

Improving courts' initial decisions: An experiment in a Mexican labor court

David S. Kaplan (Interamerican Development Bank), Joyce Sadka (Instituto Tecnol

May 26, 2014

Abstract

This paper reports on an experiment designed to improve the quality of the initial drafts of decisions made by a labor court in Mexico. A substantial proportion of final decisions made by this court are appealed successfully, causing large costs to the parties as well as to the court system. The question posed by this paper is whether a non-binding review process can substantially reduce errors and inconsistencies in the initial decision, in order to reduce the rate of successful appeals. The experiment consisted of detecting errors in all case files in the final decision process, and providing information on errors found to the law clerk responsible for the draft decision in 50% of these case files. Widespread and consistent errors were measured. Nevertheless we find that law clerks revise their decision in less than one quarter of the cases for which they receive observations indicating errors. We conclude that without a binding internal review process in this type of court, ex-post quality control of draft decisions has little chance to reduce successful appeals substantially.

1 Background and motivation

Labor law in Mexico is considered highly skewed in favor of workers, resulting in large costs to firms and inflexibility in the labor market. Little attention has been paid historically to the application of labor law in practice. In previous work we have analyzed data from one of the most important state level labor courts in Mexico, the Junta Local de Conciliacion y Arbitraje del Valle Cuautitlan Texcoco (JLCAVCT). The evidence gathered shows first that in practice labor law may be less favorable to workers due to very high litigation and enforcement costs. Second, it imposes large costs of litigation on defendant firms, and a large amount of uncertainty that may be especially relevant to the decisions of small to medium sized firms. Third, the administration of labor justice is extremely costly and inefficient. Among the specific reasons that labor justice is so costly is the rate of successful appeals that result in the court having to reopen and process case files that it closed many months earlier. Over the period 2002-2004 for example, the rate of appeals is approximately 30%. Of these appeals, close to 50% is granted by the appeals court. Generally a granted appeal will be remanded to the labor court for immediate correction, including scheduling new hearings, viewing new evidence, as well as rewriting the final decision. Given that the appeals process takes approximately 18 months, reopening the process and the decision and having to do so under time pressure implies large costs to the labor court, including the cost of deviating its attention from current cases and thus contributing to a large backlog in its work flow.

The motivation of this experiment was therefore to test the effectiveness of a low-cost quality control policy aimed at improving the quality of the court's decision, reducing successful appeals rates, and increasing the rate of enforcement. Instituting a formal and legally binding review process within the court would require reforming the code of labor law procedure, and was out of the question. Hence, we chose an administrative, non-binding process to review quality. The experiment was equivalent to have an additional «pair of eyes» to review one's decision in 50% of case files, with the results of this revision summarized in a simple and clear format. It is important to note that the experiment focused on basic errors, without questioning in any way the law clerk's proper exercise of discretion, for example in the evaluation of contradictory evidence viewed by the court. While the incorrect exercise of discretion could also give rise to successful appeals, the revision of such discretion by someone lower ranking than the law clerk would probably not

provide credible incentives for clerks to revise their decisions, whereas we believed that showing them obvious errors in information or clear violations of procedural or evidence rules in their decision had a significant chance of affecting their final decision.

This experiment provides evidence that is relevant for several important strands of literature. In the first place, while the literature on principal agent problems clearly suggests that costly monitoring of effort or costly quality control of production may help resolve the basic misalignment of incentives, there is little empirical evidence on exactly how such policies can be implemented and when in practice they are more or less effective. We provide evidence in a specific context that is generalizable to the administration of justice in developing countries.

Second, the literature on development has identified the quality of institutions as an important explanatory variable for outcomes in human and economic development. As a result the reform of institutions is an important goal in many countries, especially in the developing world. However, the impact of institutions on development outcomes is importantly determined by the performance of low and mid level bureaucrats who provide these institutions' services to the public. A generic example is labor justice: a labor reform may change specific and important elements of the costs of firing, by reducing indemnifications, increasing the relative incidence of fair dismissals, and so on. However, the impact of such changes will only affect labor markets if law clerks who write the initial decisions of labor courts apply the substantive and procedural provisions of the new law correctly. Without improving the performance of the low and mid level bureaucrat then, legal reforms may be far less useful than anticipated.

Third, there is a growing literature on the costs of attention and rational inattention. Essentially this literature posits that decision makers may make «mistakes» which are rational, as they are based on an implicit choice by the decision maker to collect limited information, due to the costs of information collection. In this context, revision of one's decision by a second pair of eyes may be beneficial, especially if this second pair of eyes is partially automated, lowering the costs of attention spent in the revision itself.

2 The experimental protocol

The methodology combined extensive data collection and quantitative techniques with knowledge of the labor law and its rules of procedure and evidence, to design a partially automated system for detecting errors and inconsistencies in the draft decision. Errors were described briefly in an observations sheet for each case file, and for 50% of the case files selected randomly, the observation sheet were given to the law clerk responsible for the file, suggesting that she revise her decision without compelling her to do so.

The experimental procedure was as follows. All case files entering the draft decision area of the court were scanned and then assigned by the office's administrator (a court bureaucrat) to a law clerk. While the clerk was writing the draft decision, detailed data was collected from the scanned case file on the names and addresses of plaintiffs and defendants, the worker's claims, facts alleged, evidence submitted, the defendant's response to all claims, and evidence submitted by the defendant. Information from court hearings including the admission and viewing of evidence was also captured. Once the clerk completed her decision, its was also scanned, and data was collected from the decision, in particular identification of the parties, claims, facts alleged, evidence submitted, admitted by the court, and viewed in hearings. Discrepancies between information in the case file and in the draft decisions were identified, and summarized in a one page «observations sheet». These sheets were produced for each case file in the decision process, but only given to the law clerk in 50% of the files, chosen randomly. Law clerks were told that the observation sheets were not legally binding but that they could benefit from paying attention to the errors or inconsistencies shown on the sheets in order to improve the quality of their initial decision. They then had the opportunity to revise their initial decision. All revised decisions were scanned and the same data was coded.

The experiment began in May 2012, with an initial period of around 6 weeks in which data was collected from draft decisions without any intervention. Starting June 26, 2012, 50% of draft decisions received observation sheets, and this continued until April 26, 2013. Over these 10 months, over 500 decisions were reviewed, and roughly half that number of observation sheets were provided to the law clerks.

3 Descriptive Statistics

Errors in the draft decisions can be separated into three major categories. First, errors in names and addresses of parties, which while merely clerical errors, can easily impede the enforcement of a judgment. Second, errors related to evidence presented, admitted, or viewed in the hearings of the case, which are closely related to a large proportion of the appeals. Third, we found errors in the quantification of the compensation to be received by plaintiffs who won their cases.

In the first place it is interesting to note that there is large variance across the 6 law clerks involved in the experiment, in the win rate for plaintiffs. Decisions are characterized by law as «in favor of the worker», «in favor of the employer», «mixed», which means the worker’s claims are only partially granted. In general a decision «in favor of the worker» results in a minority of cases; however, law clerks varied between 7% and 22% in this characterization of their decisions. This shows a large variance which could be a result of different criteria applied by the law clerks, or could be partially due to the differential impact of errors committed by each of them.

Errors in basic identification information of the parties is more common than we expected: 11% of draft decisions contained a discrepancy in the name of the defendant firm, and 7% contained a discrepancy in the address of the firm. In regards to evidence, we measured errors in two stages of the process: admission of proof and proper viewing of all admitted items of proof. Table 1 shows errors in the process of admission of proof. The rows «employer» and «worker» indicate evidence submitted by each of these parties. In most cases all items of evidence submitted are admitted by the court, and these are the observations in the first column. The rest of the columns detail what the court does in the cases in which it does not admit at least one item of evidence submitted by the party. According to the law, the court should provide legal reasoning, including law, jurisprudence, and their relationship to the specific item of evidence and case at hand, in order to reject an item of evidence. Not providing legal justification is an arbitrary act on the part of the court, which is a standard cause for an appeal on the grounds of a violation of due process. We show that in the case of plaintiffs’ evidence, slightly less than half the instances of rejected evidence are essentially not justified at all by the court, and for defendants this happens in more than half of all instances of inadmissible evidence. Notice that while this is an error committed by the court in hearings, which the law clerk who later writes the draft decision

cannot attend or control, this clerk has the prerogative to return case files to the labor courts and request corrections in procedural violations committed in hearings. However, in spite of this widespread violation of correct process, the law clerks hardly ever return cases to the courts to request corrections. Hence, they tend to carry over the procedural error committed in hearings.

Table 1: Errors in admission of evidence

Errors in unadmitted items of evidence.									
	All items of proof were admitted	%	No justification or unrelated statute/jurisprudence	%	Statute/jurisprudence without legal reasoning	%	Legal reasoning provided	%	Total
Employer	247	71.38%	53	15.31%	6	1.73%	40	11.56%	346
Worker	152	48.56%	57	18.21%	17	5.43%	87	27.79%	313

Table 2 shows errors in the draft decision with regard to the viewing of evidence which has been admitted in court. According to the procedural provisions of labor law, the draft should contain a detailed and accurate list of all viewed evidence, since this evidence must be considered in the final decision. The errors we found included not listing the evidence at all, as well as listing evidence that is inconsistent with the items of proof actually viewed in the hearings coded from the case file. We find a substantial proportion of cases with both types of errors.

Table 2: Errors in viewed evidence

# Draft decisions with errors in viewed evidence. N=625								
evidence viewed according to draft decision	No evidence submitted	%	Evidence mentioned but not listed	%	Discrepancy in viewed evidence	%	No discrepancies	%
worker	222	35.52%	85	13.60%	65	10.40%	253	40.48%
employer	88	14.08%	106	16.96%	80	12.80%	351	56.16%

Table 3 shows results on the third major type of errors in the draft decisions: incorrect quantification of the amount of compensation to be received by plaintiffs who win at trial. As shown, in more than 25% of the cases requiring a quantification by the law clerk, this calculation contains an error.

4 Initial Findings

We now investigate the relationship between four features of each case file: the exogenous characteristics of the plaintiff (gender, tenure, and daily salary), the number and types of items of proof admitted and viewed by the court for each party, the number of errors detected in the draft decision, and the

Table 3: Errors in quantification

	N= 478	Draft decisions with errors in quantification
No award to be quantified	210	43.93%
Incorrect quantification	114	23.84%
Quantification not possible due to imprecise facts	4	0.83%
Correct quantification	150	31.38%

final decision in favor of or against the plaintiff's claims. Table 4 shows the relationship between the presence of all items of proof and of particular categories of proof and the number of errors present overall in the draft decision. The main categories of evidence are confessions (depositions of parties to the case), testimony, documents provided to the court, items or documents provided through an inspection, and expert (technical) testimony. The presence of documents submitted by either party appear to have the only clear negative effect on the incidence of errors in the draft decision. This may indicate that the law clerks pay more attention to items of evidence when documents are provided by the parties, so that they make less mistakes in the draft decision. When considering the impact of each type of evidence alone, we also find that viewed confessions submitted by the plaintiff are negatively related to errors in the draft decision.

Table 5 relates errors in the draft decision to characteristics of the case, by regressing the presence of any error as well as the specific categories of errors discussed above on gender, tenure, and daily wage of the plaintiff. Few results are significant, with two notable exceptions. Having a female plaintiff is positively and significantly related to mistakes in the admission of evidence by the defendant, and plaintiff's daily salary is positively and significantly related to mistakes in the quantification of the amount to be awarded to that plaintiff.

Table 6 considers the relationship between the final decision, the evidence submitted and viewed for each party, and the characteristics of the plaintiff. The dependent variable is «decision favors the defendant». We find that tenure is highly significant and related to more favorable decisions for the plaintiff. Daily salary is unrelated to the win rate of plaintiffs, and female plaintiffs appear to do better (the coefficient is always negative and almost

Table 4: Existence of Any Error in Draft

	Existence of Any Error in Draft						
# Viewed Conf (p)	-0.0173 (-1.40)		-0.0261* (-2.36)				
# Viewed Testim (p)	-0.00651 (-0.26)			-0.0179 (-0.76)			
# Viewed Doc (p)	-0.0220* (-2.45)				-0.0210* (-2.38)		
# Viewed Insp (p)	-0.00446 (-0.86)					-0.00619 (-1.25)	
# Viewed Tech (p)	0.0204 (0.31)						0.0378 (0.56)
# Viewed Conf (d)		0.0256 (1.31)	0.0279 (1.38)				
# Viewed Testim (d)		-0.0351 (-1.31)		-0.0375 (-1.40)			
# Viewed Doc (d)		-0.0240 (-1.74)			-0.0320* (-2.46)		
# Viewed Insp (d)		-0.00715 (-0.59)				-0.00370 (-0.30)	
# Viewed Tech (d)		-0.0213 (-0.28)					-0.120 (-1.52)
_cons	0.852*** (37.64)	0.830*** (32.31)	0.824*** (31.13)	0.844*** (41.42)	0.839*** (41.78)	0.826*** (41.99)	0.819*** (41.66)
<i>N</i>	423	416	432	412	434	428	431

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Determinants of errors

Determinants of errors							
	Any Error	In Facts or Claims	Admission (p)	Admission (d)	Viewed Proof (d)	Viewed Proof (p)	Amount Awarded
Female Plaintiff	0.0585 (1.43)	0.0687 (1.31)	0.00446 (0.09)	0.108** (2.90)	-0.0125 (-0.23)	-0.0139 (-0.23)	-0.114 (-1.55)
Tenure	0.00306 (1.03)	0.00569 (1.50)	-0.00655 (-1.78)	0.000918 (0.34)	-0.00575 (-1.48)	-0.00663 (-1.44)	-0.00239 (-0.38)
Daily Salary	-0.0000262 (-0.56)	-0.0000244 (-0.41)	0.0000682 (1.18)	0.0000171 (0.41)	-0.0000262 (-0.41)	0.0000321 (0.45)	0.000256* (2.46)
_cons	0.791*** (22.88)	0.469*** (10.58)	0.378*** (8.81)	0.103** (3.29)	0.456*** (9.82)	0.454*** (8.45)	0.348*** (4.92)
N	434	434	434	434	392	301	181

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

always significant at the 90% confidence level). However, the effect of the characteristics on the final decision does not appear to be affected by the specific set of items of proof viewed by the court in the case, nor do these classifications of proof affect the final decision.

The picture we see of the evidence, errors, and characteristics seems to show little evidence of bias in the law clerk's decisions. Workers with more tenure winning more often can be easily explained by those with more years of experience with the firm having a stronger case and therefore submitting better quality evidence. Women are slightly but significantly more likely to settle a case in these labor courts, so that those women who go to court may have stronger cases. It is interesting to note that every single one of the law clerks writing draft decisions during the period of the experiment was female, and this raises the possibility of some positive bias towards women. Workers with higher salaries, though not more likely to win their cases, are more likely to have mistakes in the quantifications of their awards when they do win. Finally, though the mistakes in the draft decisions do not seem to be deliberate or to cause very clear bias, there are plenty of them, so that parties who can afford an appeal when they lose a case can easily find grounds for appeal in the court's decision. Thus, even without being convinced that errors stem from or cause bias, it is extremely important to reduce errors that provide grounds for excessive numbers of appeals.

5 Experimental impact and preliminary conclusions

The impact of the experiment can be measured in two ways. First, after law clerks received the «observation sheets» they sometimes would revise their draft decisions, and these revised drafts were scanned, coded, and compared to the original draft for that case file. Second, we plan to compare the rate of appeals and the rate of successful appeals, for cases that received observation sheets and cases that did not. However, this second and more important measurement is not yet possible since appeals take on average 18 months, and therefore the results of most appeals filed for court decisions emitted during the experimental period are not yet available.

Based on the first measurement, the experiment did not have the desired impact. Law clerks revised their decision after receiving an observations sheet

Table 6: Final Decision, Evidence, and Plaintiff Characteristics

	Final Decision Favors Defendant					
Tenure	-0.0120*** (-4.54)	-0.0120*** (-4.61)	-0.0123*** (-4.76)	-0.0118*** (-4.50)	-0.0121*** (-4.68)	-0.0125*** (-4.75)
Daily Salary	-0.0000361 (-0.88)	-0.0000341 (-0.82)	-0.0000279 (-0.68)	-0.0000393 (-0.96)	-0.0000315 (-0.78)	-0.0000336 (-0.83)
Female Plaintiff	-0.0602 (-1.65)	-0.0568 (-1.53)	-0.0565 (-1.57)	-0.0622 (-1.70)	-0.0571 (-1.58)	-0.0542 (-1.50)
# Viewed Conf (p)	-0.00374 (-0.34)		-0.00985 (-1.00)			
# Viewed Test (p)	-0.0288 (-1.26)			-0.0257 (-1.21)		
# Viewed Doc (p)	-0.00201 (-0.25)				-0.00284 (-0.36)	
# Viewed Insp (p)	0.00383 (0.83)					0.00202 (0.46)
# Viewed Tech (p)	-0.0422 (-0.73)					
# Viewed Conf (d)		0.00202 (0.11)	0.00334 (0.19)			
# Viewed Test (d)		-0.0424 (-1.70)		-0.0338 (-1.36)		
# Viewed Doc (d)		0.00198 (0.16)			0.00220 (0.19)	
# Viewed Insp (d)		-0.0101 (-0.94)				-0.0102 (-0.93)
# Viewed Tech (d)		0.0985 (1.46)				
_cons	0.245*** (7.70)	0.238*** (6.96)	0.241*** (7.07)	0.250*** (7.98)	0.229*** (7.58)	0.234*** (7.69)
<i>N</i>	416	410	425	406	427	421

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

in only 22% of the case files. Of these revisions, they only corrected clerical errors in names and addresses of parties, and errors in quantification, more than 75% of the time. Corrections of a more sophisticated nature, such as those related to the parties' claims and admitted or viewed evidence, were only undertaken between 20 and 40% of the times these errors were found and indicated to the law clerk. Therefore, even if corrections do turn out to be key in reducing appeals, their effect will be heavily diluted by the clerks' unwillingness to revise their draft decisions. On a positive note, however, the experimental procedure produced more than double the corrections that judges' review of the same draft decisions produced, during the same time period. In other words, while there were relatively few corrections in comparison to the number of corrections suggested, in absolute terms the experiment produced much more quality control than the usual minimal revision by the judges who sign the final decisions.

This leads us to two important conclusions. First, any quality control process has to have «bite» in the sense that the bureaucrats subject to the quality control must be obligated to take its results seriously and correct the errors in the task under review. This was not the case in our experiment, since the observation sheets were clearly identified as having no legal or administrative consequence for the clerks. Second, to perform quality control of a sophisticated task undertaken by a bureaucrat, it may be more useful to intervene earlier in the process and prevent mistakes rather than correct them, thus avoiding the costs of revisiting complex tasks. Based on these conclusions, we are currently testing the effects of providing law clerks with summaries of case file, previous to the writing of their draft decisions. The hope is that by providing key information from the case file, including information that would improve accuracy in the main areas of error we found in this experiment, we will be able to prevent more errors than we were able to correct through the ex-post quality control.