Erik Eckholm, *Disability Cases Last Longer as Backlog Rises*, N.Y. TIMES, Dec. 10, 2007; Astrue Statement, *supra* note 199; U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-08-40, SOCIAL SECURITY DISABILITY: BETTER PLANNING, MANAGEMENT, AND EVALUATION COULD HELP ADDRESS BACKLOGS 43-44 (2007).

²⁰⁶ Statement of Frank A. Cristaudo, Chief Administrative Law Judge, Office of Disability Adjudication and Review, Social Security Administration, Before the Subcomm. on Social Security of the H. Comm. on Ways and Means, Sept. 16, 2008; *see also* U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-09-398, SOCIAL SECURITY DISABILITY: ADDITIONAL PERFORMANCE MEASURES AND BETTER COST ESTIMATES COULD HELP IMPROVE SSA'S EFFORTS TO ELIMINATE THE HEARINGS BACKLOG 47-50 (2009).

²⁰⁷ Bice Statement, *supra* note 201. *See generally* FELIX F. BAJANDAS & GERALD K. RAY, IMPLEMENTATION AND USE OF ELECTRONIC CASE MANAGEMENT SYSTEMS IN FEDERAL AGENCY ADJUDICATION 42-52 (2018).

²⁰⁸ Jonas Statement, *supra* note 189.

²⁰⁹ Office of Appellate Operations, Executive Director's Broadcast, Vol. 3, Jan. 13, 2012, *reprinted in Social Security Disability Programs: Improving the Quality of Benefit Award Decisions: Hearing Before the Permanent Subcomm. on Investigations of the S. Comm. on Homeland Security and Governmental Affairs*, 112th Cong. 883, 885 (2013).

issues generate the most errors.²¹⁰ As areas of concern materialize, the Division of Quality can conduct selective samples of decisions, using criteria that can help screen for likely errors.²¹¹

SSA uses this pre-effectuation review for several purposes. If it identifies an error in an ALJ's decision, it may remand that decision to the ALJ for correction.²¹² Information gleaned from pre-effectuation reviews also helps SSA identify error-prone issues that warrant additional training.²¹³ But the program does not enable SSA to track the performance of individual ALJs or hearing offices, out of concern for the sort of interference with decisional independence that stymied Bellmon Review.²¹⁴

The Division of Quality also began “focused reviews” in 2011.²¹⁵ This program is the first since the Bellmon Review to scrutinize specific ALJs. It uses an “internal early warning system” with a variety of indicators, including the ALJ's disposition rate and the rate at which the Appeals Council agrees with or remands her decisions, to select an ALJ whose decisions indicate a problematic pattern.²¹⁶ If the system gets triggered, the Division of Quality may assign a team of attorneys to review a sample of the ALJ's decisions.²¹⁷ This review determines if the ALJ's decisions indeed suffer from a recurring problem, and it helps supervisory personnel identify an appropriate intervention, such as individualized counseling and feedback.²¹⁸ From 2013 to 2016, SSA conducted seventy-two of these time-intensive focused reviews of ALJs,²¹⁹ including a number of ALJs with aberrational decision patterns.²²⁰ Focused reviews have also helped SSA identify issues that tend to create errors more generally and to generate training to address them.²²¹

SSA has developed two additional quality assurance programs since 2007. Its Office of Quality Review has conducted “disability case reviews” by randomly sampling all ALJ decisions and subjecting the selected ones for additional scrutiny.²²² These post-effectuation reviews, like the pre-effectuation reviews of allowance decisions only, generate information to guide ALJ training.²²³ Pursuant to the “inline quality review” program, begun in 2009, staff in SSA's ten regional offices randomly sample cases in hearing offices at various stages before effectuation.

²¹⁰ *Id.* at 885.

²¹¹ OIG, SOC. SEC. ADMIN., A-12-15-50015, PRE-EFFECTUATION REVIEWS OF FAVORABLE HEARING DECISIONS 6 n.19 (2017).

²¹² *Id.*

²¹³ *Id.*

²¹⁴ *Social Security Disability Programs: Improving the Quality of Benefit Award Decisions: Hearing Before the Permanent Subcomm. on Investigations of the S. Comm. on Homeland Security and Governmental Affairs*, 112th Cong. 18, 18-20 (2013) (Testimony of Debra Bice, Chief Administrative Law Judge); *see also* OIG, SOC. SEC. ADMIN., A-07-12-21234, THE SOCIAL SECURITY ADMINISTRATION'S REVIEW OF ADMINISTRATIVE LAW JUDGES' DECISIONS 8 (2012).

²¹⁵ Bice Statement, *supra* note 201.

²¹⁶ OIG, SOCIAL SECURITY ADMIN., A-12-16-50106, OVERSIGHT OF ADMINISTRATIVE LAW JUDGE DECISIONAL QUALITY 1 (2017); *id.* at D-2.

²¹⁷ GAO-18-37, SOCIAL SECURITY DISABILITY: ADDITIONAL MEASURES AND EVALUATION NEEDED TO ENHANCE ACCURACY AND CONSISTENCY OF HEARINGS DECISIONS 31 (2017).

²¹⁸ Ray & Lubbers, *supra* note 84, at 1597-1598.

²¹⁹ OIG, *supra* note 216, at D-2 n.8.

²²⁰ *Id.* at D-2; Executive Director's Broadcast, *supra* note 209.

²²¹ OIG, SOC. SEC. ADMIN., A-12-15-15005, THE SOCIAL SECURITY ADMINISTRATION'S EFFORTS TO ELIMINATE THE HEARINGS BACKLOG B-2-B-3 (2015).

²²² GAO, *supra* note 217, at 35.

²²³ *Id.* at 39.

These reviews examine the extent to which hearing office personnel, including staff and decision writers in addition to ALJs, prepare cases in a manner consistent with SSA policy.²²⁴

Recognizing the proliferation of quality assurance initiatives and the need to coordinate them, SSA created an Office of Analytics, Review, and Oversight in October 2017.²²⁵ As this institutionalization of quality assurance suggests, SSA's post-2007 efforts are the most enduring since the agency began its experiments in the late 1970s. Some signs suggest an impact on ALJ decision patterns,²²⁶ but an external evaluation of its quality effects has not been conducted.²²⁷

B. EOIR's Inattention

While SSA's experimentation with quality assurance initiatives has been extensive, EOIR's has remained superficial. The near-total disregard for quality assurance initiatives may reflect immigrants' political weakness and thereby shed as much light on the pressures that affect internal regimes of administrative governance as the decades of SSA experiments do.

The earliest program appears to be EOIR's "Immigration Court Evaluation Program."²²⁸ The agency sent a team of IJs, legal assistants, and court administrators to various immigration courts for a qualitative peer evaluation. This program ended in 2008, when EOIR replaced the information it yielded with a set of quantitative measures of court performance that focus exclusively on case completion goals.²²⁹ Performance measures adopted in 2018 for IJs focus almost entirely on case production and wholly neglect decisional accuracy.²³⁰ EOIR appears to rely upon Assistant Chief Immigration Judges, who supervise regions of IJs, to determine which

²²⁴ *Id.* at 35.

²²⁵ Testimony by Bea Disman, Acting Chief of Staff, Social Security Administration, House Committee on Ways and Means, Subcommittee on Social Security, Sept. 6, 2017; *see also* GAO, GAO, *supra* note 217, at 75-76.

²²⁶ Other developments, including a rapid expansion of the ALJ corps and significant changes to ALJ training, complicate an assessment of exactly how much credit the quality assurance initiatives deserve. GAO, *supra* note 217, at 25.

²²⁷ *Cf.* Verkuil, *supra* note 24, at 244-45 (describing the early stages of SSA's efforts).

²²⁸ In 2008, a GAO report identified IJ training and the one-by-one review of IJ decisions by the Board of Immigration Appeals as EOIR's only centrally administered programs with decisional quality as a primary concern. U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-08-935, U.S. ASYLUM SYSTEM: AGENCIES HAVE TAKEN ACTIONS TO HELP ENSURE QUALITY IN THE ASYLUM ADJUDICATION PROCESS, BUT CHALLENGES REMAIN 19 (2008). We submitted a FOIA request in January 2016 asking for "[a]ll studies, memoranda, reports, or other documents produced from January 1, 2010, to the present discussing, describing, analyzing, or proposing agency-wide quality assurance initiatives designed offered or designed to improve the accuracy and timeliness of immigration judge decision-making." Letter to Office of the General Counsel, EOIR, from David Marcus, at 2 (Jan. 28, 2016). We appealed when EOIR's production in response to this request indicated no more than what we summarize here but did not receive any additional information. We also attempted to interview EOIR personnel but were not permitted to do so. We are thus reasonably confident that our summary accurately reflects the history of EOIR quality assurance, although it is possible that EOIR materials from before January 1, 2010, that are not reflected in publicly available sources memorialize more extensive efforts.

²²⁹ U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-17-438, IMMIGRATION COURTS: ACTIONS NEEDED TO REDUCE CASE BACKLOG AND ADDRESS LONG-STANDING MANAGEMENT AND OPERATIONAL CHALLENGES 61 (2017). EOIR insists that a new "Organizational Results Unit" that supposedly started at the end of 2017, will operate as a successor to this program. *Id.* at 66.

²³⁰ Memorandum from James R. McHenry III, Director, Exec. Office for Immigr. Review, to The Office of the Chief Immigr. Judge et al. Appx. A (Jan. 17, 2018) (on file with authors).

ones need some intervention to improve performance.²³¹ It does not tell these supervisory judges how to evaluate IJ performance, however, and they appear to do so idiosyncratically.²³²

One episode illustrates just how little attention EOIR has paid to quality assurance. In 2002, the U.S. Attorney General implemented “streamlining” at the BIA, the internal appellate body that reviews IJ decisions.²³³ To eliminate BIA backlogs, among other objectives, the Attorney General authorized single BIA members, instead of panels, to decide appeals much more often, and with much less reasoning.²³⁴ Shoddy agency review prompted immigrants to flood the federal courts with appeals.²³⁵ Federal judges repeatedly lashed out at poor quality decision-making within EOIR, a backlash that pushed the Attorney General to respond.²³⁶ In August 2006 he ordered EOIR to make twenty-two changes to improve immigration adjudication. These included, among others, the adoption of “mechanisms to detect poor conduct and quality.”²³⁷

A few months later, EOIR’s Director issued a memorandum detailing the status of each required change. He noted that EOIR had “implemented” the quality assurance directive by issuing an instruction that the DOJ lawyers who defend immigration orders in federal court report instances when “an immigration judge failed to display the appropriate level of professionalism”²³⁸ Even in response to a crisis of legitimacy that threatened to unmake core aspects of the immigration court system,²³⁹ one prompted by disastrous quality decision-making in the immigration courts, EOIR declined to fashion a quality assurance program of the sort that SSA had used since the 1970s.

C. BVA and Quality Review

Along the spectrum of experience and innovation, BVA fits between SSA and EOIR. Its institutional commitment to quality assurance goes back decades,²⁴⁰ but its chief program, which took shape in 1998, is basic in design. The program has generated a blunt accuracy number that invariably casts BVA performance in a favorable light – a feature that perhaps explains why the program has proven resistant to change.

²³¹ U.S. GOV’T ACCOUNTABILITY OFFICE, GAO-08-940, U.S. ASYLUM SYSTEM: SIGNIFICANT VARIATION EXISTED IN ASYLUM OUTCOMES ACROSS IMMIGRATION COURTS AND JUDGES 9 (2008).

²³² *Id.*

²³³ *E.g.*, Katie R. Eyer, *Administrative Adjudication and the Rule of Law*, 60 ADMIN. L. REV. 647,673 (2008).

²³⁴ *Id.*

²³⁵ Bert I. Huang, *Lightened Scrutiny*, 124 HARV. L. REV. 1109, 1122-23 (2011).

²³⁶ *EOIR Revises Streamlining Fact Sheet*, 83 INTERPRETER RELEASES 512 (2006).

²³⁷ Memorandum from the Attorney General to the Deputy Attorney General et al. Regarding Measures to Improve the Immigration Courts and the Board of Immigration Appeals 3 (Aug. 9, 2006) (on file with authors).

²³⁸ Memorandum from Kevin D. Rooney, Director, EOIR, to All EOIR Employees Regarding the Attorney General’s Directives 3 (Mar. 2007) (on file with authors).

²³⁹ *See generally* MARGARET MIKYUNG LEE, CONG. RESEARCH SERV., RL33410, IMMIGRATION LITIGATION REFORM (2006) (discussing changes to the review of IJ decisions prompted in part by the streamlining experience).

²⁴⁰ Memorandum No. 01-96-19 from Office of the Chairman, Board of Veterans’ Appeals, Regarding Quality Review of Board of Veterans’ Appeals Decisions 1 (May 9, 1996) (on file with authors).

Prior to the CAVC's creation in 1988, the “splendid isolation”²⁴¹ of the veterans' benefits system made external measure of quality difficult.²⁴² The CAVC added significantly more scrutiny to BVA's work and pressed it to ensure that VLJs rendered policy compliant decisions. This additional scrutiny had an immediate and dramatic impact on BVA's substantive work, resulting in a significant drop in the denial rate and increase in the rate at which VLJs remanded cases.²⁴³ As BVA's Chairman acknowledged in 1996, the CAVC “has provided the most significant guidance in defining the essential ingredients of quality appellate decisions.”²⁴⁴

That year, BVA created a new quality system that “subjected all Board decisions to screening for errors before promulgation and approximately 10 to 15 percent of those decisions to a more intense quality review.”²⁴⁵ But it failed to keep statistics or data on the results of these reviews,²⁴⁶ and thus, by BVA's own admission, this system “did not lend itself to quantifiable measurement and repeatable comparisons over time.”²⁴⁷ BVA revised the system in 1998, to create a quantitative program to evaluate and score BVA's decision-making accuracy, and to collect data to identify areas needing improvement.²⁴⁸

The quality review program that took form in 1998 ran continuously until November 2016.²⁴⁹ It substantially resembled the “Quality Review System” SSA debuted in 1976. A computer system randomly selected 5% of all original and post-remand decisions from BVA for review, and 10% of cases which had previously been remanded by the CAVC.²⁵⁰ It did so after the VLJ had completed the draft decision, but before issuing it to the appellant. Prior to May 2002, staff who remained involved in deciding cases reviewed decisions selected by the quality review program. The GAO raised concerns about this structure in 2002, suggesting that BVA members

²⁴¹ *Brown v. Gardner*, 513 U.S. 115, 122 (1994) (quoting H.R. REP. NO. 100-963, pt. 1, at 10 (1988)); see also Robert L. Rabin, *Preclusion of Judicial Review in the Processing of Claims for Veterans' Benefits: A Preliminary Analysis*, 27 STAN. L. REV. 905, 905 (1975).

²⁴² For an overview of quality reporting in the years before the CAVC's creation, see Ridgway, *supra* note 55, at 255-56.

²⁴³ See Ridgway & Ames, *supra* note 109, at 325.

²⁴⁴ Chairman's Memorandum, *supra* note 240, at 2.

²⁴⁵ *Id.* at 1.

²⁴⁶ Richard C. Thrasher, Breakout Seminar – *The BVA: Who We Are and Why We're Here* handout, CAVC 7th Judicial Conference (Sept. 19-20, 2002).

²⁴⁷ 1998 U.S. DEP'T OF VET. AFFAIRS, REPORT OF THE CHAIRMAN 25.

²⁴⁸ See generally Memorandum No. 01-98-15 from Office of the Chairman, Board of Veterans' Appeals, Regarding Quality Review of Decisions of the Board of Veterans' Appeals (May 14, 1998) (on file with authors). The stated purpose of the quality review system was to (1) “measure performance in the area of quality for the Board as a whole and for each of the four Decision Teams”, and (2) “use decision accuracy data to identify areas of strength and weakness in decisions in order to set realistic goals for improvement, and develop training programs as needed.” Thrasher, *supra* note 246. From 1995 to 2012, VLJs were divided into four Decision Teams, each managed by one Deputy Vice Chairman and two Chief VLJs. These teams divided workload along geographical lines and each Decision Team took all appeals from specific VA regional offices. 1995 U.S. DEP'T OF VET. AFFAIRS, BD. OF VET. APPEALS, REPORT OF THE CHAIRMAN 9. In December 2012, the Decision Teams were abolished and the VLJs were reorganized into a single body headed by two Deputy Vice Chairmen and ten Chief VLJs. 2013 U.S. DEP'T OF VET. AFFAIRS, BD. VET. APP. ANN. REP. 3.

²⁴⁹ E-mail from Carol DiBattiste to BVA List (Nov. 16, 2016 4:56 EST) (on file with authors).

²⁵⁰ CYNTHIA BASCETTA, U.S. GEN. ACCOUNTING OFFICE, GAO-05-655T, VA DISABILITY BENEFITS: BOARD OF VETERANS' APPEALS HAS MADE IMPROVEMENTS IN QUALITY ASSURANCE, BUT CHALLENGES REMAIN FOR VA IN ASSURING CONSISTENCY 7 (2005). Prior to May 2002, 100% of cases which had previously been remanded by the CAVC were selected for QR review. See “Quality Review and VACOLS Changes” email from Richard C. Thrasher, Chief Counsel for Policy (May 15, 2002).

directly involved in deciding veterans' appeals should not also review draft decisions for accuracy.²⁵¹ BVA revised the program in May 2002, concentrating all quality review work in its Appellate Group,²⁵² a managerial body not typically involved in direct adjudication.²⁵³ Within the group, the Office of Quality Review,²⁵⁴ a team of 4-6 competitively selected attorneys serving two year details, administered the program.²⁵⁵

Between 1998 and 2002, a case was considered erroneous, or inaccurate, if it contained at least one error from a list of pre-defined categories, internally known as "exceptions."²⁵⁶ As a result of the 2002 GAO report, BVA narrowed the definition of an error to "a deficiency that would be outcome determinative, that is, result in the reversal or remand of a Board decision by the CAVC."²⁵⁷ The error standard explicitly omitted any situation in which there could be "legitimate differences of opinion as to the outcome in an appeal, the interpretation of the law, the application of the law to the facts, or assessment of the weight and credibility of the evidence."²⁵⁸

Between 2003 and 2016, BVA's accuracy rate never dipped below 89%, even as the rate at which the CAVC remanded or reversed BVA decisions likewise never fell below 40%.²⁵⁹ This apparent disparity prompted skepticism about the accuracy rate's integrity.²⁶⁰ In 2016, one of us, who headed the Quality Review program at the time, proposed replacing the existing system with a data-driven system to identify recurring problems and generate interventions to address these problems.²⁶¹ As the proposal indicated, "the accuracy rate number . . . is essentially static, does not reflect any long-term trends or changes in quality, and is completely inconsistent with results reported by the CAVC."²⁶² BVA experimented with this alternative for just enough time to generate an in-depth study of a particular regulation that had proven difficult for VLJs to apply

²⁵¹ U.S. GEN. ACCOUNTING OFFICE, GAO-02-806, VETERANS' BENEFITS: QUALITY ASSURANCE FOR DISABILITY CLAIMS AND APPEALS PROCESSING CAN BE FURTHER IMPROVED 7 (2002).

²⁵² BVA's Appellate Group was renamed the Office of the Principal Deputy Vice Chairman in FY 2017. 2017 U.S. DEP'T OF VET. AFFAIRS, BD. VET. APP. ANN. REP. 8.

²⁵³ GAO, *supra* note 251, at 7; 2002 U.S. DEP'T OF VET. AFFAIRS, BD. OF VET. APPEALS, REPORT OF THE CHAIRMAN 7.

²⁵⁴ At various times, the Office of Quality Review was also referred to as the Quality Review Unit and the Office of Quality Assurance. For consistency, this article will only refer to it as the Office of Quality Review.

²⁵⁵ 2002 Chairman's Report, *supra* note 253, at 7.

²⁵⁶ *Id.* For a list of these error categories, see GAO, *supra* note 251, at 9. In August 2015, after overhauling its coding methodology, BVA reduced the six categories to five: (1) issues, (2) remands, (3) reasons or bases, (4) due process, and (5) miscellaneous. U.S. Dep't of Vet. Affairs, *Veterans Appeals Control and Locator System (VACOLS) 3 Data Dictionary* 49 (Jan. 26, 2018) (on file with authors).

²⁵⁷ 2002 U.S. DEP'T OF VET. AFFAIRS, BD. OF VET. APPEALS, REPORT OF THE CHAIRMAN 7; BASCETTA, *supra* note 250, at 8. This change eliminated all errors in the "Format" category.

²⁵⁸ Chairman's Memorandum, *supra* note 248, at 3.

²⁵⁹ BVA reports its accuracy rates in annual reports, and the CAVC reports remand and reversal rates in its annual reports.

²⁶⁰ See Memorandum from Vice Chairman, Board of Veterans' Appeals, to Deputy General Counsel, Regarding Monthly Performance Review Submission on the Board of Veterans' Appeals' Accuracy Rate 1 (Aug. 3, 2010) (describing VA's Office of General Counsel skepticism of the accuracy rate in light of a high CAVC remand rate); *Board of Veterans' Appeals Adjudication Process and the Appeals Management Center: Hearing Before the Subcomm. on Disability Assistance and Memorial Affairs of the H. Comm. on Veterans' Affairs*, 110th Cong. 1 (2007) (Statement of Rep. Hall) ("[A]lthough the BVA claims a 93 percent accuracy rate, the U.S. Court of Appeals for Veterans Claims sets aside or remands over 70 percent of the cases appealed indicating a much lower accuracy rate in reality.").

²⁶¹ Office of Quality Assurance, *QA Metrics Proposal* (on file with authors).

²⁶² *Id.*

properly.²⁶³ Subsequent training based on this study seems to have succeeded at changing VLJ behavior in this area of law.²⁶⁴

In late 2017, however, BVA reinstated a version of its old quality review program. As one official told us, a single erroneous VLJ decision had embarrassed VA, and in response BVA's Interim Principal Deputy Vice Chairman²⁶⁵ insisted that it restart the 2002 version of the quality review program.²⁶⁶ BVA's reconstituted program samples a smaller percentage of draft decisions; skyrocketing caseloads make sampling of a full 5% impossible absent an increase of resources to quality review.²⁶⁷ Previously attorneys assigned to the program served two-year details, as an attorney "took 6-12 months to gain expertise."²⁶⁸ The reinstated program involved attorneys on six-month details, with the "goal to achieve full productivity within one week of training."²⁶⁹

Importantly, the reconstituted program uses a new standard of review that abandons the pretense that reviewing attorneys predict CAVC remands. Instead, the new standard requires that errors be "clear and unmistakable" – ones where "the result would have been manifestly different but for the error."²⁷⁰ Such instances should be "very, very rare."²⁷¹ Using this new approach, BVA reported a 93.6% accuracy rate for FY 2018. During these twelve months, a remarkable 47% increase in staff productivity enabled BVA to decide 85,288 cases, crushing the record it had set in FY 2017.²⁷²

* * *

We can draw several lessons from this untold story of quality management in the administrative state. First, the variability in agency attention to quality is substantial. Despite the constitutional demand for accuracy and evidence of considerable challenges at each of these agencies, the variation is striking. As SSA's Office of Analytics, Review, and Oversight indicates, quality assurance has become institutionalized at SSA after decades of fitful experiments. To this day, BVA continues to use a model rooted in the 1970s, notwithstanding an internal push for more data-driven techniques. EOIR equates unprofessional conduct by IJs with low quality, suggesting a hollowing out of *Goldberg's* demand for accuracy within the agency. These varying quality efforts may be indicative of differences in motivation. SSA's numerous quality program changes and experimentations suggest an honest search for accurate quality measurement and improvement by at least some portion of the agency. BVA's utilization of the same program, without any notable

²⁶³ U.S. DEP'T OF VET. AFFAIRS, TQA 17-1, TARGETED QUALITY ANALYSIS: APPEALS OF EXTRASCHEDULAR RATINGS UNDER 38 C.F.R. § 3.321(B)(1) (2017) (on file with authors).

²⁶⁴ U.S. Dep't of Vet. Affairs, *Navigating § 3.321(b)(1) Extraschedular Analysis* (Aug. 2017) (PowerPoint presentation, on file with authors).

²⁶⁵ The Interim Principal Deputy Vice Chairman was subsequently nominated by the President as the Chairman in September 2017 and appointed in December 2017.

²⁶⁶ Telephone Interview with Interview Subject #7, Transcript at 14 (Sept. 28, 2018). There is no obvious connection between the old system's jettisoning and the erroneous VLJ decision that prompted this backtrack. The only conceivable reason why BVA responded the way it did was to be seen as appearing to do something about quality.

²⁶⁷ 2018 U.S. DEP'T OF VET. AFFAIRS, BD. VET. APPEALS ANN. REP. 11.

²⁶⁸ U.S. Dep't of Vet. Affairs, *Case Review Modernization 4* (Sept. 2017) (PowerPoint presentation, on file with authors).

²⁶⁹ *Id.* at 7.

²⁷⁰ *QR Error Standards – Summary Only* (Undated Memorandum) (on file with authors).

²⁷¹ *Case Review Modernization*, *supra* note 268, at 9.

²⁷² 2018 U.S. DEP'T OF VET. AFFAIRS, BD. VET. APP. ANN. REP. 8-9; *id.* at 11; *see also* Ridgway Presentation, *supra* note 2, at 8 (charting BVA productivity 1997-2018).

changes, for over 15 years, suggests that the rosy statistics the program produces are the true goal. EOIR's near total lack of quality monitoring efforts reflects an agency unconcerned with accountability for poor quality.

Second, quality remains a deeply contested concept. In the 1980s, SSA deployed it for political control to reduce the welfare rolls, and the quality assurance initiatives continue to challenge core notions of decisional independence. BVA's reforms illustrate the challenges of internal institutional design around quality, with GAO critiquing the use of staff attorneys to review decisions by superiors. As we show in more depth in Part V, bureaucratic incentives can distort how an agency measures and reports quality. While few would contest that agencies should strive for high-quality decisions at an abstract level, the implementation is anything but straightforward.

Third, while there has been much experimentation, we know virtually nothing about which initiatives actually work. This is unfortunate as it undercuts the ability of by all federal agencies to learn about how to design effective adjudicatory systems. SSA insiders suggest significant improvement at SSA's Appeals Council, but the effectiveness of interventions for ALJ decision-making remains uncertain.²⁷³ No quality assurance program has even been subjected to an independent evaluation. Yet if institutional and doctrinal limitations hobble the capacity of external law to ensure an adequate threshold of decisional quality in mass adjudication, and if internal administrative law offers the best alternative, it is critical to understand whether these initiatives are as effective as billed.

IV. EMPIRICAL CASE STUDY: BVA'S QUALITY REVIEW PROGRAM

Because one of us previously served as BVA's Director of the Office of Quality Review ("QR"), we are able to offer, for the first time, a rigorous assessment of whether this form of internal administrative law works. In this Part, we first spell out what makes our case study unique and then highlight the main findings. Our takeaway is straightforward: BVA's QR program failed. Cases that went through quality review were no less likely to be appealed and remanded than cases that did not. The likely reason for this stunning finding involves the standard of review QR attorneys used when they evaluated VLJ decisions. While ostensibly predicting what would happen to cases if appealed to the CAVC, these attorneys in reality were a good deal more deferential to VLJs. The QR program generated rosy numbers suggesting highly accurate decision-making that bore little relationship to any external measure of quality. It gave VLJs virtually no incentive to improve.

A. *A Critical Test Case*

Our case study is unique in several respects. First, as spelled out in Part III, BVA's QR program remains the leading prototype of how to assess and improve the accuracy of decision making through an internal agency program. Second, our examination is the first independent test of such a program.²⁷⁴ Third, in contrast the previous studies that are limited to reporting highly

²⁷³ BAJANDAS & RAY, *supra* note 207, at 47-48.

²⁷⁴ We provide the technical aspects and more extensive statistical results in our companion article for a social science audience. See Daniel E. Ho et al., *Quality Review of Mass Adjudication: A Randomized Natural Experiment at the Board of Veterans' Appeals, 2003-16*, 35 J. L. ECON. & ORG. __ (forthcoming 2019) (draft on file with authors).

aggregated data from agency output, we secure rich, internal data of every case that the agency adjudicated from 2002-2016, comprising nearly 600,000 cases. As discussed in Part III, the QR program ran formally unchanged from November 1, 2002, to November 15, 2016, giving us a long observation period to test for subsequent disposition of cases. The “Veterans Appeals Control and Locator System” (“VACOLS”) dataset includes rich information about every veteran’s case – e.g., service period, fine-grained medical diagnosis, and the subsequent disposition of the case at CAVC if appealed. It is the same that BVA used to administer the program, allowing us to construct rigorous statistical tests for its efficacy. This rich data means that we are able to provide very precise estimates of effect sizes.

Fourth, the design of the program allows us to construct an extremely credible test for its effectiveness. Most importantly, a computer *randomly* selected 5% of original cases, or VLJ decisions on cases before BVA for the first time, to quality review. Random assignment ensures that those cases selected for QR are comparable in all respects to those not selected for QR, allowing us to attribute outcome differences to the QR intervention. Fifth, our data indicates which cases were selected for quality review and what errors, if any, the QR team identified for selected cases. VACOLS thus enabled us to divide all BVA decisions into a control group (cases not selected for quality review) and a treatment group (cases selected for quality review).

Random assignment, coupled with this rich data, hence allows us to test whether the program worked as intended, by comparing QR cases and control cases. If the program worked, we would expect veterans to appeal a much smaller percentage of QR cases than control cases. To determine whether to “call” a “substantive error” – basically a problem other than a formatting mistake – the QR attorney reviewing the draft decision would decide whether the opinion exhibited “a deficiency that would be outcome determinative, that is, result in the reversal or remand of a BVA decision by” the CAVC.²⁷⁵ If she found such an error, the QR team would prepare and circulate an “exception memorandum” describing the error for the VLJ. The VLJ could revise the draft decision to correct the error, ignore the exception memorandum, or make a formal challenge to BVA’s Chief Counsel for Policy and Procedure.²⁷⁶ The express purpose was to correct errors, so as to reduce the likelihood of an appeal to, and reversal or remand by, CAVC. In addition, we would expect CAVC to reverse or remand a smaller number of issues that underwent QR. Exception memoranda would have prompted corrections in the small number of decisions with errors before they went out the door.

B. Results and Analysis

Our basic findings are damning: quality review had no causal effects for either original appeals or CAVC-remanded appeals. Veterans appealed 6% of VLJ decisions on original cases in the treatment group and 6% of the cases from the control group to the CAVC.²⁷⁷ Again, due to the size of the sample, these estimates are very precise, allowing us to rule out effects of any substantial magnitude. The top panel of Figure 1 plots the proportion of QR and control cases

²⁷⁵ See *supra* p. 39.

²⁷⁶ Paul Sorisio, *Focus on Quality: The Internal Workings of Quality Review*, at 12 (Dec. 4, 2011).

²⁷⁷ Through a series of robustness analyses we find marginal evidence that the QR program may have had an effect for cases back on remand from CAVC. QR may have caused a 1% reduction in the rate at which veterans took decisions in CAVC-remanded cases back to the CAVC. With 5,622 CAVC-remanded cases undergoing quality review, we estimate that the program has helped BVA avoid roughly sixty appeals over fifteen years. This effect is de minimis. During this time, veterans appealed 57,170 cases to the CAVC in total. These figures are in the CAVC’s Annual Reports, 2003 to 2017, which are available here: <https://www.uscourts.cavc.gov/report.php>.

appealed to CAVC in red and blue, respectively, with 95% confidence intervals that overlap. In addition, our tests show that quality review had virtually no impact on the likelihood that the CAVC would reverse or remand a BVA decision. Of original cases appealed to CAVC, roughly 75% of QR cases were vacated and remanded by CAVC, compared to 76% of control cases. The bottom panel of Figure 1 displays the set of typical outcomes for QR and control cases, in red and blue respectively, showing no differences of any substantive magnitude.

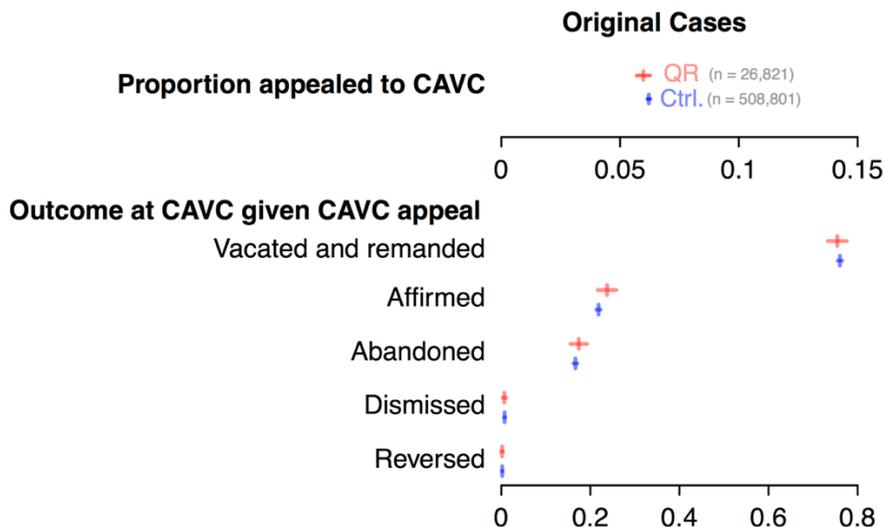


Figure 1: The impact of QR on case outcomes at CAVC. In the top panel, we plot the rate at which cases that went through QR (in red) and control cases (in blue) are appealed to CAVC. In the bottom panel, we plot the proportion of CAVC decisions in which at least one issue was subject to each disposition (e.g., vacated and remanded), again separately for cases that went through QR (in red) and control cases (in blue). Vertical dashes indicate the point estimates and horizontal lines indicate 95% confidence interval. These results show that the QR intervention had no substantive or statistically significant effect on CAVC outcomes: the QR and control outcomes are all comparable. Because the sample size involves 26,821 QR cases and 508,801 control cases, our estimates are extremely precise.

Why did the QR program fail by its own terms? We interviewed a wide range of officials, including CAVC judges, VLJs, staff attorneys, and senior officials at BVA and the VA, to understand the dynamics in more detail. One possibility is that CAVC decision-making may itself be hard to predict.²⁷⁸ If so, a QR attorney simply could not predict which cases would get remanded if appealed. By this account, the failure of quality review at BVA does not reflect poorly on BVA but on the CAVC. Or perhaps the QR example illuminates the futility of an accuracy measure pegged to expected outcomes on appeal. Yet the data refute this argument. When the QR team identified erroneous BVA decisions, those decisions were also more likely to be reversed and remanded by CAVC.

Another explanation shifts the spotlight to VLJs. Perhaps they simply ignore exception memoranda and sign decisions with errors likely to prompt remands by the CAVC. This hypothesis hardly presents BVA in a favorable light. The QR program cost BVA about \$780,000 in salaries alone each year.²⁷⁹ Moreover, QR personnel, generally thought to be among the

²⁷⁸ E.g., Telephone Interview with Interview Subject #1, Transcript at 5-6 (Aug. 27, 2018) (“It really is very judge dependent . . . at the Court of Appeals for Veterans Claims.”).

²⁷⁹ Ho et al., *supra* note 274, at 11.

agency's most successful staff attorneys, were taken off decision writing teams for the detail.²⁸⁰ Given this commitment of financial and attorney resources by an agency struggling with backlogs, routine VLJ indifference to exception memoranda would indict BVA for a serious management failure. Our interviews do not corroborate this hypothesis. According to one official, VLJs "almost always change" their drafts upon receiving an exception memorandum;²⁸¹ this impression squares with the institutional knowledge one of us acquired during his 12 years at BVA. Moreover, even if some VLJs were indeed indifferent to exception memoranda, this indifference surely would not have been uniform across all VLJs. Presumably at least some of them valued QR feedback, wanted to avoid mistakes, and gratefully incorporated suggested changes before saddling veterans with erroneous decisions.²⁸² We would thus expect to see statistically significant differences in the effect of quality review from VLJ to VLJ. But the data reveals no evidence of such heterogeneous effects across VLJs. The VLJ indifference hypothesis also does not explain why cases given a clean bill of health by the QR team still result in roughly three-quarters of cases being reversed and remanded at CAVC.

This latter fact suggests the most compelling explanation for the program's inefficacy: the standard of review. Formally QR attorneys were supposed to call errors based upon their prediction of CAVC outcomes. Yet likely due to pressures to report high accuracy rates, which the Chairman reports annually, the QR program came to grade VLJ decisions on a more forgiving curve. While the program likely corrected some errors, its relative leniency meant that a lot of cases destined for CAVC remands met the error-free standard as QR attorneys applied it. Several pieces of evidence support this explanation. First, when we examine variation in QR attorney citation rates of error, we see that higher error rates correlate with higher CAVC reversal and remand rates.²⁸³ The more stringent a standard of review a QR attorney used, in other words, the more her standard aligned with the CAVC's.

Second, we compare the types of errors that attorneys called with the reason for CAVC remands. By far the largest number of CAVC remands involve a VLJ's failure to explain the "reason or basis" of an opinion.²⁸⁴ This is veteran's law jargon for administrative law's "deeply embedded" obligation that an agency provide a "reasoned explanation" for its action.²⁸⁵ QR attorneys called reasons and bases errors in under 5% of cases they reviewed. This rate increased to 10% for cases these attorneys reviewed that veterans subsequently appealed to the CAVC (a difference that further corroborates that QR attorneys could meaningfully identify lower quality cases). By contrast, the CAVC remanded 62% of appeals for inadequate "reasons or bases."

A 2010 memorandum written by BVA's Deputy Vice Chairman confirms what these numbers suggest, that QR attorneys used a far less stringent metric to judge the adequacy of a

²⁸⁰ *Id.*

²⁸¹ Interview Subject #4, *supra* note 4, at 19; *see also* Interview Subject #2, *supra* note 80, at 15; Interview Subject #3, *supra* note 22, at 5.

²⁸² *Cf.* Interview Subject #1, *supra* note 278, at 3 ("[S]ome [VLJs] welcomed . . . correcting a decision that was not legally correct before it went out.").

²⁸³ In the vast majority of instances, the QR program assigned cases to attorneys in chronological order. In other words, when an attorney was ready to receive her next decision, she simply took the decision next up in the QR queue. No QR attorney, then, was more or less likely to encounter an error. This random assignment enabled us to measure the stringency of each QR attorney's standard of review. In rare instances, the chief of the QR office would make an adjustment to this workflow to ensure balance of workloads across attorneys.

²⁸⁴ *See generally* Ridgway & Ames, *supra* note 109.

²⁸⁵ Daniel A. Farber & Anne Joseph O'Connell, *The Lost World of Administrative Law*, 92 TEX. L. REV. 1136, 1152 (2014).

decision's "reasons or bases." The Deputy Vice Chairman apparently wrote the memorandum to defend the FY 2009 accuracy rate the QR program generated after the VA's Office of General Counsel ("OGC") expressed "'reservations regarding whether external data support'" it.²⁸⁶ The OGC had pointed out that the QR program estimated errors in slightly less than 3,000 cases VLJs decided in FY 2009, more than 1700 fewer cases than veterans appealed to the CAVC that year, and about 100 fewer than the number the CAVC remanded. The Deputy Vice Chairman included an argument about "reasons or bases" remands among his responses to OGC:

While some of [the "reasons or bases"] remands are necessary, in many instances the Court remands the case for an additional explanation as to matters that are not essential to the final factual or legal conclusions already reached by the Board or matters as to which the Board's views reasonably may be discerned from its decision. A "reasons or bases" remand does not necessarily indicate that the Board committed a legal or factual error, but may only reflect the Court's judgment that an amplified explanation is desired with respect to a particular issue. The "reasons and bases" standard is both highly subjective and inconsistently applied. Therefore, it is often difficult to predict when the Court will remand a case for this reason.²⁸⁷

This statement is remarkable to the extent that it suggests BVA's disavowal of its obligation to provide reasoned explanations for its decisions – a core duty for any agency. Regardless, the statement also indicates that a lesser standard for adequate "reasons or bases" prevailed at BVA. The CAVC may demand an "amplified explanation," but at BVA, a VLJ's explanation is sufficient if it addresses "matters . . . essential to . . . final factual or legal conclusions," or if the VLJ's "views reasonably may be discerned"

Third, our interviews corroborate the standard of review explanation and describe informal pressures to soft-pedal the QR process. As one official put it most pointedly,

The standards of quality review . . . have loosened up over time. The standard that we used to be following was essentially [whether CAVC] would be likely to vacate and remand [We] moved off of that standard quite some time ago to a lesser standard.²⁸⁸

Another told us that "the QR office knew that it was much less strict than the [CAVC], and that [the office's] standard was much more generous than what the [CAVC] was doing."²⁸⁹ A third expressed the difficulty of reviewing VLJ work in the role of a staff attorney, who might later serve as an employee to the VLJ. This person linked pressure not to call too many errors with an attorney's time-limited stint on the QR detail:

It was informal information . . . [a]long the lines of . . . "this is just a detail and you're going to be coming back to the decision teams [supervised by VLJs.]

²⁸⁶ Memorandum from Vice Chairman, *supra* note 276, at 1.

²⁸⁷ *Id.* at 2.

²⁸⁸ Interview Subject #4, *supra* note 4, at 2-3.

²⁸⁹ Interview Subject #2, *supra* note 80, at 12.

[T]hen you're not going to be very popular That could impact your career going forward

“There was a sense of do you want to poison the well,” this official continued: “[I]f you do your job too well, you're going to anger or upset” VLJs.²⁹⁰

Fourth, the evolving standard of review is consistent with BVA's decision in 2017 to jettison a QR error standard pegged to predictions of CAVC outcomes in favor of the more explicitly lenient “clear and unmistakable” standard. Just as BVA's backlog exploded, it codified the extremely deferential standard. This change gives ample reason to doubt the integrity of the currently reported accuracy rate, and it suggests how BVA prioritizes the appearance of accuracy over more rigorous evaluation of decisional quality.

* * * *

Our analysis shows that there no reason to believe that the QR program improved the overall accuracy of VLJ decision-making. As one official confided to us, the open secret was that “Everyone . . . knows it's kind of a sham.”²⁹¹ Both quality assurance initiatives and individual procedural rights can yield error avoidance or error correction in particular cases. The former's supposed institutional advantage lies with its capacity to improve decisional quality systemically. Yet the lenient standard of review QR attorneys used changed the program from one of quality improvement to public relations.

There is little evidence of systematic quality improvement. Given the rate at which QR attorneys called errors, a VLJ who decided 700 cases annually could expect an exception memorandum for less than 1% of her decisions.²⁹² Not only did the program generate little targeted feedback for VLJs on their specific error tendencies. It also gave them little incentive to improve. It comes as little surprise that BVA has not factored QR results into evaluations of VLJ performance.²⁹³ A VLJ confronted with an exception memorandum could point to the dozens of his decisions given a clean bill of health and reject the error as nothing more than a one-off mistake.

On the other hand, there is abundant evidence that the high accuracy rate was useful as a matter of public relations. To judge by its performance before the CAVC, BVA struggles to decide cases accurately. By one measure, the government's position is not “substantially justified” in 70% of CAVC cases, a figure Chief Justice Roberts found “really startling” when it came up in a 2009 case before the Supreme Court.²⁹⁴ The QR program generates a helpful counterweight, one BVA deploys to its advantage.²⁹⁵ Before Congress and in its budget requests, BVA officials repeatedly tout the high accuracy rate as demonstrating how well the agency operates. In an oversight hearing by the House Veterans Affairs Committee in 2007, for instance, the BVA

²⁹⁰ Interview Subject #3, *supra* note 22, at 7-8.

²⁹¹ Interview Subject #7, *supra* note 266, at 14.

²⁹² If about 6% of a VLJ's decisions were sampled for quality review, and if QR attorneys wrote staff attorneys for about 7% of the cases they reviewed, then the VLJ would receive about three exception memoranda per year. Of course, presumably exception memoranda were not evenly distributed, but even a VLJ who received twice the number of exception memoranda than the average would receive them in less than 2% of his decisions.

²⁹³ Interview Subject #3, *supra* note 22, at 10 (“I was never aware” of QR results factoring into performance evaluation “and that was something that always bugged me . . .”).

²⁹⁴ Transcript of Oral Argument at 52, *Astrue v. Ratliff*, 558 U.S. 987 (2009) (No. 08-1322).

²⁹⁵ One official agreed that the accuracy rate “is kind of an effective way” for BVA “to respond” when it gets public criticism for poor quality decision-making. Interview Subject #7, *supra* note 266, at 14.

Chairman described the quality review system as a “fine program,” boasted that GAO had “applauded our program and said it was actually exemplary,” and reported a “93 percent error-free rate.”²⁹⁶ Indeed, BVA’s public use of the accuracy rate to respond to criticism of its quality has been consistent throughout the entirety of the program, even after an internal study found that the accuracy rate “will be difficult to justify as a valid measurement if confronted by external stakeholders.”²⁹⁷ In another oversight hearing by the House Veterans Affairs Committee in 2018, the BVA Chairman reported that its accuracy rate was “over 92%.”²⁹⁸ This is what the accuracy number is for, an official remarked: it “goes in the annual report” BVA submits to Congress, and thus the number “has to be as good as possible.”²⁹⁹

V. IMPLICATIONS

BVA’s quality review program is a case study in failure. It and the checkered history of quality assurance at SSA and EOIR have implications for theories of internal administrative law, particularly for scholarship that attempts to describe the relationship between internal and external sources of legal governance. This scholarship tends to criticize judicial and congressional interventions for their interference with internal agency attempts to achieve consistent, reasoned decision-making. It has neglected the pressures that can lead internal administrative law into dysfunction. Experiments with quality assurance help identify and categorize these pressures. They also beg the question of whether a more nuanced relationship between external and internal law than what the existing literature describes might produce better overall results.

Our findings’ policy implications are our effort to answer this question. We suggest several interventions that can improve an agency’s quality assurance program, perhaps the best chance agencies have to achieve their constitutional obligation to an adequate threshold of decisional quality. These interventions include not only best practices for agency governance, but also legal changes that legislatures and courts can make.

A. Theoretical Implications

1. Internal Administrative Law and the Threat from Outside

Gillian Metzger and Kevin Stack describe the relationship between external and internal administrative law as a largely jurispathic one. The more “lawlike or manifestly binding [an] agency’s self-regulation is,” the more likely a court is to transform that regulation into an externally enforceable obligation or to invalidate it.³⁰⁰ The paradigmatic example is the guidance document

²⁹⁶ *Board of Veterans’ Appeals Adjudication Process and the Appeals Management Center: Hearing Before the Subcomm. on Disability Assistance and Memorial Affairs of the H. Comm. on Veterans’ Affairs*, 110th Cong. 39 (Testimony of James P. Terry, Chairman, Board of Veterans’ Appeals).

²⁹⁷ Office of Quality Assurance, *QA Metrics Proposal* (on file with authors).

²⁹⁸ *Is VA Ready for Full Implementation of Appeals Reform?: Hearing Before the Subcomm. on Disability Assistance and Memorial Affairs of the H. Comm. on Veterans’ Affairs*, 115th Cong., Video at 1:16:40-1:16:48 (2018) (Statement of Cheryl Mason, Chairman, Board of Veterans’ Appeals), <https://republicans-veterans.house.gov/calendar/eventsingle.aspx?EventID=2276> (last visited Jan. 24, 2019).

²⁹⁹ Interview Subject #7, *supra* note 266, at 12.

³⁰⁰ Metzger & Stack, *supra* note 138, at 1278.

the Obama Administration crafted to afford undocumented parents of U.S. citizen children relief from deportation.³⁰¹ The Department of Homeland Security took this action to ensure consistency in the exercise of discretion among frontline immigration officers, by binding them to make certain decisions when confronted with certain classes of immigrants. This feature made the guidance document law-like, but it also sealed the guidance document's doom. Because it bound these officials, the Fifth Circuit reasoned, the guidance document was really a legislative rule and thus procedurally deficient for failure to go through notice-and-comment.³⁰² The result of such external intrusions into internal governance, Metzger and Stack argue, are "incentives for agencies to be less specific, less decisive, and less clear in their internal documents" – that is, to create less internal administrative law.³⁰³

Metzger and Stack allow that, in theory, Congress or the courts could govern agencies in a way that facilitates the development of sound internal governance.³⁰⁴ But they are skeptical that this result will obtain in practice.³⁰⁵ Mashaw's doubt about a copacetic relationship between external and internal law, especially for mass adjudication, runs even deeper. He concludes his canonical account of disability benefits adjudication not only with optimism about SSA's capacity to govern itself justly, but also convinced that "[t]he external legal order" poses "*the* threat" to this outcome.³⁰⁶ The Due Process Revolution, for instance, forced SSA to accommodate a model of justice rooted in individual participation and desert, one epitomized by the sorts of processes that litigants enjoy in court.³⁰⁷ The result was "a tale of stress and woe,"³⁰⁸ with a shift of significant control over disability benefits adjudication to a cadre of Article III-like officials who resist managerial control and fall prey to inefficiency.³⁰⁹ Reflecting more generally on external legal control of disability benefits adjudication, Mashaw describes possible interventions as "essentially bankrupt."³¹⁰ Congress and the courts should leave the agency to govern itself.³¹¹

2. Internal Pressures and Distortions

While enumerating the ways that external oversight and review can distort internal governance, leading accounts have spent less time identifying the pressures that can distort internal administrative law from within.³¹² Ironically, Metzger's and Stack's optimism for internal administrative law may reflect an insufficiently internal inquiry into relevant governance strategies. To them, guidance documents and other formal expressions of internal governance – material readily observable outside the agency – constitute this corpus. But this sort of instrument's capacity to generate reasoned, consistent decision-making depends as much on

³⁰¹ *Id.* at 1241.

³⁰² *Texas v. United States*, 809 F.3d 134, 170-177 (5th Cir. 2015).

³⁰³ Metzger & Stack, *supra* note 138, at 1288.

³⁰⁴ *Id.* at 1291; *id.* at 1295-1296. Indeed, in other work, Metzger argues for a judicially enforceable constitutional "duty to supervise" that would entail some sort of judicially administered structural remedy devised to improve agency processes at a wholesale level. Metzger, *supra* note 99, at 1915.

³⁰⁵ Metzger & Stack, *supra* note 138, at 1288-1289; *id.* at 1292; *id.* at 1296.

³⁰⁶ MASHAW, *supra* note 116, at 222 (emphasis added).

³⁰⁷ *Id.* at 29-31; *id.* at 192-193.

³⁰⁸ *Id.* at 41.

³⁰⁹ *Id.* at 192.

³¹⁰ *Id.* at 192.

³¹¹ Merrill, *supra* note 11, at 53.

³¹² MASHAW, *supra* note 116, at 158-163; Metzger & Stack, *supra* note 138, at 1303-1304.

agency culture, agency personnel, the agency’s political environment, and other harder-to-observe variables as on the form that it takes.

The contrast between the standard of review as written and the standard of review Board attorneys actually used for quality review nicely illustrates how administrative law-in-action can diverge from administrative law-in-books. While standards of review “are not precision instruments,”³¹³ they “express[] a mood,” as Felix Frankfurter insisted, and “that mood must be respected.”³¹⁴ Our findings lend weight to a classic critique of these standards, that they “have no more substance at the core than a seedless grape”³¹⁵ Two sets of experts in veterans’ law, Board attorneys and CAVC judges, differed considerably in their decisions while ostensibly applying the same standard of review. The institutional context mattered significantly to the attorneys’ application of the law.

Quality assurance is a useful example of internal administrative law-in-action, because experiences with it owe their success or failure as much to context as to the merits of some formally stated policy. The initiatives described here illuminate at least three pressures that can distort a regime of internal administrative law: resistance problems, incentive problems, and problems of conflicting goals.

Resistance Problems. Those governed by a regime of internal law may resist its control. ALJs responded to SSA’s early experiments with lawsuits, obtaining settlements and eventually rules that handcuffed the agency’s power to use its programs to target low-performing adjudicators. Formal accountability for low quality decision-making remains extremely rare to this day. Since 2007, for instance,³¹⁶ SSA has imposed formal discipline in several instances based on ALJ “workload performance,” but each has involved case processing, not decisional accuracy.³¹⁷ VLJs and IJs have less law they can use to shield themselves from efforts to control their decision-making.³¹⁸ Quality review at BVA, however, suggests that agency culture may provide as solid a foundation for resistance as formal legal authority.³¹⁹ VLJs are judicial officers, and they stand higher on VA’s bureaucratic ladder than quality review personnel. Both are institutional facts that may prompt resistance.³²⁰ One official involved with quality review recalled a conversation about the program with a VLJ early in his tenure. “The thing you need to understand,” this VLJ told the official, hinting at how much feedback the VLJ was willing to receive, “is that I have a piece of paper signed by the President on my wall and you do not.”³²¹

³¹³ Kristen E. Hickman & Matthew D. Krueger, *In Search of the Modern Skidmore Standard*, 107 COLUM. L. REV. 1235, 1250 (2007).

³¹⁴ *Universal Camera Corp. v. NLRB*, 340 U.S. 474, 487 (1951).

³¹⁵ Ernest Gellhorn & Glen O. Robinson, *Perspectives on Administrative Law*, 75 COLUM. L. REV. 771, 780 (1975); see also David Zaring, *Reasonable Agencies*, 96 VA. L. REV. 135, 153-154 (2010). Cf. *United States v. Kinney*, 919 F.2d 405, 422-423 (7th Cir. 1990) (Posner, J., dissenting) (discussing standards of review more generally).

³¹⁶ On the paucity of disciplinary action before 2007, see, e.g., *Delays in Social Security Appeals*, *supra* note 152, at 81; Lubbers, *The Federal Administrative Judiciary: Establishing an Appropriate System of Performance Evaluation for ALJs*, 7 ADMIN. L.J. AM. U. 589, 599-601 (1994). Cf. *Shapiro v. Social Security Administration*, 800 F.3d 1332, 1338 (Fed. Cir. 2015) (commenting on the “virtually insurmountable” barrier to ALJ removal).

³¹⁷ OIG, SOC. SEC. ADMIN., A-12-11-01138, OVERSIGHT OF ADMINISTRATIVE LAW JUDGE WORKLOAD TRENDS 15 (2012); *id.* at E-1; see also GAO, *supra* note 217, at 33.

³¹⁸ See *supra* p. 7.

³¹⁹ On culture and resistance, see Nina Rabin, *Victims or Criminals? Discretion, Sorting, and Bureaucratic Culture in the U.S. Immigration System*, 23 S. CAL. REV. L. & SOC. JUST. 195, 225-226 (2014).

³²⁰ William H. Simon, *Where is the “Quality Movement” in Law Practice?*, 2012 WIS. L. REV. 387, 402-403.

³²¹ Interview Subject #2, *supra* note 80, at 6; see also Interview Subject #3, *supra* note 22, at 3-4 (“Who, as they would say, were approved by the president and who the heck was I to be saying that they were wrong”). As noted in

A recognition that officials subject to an internal governance regime may resist may lead the agency to account for this resistance when it designs the regime, possibly at the expense of sound administration. The highly deferential standard of review SSA adopted for its first quality assurance program, for example, resulted in part from the agency's concern that a more searching one would provoke ALJ backlash. BVA's official standard of review adopted in 2002 asked reviewers to predict how a particular case would fare at the CAVC if appealed. According to one person familiar with the program, however, within a decade reviewers no longer attempted to predict CAVC outcomes when they looked for errors.³²² This evolution fits our findings, that quality reviewers used a more deferential standard than the CAVC did notwithstanding the stated equivalence between the two.³²³

One possible reason for this deference has to do with the strained relationship between BVA and the CAVC. No CAVC judge has ever served at BVA and thus none has ever experienced first-hand BVA's docket and processes.³²⁴ As a result, one official told us, "there's definitely an attitude [among VLJs], and I think very, very much justified, that [CAVC judges] have no idea what's going on."³²⁵ "I personally have no respect for the Court," an official declared, explaining that its judges "nitpick everything."³²⁶ A standard of review that rigorously graded the quality of VLJ decisions by a metric that predicted what CAVC judges would do might have prompted resentment, and then resistance, by VLJs.

Incentive Problems. Incentives can operate at both the individual official and the agency level to distort regimes of internal administrative law. First, a regime can fail if officials responsible for its administration have an incentive to administer the regime in a manner that departs from its stated purpose. BVA staff attorneys, for example, may have had an incentive to review VLJ decisions lightly. As one official told us, "everybody . . . who had ambitions of becoming selected to be a judge one day" "competed for" selection for a quality review detail.³²⁷ In fact, of the 43 staff attorneys who served on quality review details from April 2002 to November 2016, 19 have become VLJs and another 5 have become Senior Counsel, a highly competitive position and stepping stone to a VLJ appointment.³²⁸ Knowing that she is on track to win further promotion, a staff attorney assigned to quality review might consciously or subconsciously worry about angering VLJs, a constituency with influence over her future,³²⁹ by administering too

Part IV, officials told us that VLJs generally revised opinions after receiving exception memoranda. This practice does not rebut concerns about resistance at BVA. It is possible – and indeed consistent with what interview subjects told us – that the QR program ensured that VLJs would accept exception memoranda so long as the program issued few of them. By anticipating and dodging VLJ resistance, in other words, the QR program could maintain the effectiveness of the relatively small number of exception memoranda.

³²² Interview Subject #1, *supra* note 278, at 6.

³²³ See also *Board of Veterans' Appeals Adjudication Process and the Appeals Management Center: Hearing Before the Subcomm. on Disability Assistance and Memorial Affairs of the H. Comm. on Veterans' Affairs*, 110th Cong. 39 (Statement of Barton Stichman).

³²⁴ Interview Subject #2, *supra* note 80, at 7.

³²⁵ *Id.*

³²⁶ Interview Subject #1, *supra* note 278, at 10.

³²⁷ Interview Subject #2, *supra* note 80, at 5.

³²⁸ One of us with substantial experience at BVA compiled these figures, using his own institutional knowledge supplemented by information from BVA annual reports and internal agency documents obtained through FOIA. Documentation for these figures is available from the authors.

³²⁹ For a description of the VLJ appointments process, see *Cheryl Mason, Chairman of the Board of Veterans' Appeals within the U.S. Department of Veterans Affairs*, THE BUS. OF GOV'T RADIO HOUR (Jan. 14, 2019) (downloaded using iTunes).

searching a standard of review.³³⁰ Even if a staff attorney does not harbor judicial ambitions, she might still review decisions lightly. The time limited nature of a quality review detail means that the attorney will cycle back to writing decisions under the supervision of a VLJ, an official who can make or break her job.³³¹

If an incentive to defer does affect BVA quality review, it results from a long-recognized and obvious design flaw – the use of bureaucratic subordinates to evaluate the work of bureaucratic superiors.³³² But the flaw has persisted for two decades, a stubbornness that may indicate the operation of a different set of incentives. Depending on how the errors skew, a malfunctioning quality review program may actually serve the agency’s political interests. An entrenched political consensus favors generosity in the veterans’ benefits system.³³³ BVA may thus face no political consequences if its adjudicators tend to err in veterans’ favor. EOIR’s failure to adopt any meaningful quality assurance initiatives for immigration courts in the face of longstanding criticism of their systemic accuracy makes no sense if the agency prioritizes sound IJ administration of immigration policy. But it does jive with political forces if IJs tend to err in ways that disfavor immigrants. An increased pace of deportations fits the pro-enforcement preferences of recent administrations and the institutional commitments of the Attorney General as the country’s chief law enforcement officer.³³⁴

Politics in the form of inter-branch relations may disincentivize improvement as much as politics in the form of ideological commitment. The Government Performance and Results Act requires federal agencies to list performance goals in annual reports presented to Congress, and to identify and discuss performance measures that indicate the agency’s progress toward the goals’ fulfillment.³³⁵ Ostensibly an accountability mechanism, the GPRM, as Sidney Shapiro and Rena Steinzor argue, gives agencies an incentive to devise performance goals and metrics to “ensure that they can ‘pass’ their own grading criteria.”³³⁶ BVA’s accuracy rate, never dipping below 89% from 2003 to 2018 regardless of its caseload, certainly seems like “a sunny set of invented statistics designed to reassure [its] overseer[] that [it is] doing fine”³³⁷ Congress has little ability to dissect the process generating these numbers or to conduct its own quality audit of something as complex as veterans’ benefits decision-making. It effectively has to take the results as true.³³⁸

³³⁰ Interview Subject #3, *supra* note 22, at 8 (“You want to do your job, but if you do your job too well, you’re going to anger or upset [VLJs]. That could have a significant impact on your career”). *But see* Telephone Interview with Interview Subject #5, Transcript at 13-14 (disagreeing that concern for career prospects impacted the QR process).

³³¹ One official described to us how a VLJ needs to manage his docket with care to ensure that all staff attorneys under the VLJ’s supervision can meet their annual production quota. Interview Subject #4, *supra* note 4, at 13.

³³² *Cf.* MASHAW, *supra* note 150, at 176 (insisting that “considerable care must be taken to ensure the independence of the quality assurance staff”).

³³³ *E.g.*, Linda D. Jellum, *Heads I Win, Tails You Lose: Reconciling Brown v. Gardner’s Presumption that Interpretive Doubt Be Resolved in Veterans’ Favor with Chevron*, 61 AM. U. L. REV. 59, 64 (2011).

³³⁴ AMERICAN BAR ASSOC. COMM’N ON IMMIGR., REFORMING THE IMMIGRATION SYSTEM 2-24/2-25 (2010); Stephen H. Legomsky, *Deportation and the War on Independence*, 91 CORNELL L. REV. 369, 372-375 (2006).

³³⁵ *E.g.*, CLINTON T. BRASS, CONG. RESEARCH. SERV., R42379, CHANGES TO THE GOVERNMENT PERFORMANCE AND RESULTS ACT (GPRA): OVERVIEW OF THE NEW FRAMEWORK OF PRODUCTS AND PROCESSES 5 (2012).

³³⁶ Sidney A. Shapiro & Rena Steinzor, *Capture, Accountability, and Regulatory Metrics*, 86 TEX. L. REV. 1741, 1744 (2000).

³³⁷ *Id.* at 1744.

³³⁸ Interview Subject #4, *supra* note 4, at 5-6 (describing the complexities of veterans’ law and the difficulties they pose for congressional oversight).

BVA has every incentive to continue with a program that helps it with congressional relations, however ineffective it is as a strategy to protect decisional quality.

The Problem of Conflicting Goals. Agencies handling anything more than a modestly complicated task will have multiple, potentially conflicting goals to pursue.³³⁹ When agencies are left to govern themselves, at least three related phenomena may result. First, the agency may prioritize which goals it pursues in a manner that best serves its institutional interests and not necessarily the interests of its constituents or the larger public good.³⁴⁰ Second, the agency may prioritize goals that have straightforward performance metrics over others whose measurement is more difficult.³⁴¹ Third, the agency may devise metrics or ways of presenting data about its accomplishments with its own short-term interests in mind, not necessarily to present the most accurate portrait of its achievement.³⁴²

The enduring conflict between quality and quantity illustrates these phenomena and how they can distort a regime of internal law. Because an agency can more readily measure the number of cases it decides than decisional quality,³⁴³ it will tend to favor the pursuit of quantity goals over quality ones.³⁴⁴ An agency may therefore sacrifice its quality assurance initiatives when backlogs increase. SSA's first experiment with own motion review ended in 1975 when the agency, facing an exploding docket, reassigned staff administering the program to case production.³⁴⁵ More recently, as backlogs surged and ALJ productivity declined,³⁴⁶ SSA suspended its inline quality reviews and disability case reviews and reallocated staff assigned to them to deciding cases.³⁴⁷ BVA did not respond to the spike in its case load by increasing its quality review program's capacity. Instead, it effectively halved the percentage of cases the program sampled for review and reduced the number of attorneys assigned to the office.

BVA's behavior over the past decade exemplifies how an agency's institutional interests affect its goal prioritization. A BVA official familiar with congressional relations told us that Congress's "concern is more focused on 'what are you doing to get rid of the backlog'" than quality.³⁴⁸ Even so, presumably a dramatic drop in the self-reported accuracy rate would spark oversight concerns.³⁴⁹ If rising caseloads do impact quality, then this sort of drop should happen – unless BVA grades itself on a more forgiving curve. Just this sort of inflation appears to have happened. The shift to the "clear and unmistakable error" standard and the shortening of quality review details to six months from two years, whatever their legitimate motivations, have helped

³³⁹ Eric Biber, *Too Many Things to Do: How to Deal with the Dysfunctions of Multiple Goal Agencies*, 33 HARV. ENV'T L. REV. 1, 12 (2009).

³⁴⁰ Donald P. Moynihan, *The Promises and Paradoxes of Performance-Based Bureaucracy*, in THE OXFORD HANDBOOK OF AMERICAN DEMOCRACY 279, 286-287 (Robert F. Durant ed., 2010).

³⁴¹ *Id.*; see also Biber, *supra* note 339, at 12.

³⁴² Moynihan, *supra* note 340, at 286-87.

³⁴³ Cf. MASHAW, *supra* note 116, at 158-60 (discussing the "normative ambiguity" of defining successful performance).

³⁴⁴ Moynihan, *supra* note 340, at 288 ("As a result" of an agency's emphasis on readily measured goals, "efficiency goals are often pursued at the expense of program quality . . .").

³⁴⁵ SURVEY AND ISSUE PAPER, *supra* note 85, at 25.

³⁴⁶ E.g., OIG, *supra* note 221, at 3.

³⁴⁷ GAO, *supra* note 217, at 40.

³⁴⁸ Interview Subject #5, *supra* note 330, at 13; see also Interview Subject #4, *supra* note 4, at 17-18 (insisting that Congress is almost wholly interested in the number of cases decided and timeliness indicators).

³⁴⁹ Cf. Interview Subject #4, *supra* note 4, at 18 (commenting on congressional interest in quality and insisting that "[i]f there's that dramatic of a drop then they'd start to say something").

BVA continue to report a very high accuracy rate even as it maintains what one congressperson called an “absolutely alarming” pace of decision-making.³⁵⁰

Also, as BVA’s backlog has grown, VA has subtly de-emphasized quality as a key performance measure. In 1998, BVA’s chairman declared that “[q]uality in appellate decision-making . . . is . . . the Board’s single most important goal.”³⁵¹ VA’s annual budget requests to Congress from 2008 to 2010 reflected this emphasis; they included a “deficiency-free decision rate” as BVA’s first performance measure.³⁵² From 2011 to 2014, the requests stopped reporting the accuracy rate, although they did refer to “expectations for quality, timeliness, and responsiveness” as an “objective.”³⁵³ Its requests from 2015 to the present, when caseloads began to soar, have omitted quality as a performance measure altogether. Instead they refer exclusively to case production.³⁵⁴

To be sure, a preference for quantity over quality is not itself problematic. An agency transparent about the impact that accelerated decision-making has on decisional accuracy could put the onus on Congress to respond, forcing it either to decide which goal it wants the agency to prioritize, or to commit sufficient resources so that the goals no longer conflict. But top officials anxious to avoid Congressional scrutiny or public outcry may prefer to soft-pedal any problem, even if a more frank acknowledgement might buttress a request for additional resources.³⁵⁵ As one major account of agency behavior posits, the agency stands to gain power and autonomy to the extent it can strengthen or defend its reputation for competent performance.³⁵⁶ Put conversely, “politicians do not respond to a structure fraught with three-alarm fires by supplying more fuel; they are just as likely . . . [to] begin choosing other venues in which to vest authority and power.”³⁵⁷ Recent testimony by the BVA Chairman at an oversight hearing illustrates this tendency. She disclaimed a need for additional resources, even as VLJ caseloads are anticipated to more than double over the next five years.³⁵⁸

³⁵⁰ *Assessing Whether VA is On Track to Successfully Implement Appeals Reform: Hearing Before the H. Comm. On Vet. Affairs*, 115th Cong., Video at 1:48:13 (2018) (Statement of Cheryl Mason, Chairman, BVA), <https://www.youtube.com/watch?v=AACgoVzBKKg> (last visited Jan. 30, 2019). Compare U.S. DEP’T OF VET. AFFAIRS, *supra* note 70, at 5 (projecting 81,033 decisions for FY 2018), with *id.* at 17 (reporting the FY 2018 to date accuracy rate). See also *Is VA Ready for Full Implementation of Appeals Reform?: Hearing Before the Subcomm. on Disability Assistance and Memorial Affairs of the H. Comm. on Veterans’ Affairs*, 115th Cong., Video at 1:16:40-1:16:48 (2018) (Statement of Chairman Mason), <https://republicans-veterans.house.gov/calendar/eventsingle.aspx?EventID=2276> (last visited Jan. 8, 2019).

³⁵¹ Memorandum No. 1-98-15 from the Office of the Chairman, Board of Veterans’ Appeals (May 14, 1998) (on file with authors).

³⁵² 3 U.S. DEP’T OF VET. AFFAIRS, CONG. SUBMISSION, FY 2010, at 5C-5; 3 U.S. DEP’T OF VET. AFFAIRS, CONG. SUBMISSION, FY 2009, at 5C-2; 2 U.S. DEP’T OF VET. AFFAIRS, CONG. SUBMISSION, FY 2008, at 7C-2.

³⁵³ 3 U.S. DEP’T OF VET. AFFAIRS, CONG. SUBMISSION, FY 2014, at 4C-4; 3 U.S. DEP’T OF VET. AFFAIRS, CONG. SUBMISSION, FY 2013, at 5C-4; 3 U.S. DEP’T OF VET. AFFAIRS, CONG. SUBMISSION, FY 2012, at 5C-4; 3 U.S. DEP’T OF VET. AFFAIRS, CONG. SUBMISSION, FY 2011, at 5C-5.

³⁵⁴ 3 U.S. DEP’T OF VET. AFFAIRS, CONG. SUBMISSION, FY 2017, at BVA-273-BVA-288-5; 3 U.S. DEP’T OF VET. AFFAIRS, CONG. SUBMISSION, FY 2016, at 273-282; 3 U.S. DEP’T OF VET. AFFAIRS, CONG. SUBMISSION, FY 2015, at BVA-9.

³⁵⁵ Cf. Moynihan, *supra* note 340, at 288 (noting that conflicting goals and goal ambiguity will lead agencies to privilege “short-term goals over long-term measures of effectiveness”).

³⁵⁶ DANIEL CARPENTER, REPUTATION AND POWER: ORGANIZATIONAL IMAGE AND PHARMACEUTICAL REGULATION AT THE FDA 46 (2010).

³⁵⁷ *Id.* at 54.

³⁵⁸ *Is VA Ready*, *supra* note 350, at 1:12.

This link between reputation and power may explain an arguably counterintuitive pattern in the history of SSA's quality assurance programs. SSA pursued three major quality assurance initiatives at times of rising backlogs, pairing a push for an increased rate of decision making with centrally administered efforts to ensure that ALJs met an adequate threshold of decisional quality. Perhaps this coupling indicates that SSA appreciates the threat that an emphasis on quantity poses to quality, and it has attempted to mitigate the impact of faster decision-making on decisional accuracy. A more cynical take views SSA's quality assurance efforts as window dressing, to give the agency institutional cover as it really pursues the more easily measurable – and therefore favored – quantity goal.

B. Improving Quality Assurance: Externally Influenced Internal Administrative Law

Dysfunction in internal administrative law is not irreducible. An agency could implement meaningful quality assurance programs without action from Congress or the courts. With little apparent prodding from outside the agency,³⁵⁹ SSA has recently experimented with an extensive set changes to its quality assurance programs. Its chief designer claims that they have improved decisional quality even as ALJs decide more cases.³⁶⁰ External observers need to confirm these changes' efficacy. But they suggest something promising.

One of us likewise spearheaded a change to BVA's quality review program without any external mandate to do so. Between November 2016 and September 2017, the revised program did not generate a crude, effectively meaningless overall accuracy rate. Rather, BVA staff attorneys used their review of draft decisions, along with other sources of information like CAVC remands, to identify recurring problems that tend to generate a disproportionate number of errors in VLJ decisions. This information was intended to support systemic interventions, such as focused training for all VLJs, or even broad policy changes. The only such "Targeted Quality Analysis" the revised program could complete before its abandonment determined that a particular requirement involving "extraschedular ratings" produced significant disparities between BVA understandings and CAVC case law and generated a large number of appeals. But these appeals when remanded almost never actually resulted in increased benefits for veterans. The report thus recommended that VA replace the regulation, which simply generated more "churn," with more specific, easily applied eligibility criteria. A category of mistakes would disappear, and deserving veterans would get their benefits more efficiently.

The fate of this revised program underscores the fragility of internal administrative law. When a single erroneous decision caused VA embarrassment, the BVA Chairman scuttled the program and insisted that quality review go back to generating the meaningless accuracy number. The old program only sampled about 6% of VLJ decisions for review, so in all likelihood BVA would have issued the embarrassing decision even had it remained in place. The only conceivable reason to bring back the accuracy number is to give BVA a rosy statistic to cite when someone castigates the quality of its work in the future. This development echoes SSA's decision to invigorate its quality assurance efforts after reports of about biased decision-making surfaced in 1992. The agency's action suggested its seriousness about quality, but it didn't grapple with the underlying problem that prompted criticism in the first place.

³⁵⁹ In fact, according to one observer, the person responsible for these changes implemented them without seeking any go-ahead from top agency officials. Interview Subject #2, *supra* note 80, at 17.

³⁶⁰ Ray & Lubbers, *supra* note 84, at 1606.

When pressures like resistance, incentives, and goal conflict interfere with internal reform, external intervention by Congress and the federal courts may help. None of what we now suggest would require an agency to do anything specific and instead leaves considerable room for internally controlled design and experimentation. Our proposals for better quality assurance illustrate how external legal interventions can facilitate, rather than retard, the development of good internal administrative law.

1. Improving Oversight

BVA's accuracy rate shows how an agency can use a meaningless metric to blunt oversight and public scrutiny. BVA has invoked it repeatedly to assuage overseers that its adjudication meets an adequate threshold of systemic quality.³⁶¹ To preserve the possibility of meaningful oversight, Congress should require agencies to disclose sufficient information about quality review to enable outsiders to determine what is being measured. The fact that BVA formally advertised one standard of review (prediction of CAVC disposition) and functionally deployed a more lenient one illustrates the risk of a lack of transparency. Just as scientists cannot publish results without disclosing methodology, agencies should not be able to publish performance measures without disclosing their measurement methods.

An agency ought to make public data that observers could use to evaluate a program's efficacy. Although we accessed data sufficient to evaluate BVA's quality review program, we only knew what to ask for because one of us worked for the agency. FOIA requests involve considerable guesswork and delays, ones that might defeat all but the most dogged efforts. The FOIA Improvement Act of 2016 has implemented a "Rule of 3," whereby three requests for the same disclosure triggers an agency's duty to make the disclosure without additional requests.³⁶² Given the obvious interest in information like the VACOLS data, an agency should not wait for multiple FOIA requests to discharge this obligation.³⁶³

To make this data truly meaningful, an agency should also be required to disclose details about how it designs and administers its quality assurance programs. The institutional setup of quality review at BVA alone should prompt skepticism about the accuracy rate. The use of bureaucratic subordinates to review their superiors' work, the weakening of the standard of review, and the shortening of attorney terms on quality review details all suggest that BVA grades itself on an increasingly lenient curve.

Adequate disclosure would help to blunt incentives an agency might have to design or continue with a quality review process that serves its short-term institutional interests and not necessarily those of the population whose cases it adjudicates. It can compensate for Congress's institutionally determined propensity to accept an agency's own measure credulously. External scrutiny may also force the agency to clarify which goals conflict, and thus better to force Congress

³⁶¹ *E.g.*, *Is VA Ready*, *supra* note 350, at 1:16:40; *Board of Veterans' Appeals Adjudication Process and the Appeals Management Center: Hearing Before the Subcomm. on Disability Assistance and Memorial Affairs of the H. Comm. on Veterans' Affairs*, 110th Cong. 30 (2007) (statement of James Terry, Chairman, Board of Veterans' Appeals) (using the accuracy rate from the Quality Review program to defend against claims that the quality of BVA's work was suffering).

³⁶² *E.g.*, U.S. Dep't of Justice, Office of Information Policy, *OIP Summary of the FOIA Improvement Act of 2016*, <https://www.justice.gov/oip/oip-summary-foia-improvement-act-2016> (last visited Jan. 8, 2019).

³⁶³ EOIR, for instance, now makes data on IJ decision-making available to the public without a FOIA request. See *EOIR Case Data*, <https://stg.justice.gov/eoir> (last visited Jan. 8, 2019).

to assume responsibility for articulating which goal it prioritizes and why. BVA's accuracy measure excuses Congress from the difficult task of ranking case production relative to decisional quality. Legislators can pressure VA to reduce backlogs without any consideration for the effect increased production has on decisional quality. Agency personnel can object to the ever-increasing demands on their time. But if quality does not suffer, then the result is simply that government employees work harder, hardly an outcome any member of Congress would likely lament.³⁶⁴ If confronted with the costs that the pressure to produce can have, Congress at the least may be forced to assume responsibility for goal prioritization.

Congress could institutionalize external scrutiny of an agency's quality review program by requiring an independent body, such as the agency's inspector general, to audit quality assurance initiatives on some regular basis. Without an instruction to pay attention to decisional quality, however, an inspector general may neglect it. SSA's IG has evaluated some of the agency's quality assurance initiatives, but VA's IG, when he has paid attention to BVA at all, tends to focus on picayune matters of little significance to the systemic integrity of its operations.³⁶⁵ Moreover, inspector general offices can go for long periods without leadership,³⁶⁶ and they may face funding shortfalls.³⁶⁷ In light of such institutional limitations, an IG appointed to oversee an agency the size of VA may sensibly decide to pay only irregular attention to a comparatively small division within it. The GAO is another option, but again, meaningful scrutiny would require some statutory mandate requiring it to scrutinize decisional quality regularly. The GAO has helpfully studied SSA quality assurance programs, and its critique of BVA's in 2002 prompted meaningful change.³⁶⁸ But without a statutory mandate the GAO takes its marching orders from congressional requesters, and its attention to decisional quality is episodic at best for that reason.³⁶⁹

2. Improving Incentives and Blunting Resistance

Simply empowering the oversight capacity of Congress could help decisional quality improve systematically, even if Congress provides no actual instruction about how the agency should assess quality. But Congress could go a step further without interfering with an agency's capacity to develop good internal governance strategies. It could mandate very basic design principles, and it could legislate good reporting requirements. By doing so, Congress could protect against distortion that resistance and incentive problems cause.

BVA's program is institutionally flawed because it tasks bureaucratic subordinates with the responsibility for reviewing the work of their bureaucratic superiors. SSA's programs, in contrast, mostly have relied on work done by Appeals Council personnel, who presumably have

³⁶⁴ Interview Subject #4, *supra* note 4, at 14 ("Congress doesn't care if a bunch of federal bureaucrats . . . have to work 100 hours a week.").

³⁶⁵ *E.g.*, OIG, U.S. DEP'T OF VET. AFFAIRS, 15-02747-314, BOARD OF VETERANS APPEALS: REPORT OF ADMINISTRATIVE INVESTIGATION ALLEGED PROHIBITED PERSONNEL PRACTICE (2016) (investigating a charge that members of a VLJ appointment panel disclosed confidential information about their process).

³⁶⁶ Anne Joseph O'Connell, *Vacant Offices: Delays in Staffing Top Agency Positions*, 82 S. CAL. L. REV. 913, 957 (2009).

³⁶⁷ *Strengthening the Unique Role of the Nation's Inspectors General*: Hearing Before the S. Comm. on Homeland Security and Governmental Affairs, 110th Cong. 46 (2007) (Statement of Glenn Fine) (claiming that IG offices' biggest challenge is chronic underfunding).

³⁶⁸ *See supra* p. 40-41.

³⁶⁹ On GAO work priorities, see generally U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-17-767G, GAO'S CONGRESSIONAL PROTOCOLS (2017).

no professional incentive to go easy on ALJs. On the surface, the structure of quality review at SSA seems superior and suggests an obvious change BVA could make. The issue of institutional design, however, raises more complex questions. ALJs can resent direction from SSA's Appeals Council, a body that some ALJs view as arbitrary and disconnected from the day-to-day realities of benefits adjudication.³⁷⁰ Review done by bureaucratic superiors, while blunting the incentive effect, might increase resistance. It also might deny a quality review program of the benefits of true peer review, where adjudicators with the same basic job duties review each other's work.³⁷¹

Congress does not need to specify a particular institutional structure for quality review. It could simply require that the structure ensure that reviewers lack any personal incentive to weaken their scrutiny, and that those reviewed have the least reason to resist feedback. A quality review initiative might, for instance, generate results to be used in performance evaluation, but also give adjudicators particular credit in their evaluations if they show improvement on identified areas of concern. An initiative could task personnel from the ranks of the adjudicators themselves with the review of peers' work; VLJs relieved of deciding cases would review other VLJs' decisions, for instance. The regime could anonymize review. If a quality review attorney at BVA, for instance, flags an error, the exception memorandum could issue in the program's name and not the attorney's. The reviewed VLJ would have no reason to resent a particular colleague.

Congress could also legislate reporting requirements that prompt an agency to design good quality assurance initiatives. To the extent that the GPRM requires something like BVA's accuracy rate, it demands a blunt measure of decisional quality subject to manipulation. A reporting requirement ought to place "less emphasis on the objective of precisely measuring government performance," and instead require agencies to "develop . . . more effective policy tools for guiding program management . . ." ³⁷² Congress could incentivize an agency to do so by insisting that it report not a crude numerical indicator of overall quality, but instead a description of issues that arise systemically in decision-making, steps the agency is taking to respond, and measures of these responses' efficacy.

3. A Role for Structural Reform Litigation

Our discussion of Congress's role rests on an asymmetric assumption: an agency on its own lacks sufficient motivation to adopt a set of good government reforms, but the promise of good government could spur Congress to act. The politics of decisional quality can fail, of course, and a systemic skew in error patterns may cause them to do so. A decade ago, errors in disability benefits adjudication probably skewed to favor claimants, a situation that invited scrutiny from congressional Republicans. ALJ allowance rates have since fallen. If anything, errors may now systematically disfavor claimants, a situation that will not trouble budget hawks or legislators skeptical of social welfare. If VLJ errors presently skew in veterans' favor, a political constituency for improved decisional quality probably will not materialize. To the contrary: "if you correct that and you start granting less," one official told us, "people will scream."³⁷³ The total dearth of meaningful quality assurance at EOIR suggests that, if errors skew at all, they probably disfavor

³⁷⁰ Gelbach & Marcus, *supra* note 33, at 126.

³⁷¹ On the potential of peer review, see Daniel E. Ho, *Does Peer Review Work? An Experiment of Experimentalism*, 69 STAN. L. REV. 1, 79-82 (2017).

³⁷² Carolyn J. Heinrich, *Outcomes-Based Performance Management in the Public Sector: Implications for Government Accountability and Effectiveness*, 62 PUB. ADMIN. REV. 712, 722 (2002).

³⁷³ Interview Subject #2, *supra* note 80, at 21.

immigrants.³⁷⁴ A powerful constituency exists to pressure for depressed relief, while immigrants with questionable claims to stay in the country have less political power. So long as immigrants do not enjoy the upsides of IJ error, EOIR will pay no political cost for its indifference to decisional quality.

Congress, VA, and veterans advocates may prefer a system riddled with error – erroneously, in our view.³⁷⁵ To them, a supposed crisis of decisional quality may be a win-win, with veterans enjoying what amounts to a *sub rosa* expansion of benefits schemes and only taxpayers suffering the downside. If errors skew to disfavor immigrants and disability benefits claimants, however, the lack of a political response may not reflect an accepted equilibrium among stakeholders. Rather, indifference to a crisis of decisional quality betrays the political impotence of one set of these stakeholders.

Canonical accounts of judicial review justify court involvement precisely under such circumstances.³⁷⁶ But judicial intervention to protect decisional quality prompts at least two concerns. The first involves the claims of critics like Mashaw and Metzger and Stack, that judicial review tends to harm the development of internal administrative law. But public law litigation of the sort a systemic challenge involving decisional accuracy would take can preserve room for internal agency governance within the confines of a court-imposed remedy. As William Simon and Charles Sabel argue, the prevailing approach to public law remedies has “move[d] away from command-and-control toward experimentalist methods.”³⁷⁷ Upon a determination of liability, the judge does not dictate remedial terms but rather “facilitate[s] a process of deliberation and negotiation among stakeholders,” including the defendant agency itself.³⁷⁸ The result, a “rolling-rule regime” that “incorporate[s] a process of reassessment and revision,”³⁷⁹ is not intended “to coerce obedience” but to “induce internal deliberation and external transparency.”³⁸⁰ A court could insist that an agency protect decisional quality, but it can let the agency, along with stakeholders, devise institutionally acceptable methods to do so.

A second concern involves the related problems of jurisdiction-channeling and cause of action. As described in Part II, a claim that an agency has taken inadequate steps to protect decisional quality ultimately implicates the merits of individual cases. As such, the claim may not be sufficiently “collateral” to evade jurisdiction-channeling provisions that would steer it to a forum disinclined to grant a systemic remedy.

One response is to recast the cause of action not as a claim effectively alleging a lot of individual errors but rather as something more obviously collateral to the merits of particular cases. The Information Quality Act, for example, obliges federal agencies to issue and abide by guidelines “for ensuring and maximizing the quality, objectivity, utility, and integrity of

³⁷⁴ TRAC Immigration, *Findings of Credible Fear Plummet Amid Widely Disparate Outcomes by Location and Judge* (July 30, 2018), <http://trac.syr.edu/immigration/reports/523/> (last visited Jan. 8, 2019).

³⁷⁵ Veterans advocates should demand protections for decisional quality, regardless of the direction of the skew in VLJ errors. Errors cause appeals and remands, and thus they add to the “hamster wheel” of delay in which veterans find themselves. Moreover, while perhaps more veterans erroneously receive relief than erroneously get rejected, the distribution of unwarranted largesse is arbitrary. In the end, poor quality decision-making simply creates delay and makes it more difficult for truly deserving veterans to get the benefits they are entitled to.

³⁷⁶ JOHN HART ELY, *DEMOCRACY AND DISTRUST* 135-79 (1980).

³⁷⁷ Charles F. Sabel and William H. Simon, *Destabilization Rights: How Public Law Litigation Succeeds*, 117 HARV. L. REV. 1015, 1052 (2004).

³⁷⁸ *Id.* at 1055; *id.* at 1062; *id.* at 1067-68.

³⁷⁹ *Id.* at 1068.

³⁸⁰ *Id.* at 1071.

information” the agency disseminates publicly.³⁸¹ Perhaps a plaintiff could challenge the integrity of something like BVA’s accuracy number and thereby force it to improve its quality review program. The question of whether the IQA provides a private right of action remains open, and thus far litigation to force the agency to improve its information disclosures has not fared well.³⁸² But no one has yet tested a theory of the sort we propose.

Another response builds on an important development in the CAVC, the jurisdictionally preferred forum for cases implicating the merits of VLJ decisions. In 2017, the Federal Circuit ruled that the CAVC has authority to adjudicate class actions challenging systemic shortcomings at BVA.³⁸³ The case at issue involved allegations of unconstitutional and arbitrary delays in VLJ decision-making, a theory that may present fewer obstacles to systemic reform than one rooted in quality concerns. But the decision at least recognizes the possibility that a reviewing court is not institutionally disabled from adjudicating a case involving much more than an effort to correct an error in an individual case.

A final response is for the federal courts to appreciate something they have recognized in other structural reform contexts involving constitutional rights. A systemic challenge to widespread agency dysfunction is not simply a bundle of individual claims alleging specific harms. When a prisoner brings a class action to challenge a state’s failure to provide adequate medical care at its facilities, he does not purport to litigate a lot of discrete harms to fellow inmates. Rather, he seeks to remedy the policies and practices that create a systemic risk of harm that all inmates bear.³⁸⁴ Likewise, a child in foster care challenging systemic mismanagement in a class action does not litigate the state’s failure to provide her with a competent caseworker. Rather, her individual experience is merely a representative manifestation of the state-wide risk of harm that mismanagement creates for all children in foster care.³⁸⁵ An equivalent doctrinal move would cast errors in individual cases not as individually fixable, but as illustrative of the harm that results from the systemic risk the agency creates when its overall decisional quality erodes.

CONCLUSION

BVA’s Quality Review story does not end happily. Given the little regard many external observers have for IJ decision-making, EOIR’s failure to implement even a rudimentary quality assurance program is shocking. SSA’s recent efforts are promising, but its decades of fitful experimentation and the persistent temptation to favor quantity over quality make rigorous critique by impartial outsiders all the more essential. Prompted by the right mix of external oversight and internal incentives, however, agencies may yet vindicate *Goldberg*’s commitment, not just for those individuals dogged enough to exercise procedural rights, but for everyone enmeshed in a system of mass adjudication. Two of us have helped to develop, test, and implement programs that have improved decisional quality without the commitment of significant additional

³⁸¹ 44 U.S.C. § 35016 note (2000).

³⁸² *E.g.*, *Zero Zone, Inc. v. U.S. Dep’t of Energy*, 832 F.3d 654, 678 n.25 (7th Cir. 2016).

³⁸³ *Monk v. Shulkin*, 855 F.3d 1312, 1318 (Fed. Cir. 2017).

³⁸⁴ *Parsons v. Ryan*, 754 F.3d 657, 676 (9th Cir. 2014).

³⁸⁵ *M.D. v. Perry*, 294 F.R.D. 7, 45 (S.D. Tex. 2013).

resources.³⁸⁶ There remain plenty of veins to tap for an agency willing to experiment and subject results to external evaluation.

A due process balancing test from the 1970s will not defuse the crisis of decisional quality agencies face as they buckle under the strain of large caseloads. Internal administrative law, properly shaped by external oversight and intervention, still might.

³⁸⁶ See *supra* p. 41-42 (describing BVA's short-lived experiment in 2016); Daniel E. Ho, *Does Peer Review Work? An Experiment of Experimentalism*, 69 STAN. L. REV. 1 (2017) (describing a peer review program and its promising impact on the quality of food safety inspections in King County, Washington).