Measuring the Strength of Teachers' Unions: An Empirical Application of the Partial Independence Item Response Approach

Author(s): Katharine O. Strunk and Sean F. Reardon

Source: *Journal of Educational and Behavioral Statistics*, Vol. 35, No. 6 (December 2010), pp. 629-670

Published by: American Educational Research Association and American Statistical Association

Stable URL: https://www.jstor.org/stable/40959473

Accessed: 17-01-2020 19:04 UTC

Journal of Educational and Behavioral Statistics
December 2010, Vol. 35, No. 6, pp. 629–670
DOI: 10.3102/1076998609359790

Measuring the Strength of Teachers' Unions: An Empirical Application of the Partial Independence Item Response Approach

Katharine O. Strunk
University of Southern California

Sean F. Reardon
Stanford University

The literature on teachers' unions is relatively silent about the role of union strength in affecting important outcomes, due in large part to the difficulty in measuring union strength. In this article, we illustrate a method for obtaining valid, reliable, and replicable measures of union strength through the use of a Partial Independence Item Response (PIIR) model. This method uses the individual regulations found within collective bargaining agreements (CBAs) and models their existence as a function of a contract-specific latent level of restrictiveness. This allows for the estimation of (a) a measure of the latent level of relative union strength and associated standard errors of measurement; (b) the reliability of the measure; (c) item severities for each item addressed in the contract, allowing for assessment of item restrictiveness; and (d) contract and item information.

Keywords: *teachers' unions; union strength; IRT*

1. Introduction

Labor unions are a popular topic in the general media. The economic turmoil in the automobile industry has been attributed in part to the "strong" unions in the sector, and flagging educational outcomes have been ascribed to public school teachers' unions. Although many characteristics and actions are attributed to these labor unions, little is in fact known about their recent impacts on important market outcomes. Because labor unions were given the legal right to collectively bargain with employers through the 1947 amendment of the National Labor Relations Act, labor unions have exerted substantial amounts of control

We would like to thank Susanna Loeb, Steve Raudenbush, Tony Bryk, Katherine Masyn, Jamal Abedi, and two anonymous reviewers for their helpful comments. All remaining errors are our own.

over their own employers, with whom they collectively bargain over compensation, benefits, and work rules. Although union membership has declined in the private sector since its peak in the 1960s, they still remain an important force. Twenty-four percent of private transportation and utilities workers were unionized in 2005, as were 13.6% of private information workers and 13% of manufacturing and construction workers. Public sector union membership remains stable, representing 42% of public sector workers in 2005 (U.S. Department of Labor, Current Population Survey, 2005).

Given that unions affect so many workers and firms in the United States, questions regarding the cause and effects of these unions hold an important place in the labor economics and policy literatures. Teachers' unions, which represent 4.6 million teachers and paraprofessionals in U.S. public schools, specifically have become a popular area of research and discussion as politicians, researchers, and the public debate the appropriate role of unions in public education. Although the literature thoroughly explores the relationships between union members or unionization and various economic and social outcomes, it is largely silent about questions of the role of union strength in affecting these important outcomes. This is principally a result of the difficulty researchers face in measuring union strength. This article is motivated by an interest in empirically determining union strength relative to their employers so that we may assess the impact of unions, and union strength, on relevant outcomes. In it, we illustrate how to obtain a valid, reliable, and replicable measure of union strength with which researchers can explore the contexts in which strong unions thrive and the impacts of unions on relevant outcomes.

By extending the Partial Independence Item Response (PIIR) model developed by Reardon and Raudenbush (2006), we generate a probabilistically based measure of the inherent restrictiveness of collectively bargained union-management contracts that may serve as an intuitive and substantiated proxy for relative union strength. This method uses the individual regulations found within a contract and models their existence in union contracts as a function of a contract-specific latent level of restrictiveness. This allows for the measurement of both general and specific contract information, including (a) the latent level of restrictiveness or relative union strength inherent in each contract as well as associated standard errors of measurement; (b) the reliability of the measure of contract restrictiveness; (c) conditional and marginal item severities for each item addressed in the contract that allow for the ranking of item restrictiveness within the contract; and (d) total contract and specific item information. Our goal is to generate a reliable, valid, and usable measure of contract restrictiveness that can be applied in future research to assess the relationships between contract restrictiveness, important covariates, and relevant outcomes. Given that collective bargaining agreements (CBAs) are the tangible equilibrium outcomes of negotiations between unions and their employers, this

measure of contract restrictiveness can be considered a reasonable proxy for relative union strength.

This methodology is also applicable to problems that go beyond the measurement of union strength. Many questions in the fields of economics, politics, health, education, and sociology may be concerned with the level of flexibility of an organization under study. The method outlined in this article provides a tool with which researchers can examine institutional documents, like contracts or another set of organization policies, and from them extract a reliable, valid, and usable measure of the inherent level of flexibility, restrictiveness, or another underlying characteristic of the organization.

This article proceeds as follows. Section 1.1 addresses the methodological issues that face measures of union or contract strength used in the extant literature that leave a need for a more objective, informative, and reliable measure of union strength. Next, Section 1.2 outlines how we use the PIIR model to measure the inherent level of restrictiveness in the union contract, using the case of teachers' union contracts to explicate our method. Section 2 provides the empirical application of the PIIR model to California teachers' union contracts, explaining how we use the PIIR method to measure contract restrictiveness. Section 3 clarifies how we improve our measure of union strength to ensure that the measure is valid, reliable, and usable. Section 4 concludes with a discussion of the distinctive advantages of using the PIIR approach to model union strength.

1.1. Methodological Issues Facing Existing Measures of Unions, Union Strength, and Contract Restrictiveness

The majority of the work on labor unions focuses on the impact of labor unions on firm productivity and employee wages. For the most part, it has found that earnings and earnings inequality tend to be lower in industries and firms that are unionized or in which there is a high fraction of unionized employees (see, for example, Blau & Kahn, 1996; Freeman & Medoff, 1984). However, studies have found little difference in worker or firm productivity in unionized versus nonunionized industries or firms or in industries with a higher fraction of unionized employees (Clark, 1980; Pencavel, 1997). In addition, work on public sector unions has found that cities with higher proportions of unionized workers have higher total municipal employment (Benecki, 1987; Freeman & Valletta, 1988; Trejo, 1991; Zax, 1989), but that unionized public sector employees are not significantly less productive than nonunionized employees (Ehrenberg, Sherman, & Schwarz, 1983). The public sector union that has come under the most scrutiny in the literature is the teachers' union. Although there is agreement in the economics and politics of education literatures that teachers covered by teachers' unions' CBAs tend to earn more and have better working conditions than those who are not covered by union contracts, there is little consensus about the overall impact of

teachers' unions on student achievement (Baugh & Stone, 1982; Eberts, 2007; Eberts & Stone, 1987; Goldhaber, 2006; Hoxby, 1996; Kleiner & Petree, 1988).

Although this long and varied literature delves deeply into questions pertaining to the effects of union membership and union presence on various outcomes, important questions regarding how strong versus weak unions affect these outcomes have yet to be satisfactorily investigated. These questions surrounding the differential effects of union strength are more important now, when most employees in given firms or industries are unionized, than they were at earlier times, when whether an organization was unionized or the fraction of employees who belonged to a union was a more relevant policy issue. Today we understand that not all unions are created equal, and some unions can exert more influence on their employers than others. To determine how heterogeneous unions may differently affect important outcomes, we must determine a way to measure union strength.

Most of the extant literature does not address questions of union strength because researchers have had difficulties determining a measure with which to adequately gauge this attribute. Existing research addresses the measurement of union strength in a variety of ways. While each approach is useful and provides some interesting analyses, the collective literature that attempts to measure union strength has four main problems that have yet to be solved. First, much of the extant literature uses simple dichotomous indicators of unionization, effectively comparing unionized to nonunionized firms or industries, or in the case of the public sector, districts or municipalities (see, e.g., Blau & Kahn, 1996; Chambers, 1977; Eberts & Stone, 1987; Trejo, 1991). Other studies attempt to determine the effects of unions by measuring union strength through the use of proxies such as union membership, union size, or some more complex indicator of the extent of unionization (Baugh & Stone, 1982; Easton, 1988; Hoxby, 1996; Kleiner & Petree, 1988; Milkman, 1997; Steelman, Powell, & Carini, 2000). None of these measures of unionization or union strength adequately differentiate between unions to determine the relative strength or power that unions may exert on school districts to achieve specific outcomes.

More recently, researchers have turned to the CBAs (or contracts) negotiated between unions and management to generate indicators of union strength. In contract negotiations, unions bargain with management over working conditions, compensation, and benefits. These contracts contain myriad regulations that govern detailed aspects of workers' and employers' rights and responsibilities and in many ways are the most influential product of union strength. In effect, CBAs are the tangible equilibrium result of negotiations between unions and employers (Nash, 1953). It follows that contracts that result from negotiations in which unions are stronger than the management should theoretically contain more union-friendly outcomes, and contracts that result from negotiations in which firms are more powerful than unions should contain more employer-friendly

results. The contracts, then, if explored carefully, should signal the strength of the union relative to the employer (Eberts, 2007).

Because the large majority of these studies concentrate on the specific case of teachers' unions and district administrations, which is also the focus of our empirical application, we highlight the extant work that uses teachers' union contracts to understand teachers' union strength (Hess & Kelly, 2006; Hess & Loup, 2008; Koski & Horng, 2007; Levin, Mulhern, & Schunck, 2005; Moe, 2006, 2009). Although the use of aspects of teachers' union contracts to express union strength may allow for a more accurate measure of union strength than simple dichotomous indicators, these methods of assessing union contract strength suffer for three main reasons. First, many of them concentrate on only one or a few small areas of contracts, such as employee evaluation procedures, and use a limited number of individual contract provisions from within those areas to draw conclusions about the strength of contracts overall (Hess & Kelly, 2006; Koski & Horng, 2007; Levin et al., 2005; Moe, 2006; Woodbury, 1985). These single-aspect approaches to measuring union "strength" may isolate the effects of specific clauses or articles, but they lack content validity; by only covering content from a part of the contract, they do not capture union strength expressed across the entire contracts, which may lead to biased and invalid estimates of "union strength." This is especially important because, as the outcomes of negotiations between teachers' unions and school boards, CBAs include within them a series of trade-offs. Unions and boards will each give up provisions that are important to them to gain others that are more important. By focusing on only one or a few items or areas of a contract, researchers may only assess part of the trade-off, which will indicate a different level of strength than would have been assumed by viewing another area of the contract.

Second, no researchers to date have considered the amount of noise relative to true signal present in their measures of union strength that would bias any empirical results that stem from their use. Many researchers have attempted to generate their measures of contract or union strength by generating simple index variables that tally the number of items that exist in a contract or that subjectively assess the strength of contract provisions on an index scale (Hess & Kelly, 2006; Koski & Horng, 2007; Moe, 2006). Others "grade" contracts on their restrictiveness based on only a small number of contract provisions deemed the most potentially harmful to district flexibility (Hess & Loup, 2008). Eberts and Stone (1987) use the number of items in teachers' union-district contracts to measure union strength. This research assumes that the contract variables are measured without error and makes no effort to ascertain or adjust for the reliability of their measures.

Third, many of these measures are not replicable by the larger research community either because they are based on a subjective understanding of "restrictive" that may be particular to a given researcher or because the specific manner in which the measure is created is not explained in a way that allows for

replication. The adequacy of a measure of union strength matters very little if it is not transparent enough for exploration and replication by other researchers.

A notable exception to this literature is the recent work by Terry Moe (2009) in which he addresses many of the considerations posed above by using a number of items from a breadth of areas in school-district contracts to generate a measure of the latent level of contract restrictiveness. He uses a factor analytic approach to measure the latent level of restrictiveness toward districts in teachers' union-district contracts. This article, which is the only one in the literature to use a statistically based approach to measuring restrictiveness, avoids the problem of subjectively "grading" contracts or items faced by many researchers who have attempted to use contracts to capture union strength or contract restrictiveness. However, Moe still subjectively chooses the contract items he includes in the model, selecting those that he believes are "potentially important to the everyday operation of schools and typical of the kinds of restrictions unions fight for" (Moe, 2009, p. 6). Similar to earlier work in the field, Moe's recent paper also fails to provide an estimate of his measure's reliability or measurement error.

This study attempts to address these methodological problems found in the labor union literature by using a method that allows for the unbiased and statistically based examination of labor union-management contracts. We use many regulations from within teachers' union contracts to assess the latent level of relative union strength expressed in teachers' union contracts. As we will explain in greater detail in the following section, we refrain from subjectively selecting items for inclusion by defining, a priori, the complete set of contract provisions without interpretations of restrictiveness or flexibility. Furthermore, we use accepted statistical rationales to winnow down a set of 334 contract items to a set of 39 that are sufficient to generate a valid, reliable, and useful measure of contract restrictiveness. This method of selecting items for inclusion in the final model of contract restrictiveness allows us to refrain from making subjective judgments about including items that "best" reflect contract restrictiveness and provides a replicable model for use in later research. The PIIR method also allows us to generate a probabilistically based measure of union strength as well as acknowledge the amount of measurement error that exists in our measure of contract restrictiveness and improve the reliability of our measure. All of this enables us to generate a reliable, valid, and usable measure of contract restrictiveness as well as to determine which items within contracts are the most and least restrictive to the school district based on a probabilistic, objective measurement model.

1.2. Using a PIIR Model to Measure Union Strength

The rationale for examining union-management CBAs for indicators of union strength (relative to the management teams with whom they bargain) rests in the knowledge that these contracts result from negotiations in which both the union and the management bargain for the inclusion of policies that are meant to serve

their constituents. For example, the union may try to negotiate for higher pay, which benefits the workers, and the management team may try to include stricter evaluation policies, which serve the management team. Because the CBA is simply a collection of all the policies attained by both groups, it contains within it a measure of latent union strength relative to the management.

The important methodological question, then, is how best to use the information provided within these contracts to measure the latent strength characteristic. Rather than selecting specific items to include in a series of dichotomous or index measures that are intended to proxy for union strength, we generate a statistically sound, probabilistic measure of the latent relative union strength characteristic expressed in labor union-management contracts through the use of a one-dimensional item response theory (IRT) Rasch model adjusted to account for the conditional structure of contract regulations.

IRT has traditionally been used to construct measures of examinees' latent cognitive skill levels (Hambelton, Swaminathan, & Rogers, 1991; Rasch, 1980; Wright & Masters, 1982). IRT assumes that characteristics of a test item (item j), such as its difficulty (γ_j), interact with an examinee's (examinee i) ability, or latent trait (θ_i), to determine the probability of a correct response to that particular item such that examinees of a given ability level will correctly answer more difficult items less frequently than easier items. The Rasch model, the most simple IRT model, predicts the probability that a randomly chosen examinee with ability θ_i will answer test item j correctly given item j's level of difficulty, γ_j. In the end, IRT models produce measures of the difficulty of each item, the reliability of the scale/test, the ability of each examinee, and a standard error of measurement for each examinee. The IRT model uses the data from the item responses to generate an interval scale along which every item and every examinee can be located (Wright & Stone, 1979).

Social scientists have begun to adapt IRT methods to non-testing contexts such as the analysis of survey responses to measure latent levels of behaviors, assuming that an individual's latent behavioral characteristic is expressed through his or her survey responses (Cheong & Raudenbush, 2000; Osgood, McMorris, & Potenza, 2002; Raudenbush, Johnson, & Sampson, 2003; Raudenbush & Sampson, 1999; Reardon & Raudenbush, 2006). This approach enables researchers to construct meaningful interval scales for the latent behaviors measured by the surveys. For instance, Cheong and Raudenbush (2000) examine how a behavioral rating scale measures different behavioral problems in children. Rather than interpret γ_i as item difficulty, the rating scale of Cheong and Raudenbush incorporates items that reflect varying levels of severity in underlying problem behaviors, including aggressiveness, delinquency, and anxiety.

This study differs from social research that has used Rasch or other IRT models to measure latent constructs because, instead of using survey data, we examine institutional documents (teachers' union contracts) and instead of measuring some person ability or trait, we measure an institutional construct:

contract restrictiveness defined as relative union strength. In this way, we conceive of a contract as similar to a test, consisting of a set of items that, if endorsed, restrict districts' freedom to take certain actions. Particularly strong unions will succeed in compelling districts to agree to endorse more of the items. In this sense, a contract is like a test with many items, and union strength is analogous to student ability. This conception of union contracts suggests the use of IRT to estimate latent union strength just as we use IRT to estimate latent ability of students.

The problem with using standard IRT models to measure a latent trait captured by contracts is that the items in union contracts are not conditionally independent of each other. IRT models assume that the probability of an examinee answering one item in the affirmative is not predicated on his or her answering any other items in the affirmative (items are locally independent). However, contracts, like many surveys, are not structured in such a way as to make this assumption plausible: many contract regulations are conditionally structured such that certain items in contracts may be present only if another item also exists within the contract, immediately violating IRT's conditional independence assumption.

Reardon and Raudenbush (2006) address this issue of partial independence by acknowledging that many surveys contain "gate items" and "skip patterns," causing values of some items to be strictly determined by positive or negative responses to other items. They propose a generalized hybrid of a proportional odds hazard model and a Rasch model: a model for item responses that allows item responses to be conditional on prior responses while being independent of responses to other items. Similar to other latent trait models, the model by Reardon and Raudenbush (2006) produces four important values, this time incorporating conditional as well as marginal probabilities and severities. They are (a) person-specific, interval-scale measures of a latent traits; (b) estimates of conditional and marginal item severity for each item in the survey; (c) person-specific conditional and marginal probabilities of an affirmative response to each item in a protocol; and (d) measures of item information and total test information.

Applying the PIIR model of Reardon and Raudenbush (2006) to teachers' union contracts allows us to measure the inherent restrictiveness or flexibility toward districts in teachers' union contracts along with each contract's standard error of measurement, while also measuring the "severity" or "restrictiveness" of each separate regulation within the contract. This partial independence model also enables us to evaluate the amount of measurement error that exists in our measure of union strength. Because contract items are not necessarily locally independent, we depart from the traditional test format by our use of the PIIR model, which allows for the relaxation of the local independence assumption so that we can model certain items within a contract as conditional on a preceding item.

Measuring the Restrictiveness of California Teachers' Union Contracts

In this article, we focus on measuring union strength by measuring the restrictiveness of the contracts negotiated between teachers' unions and district school boards in California. In California, one of the 35 states in the United States that allows K–12 public school teachers to collectively bargain, contracts are negotiated approximately once every 3 years between representatives from each district's local teachers' union and from the local school boards. The teachers' union bargains on behalf of all the teachers in its local district, and the school board negotiates on behalf of the students and the district's own interests. The contracts that result from the union-district negotiations contain policies that regulate virtually every aspect of a teacher's work life, including guidelines regarding how teachers can be evaluated, compensated, and assigned to specific positions, as well as about class sizes, permissible leave time, grievance procedures, and the precise rights of the teachers' union.

2.1. Sample and the Generation of Contract Data

This article uses data from 466 California public school-district teachers' union contracts.[1] Of the 565 districts with four or more schools in California, 466 provided copies of their contracts in response to researchers' request, for a response rate of 82.5%.[2] To apply the PIIR model of Reardon and Raudenbush (2006) to contracts, we treat the contracts as "responses" to "surveys" or "tests" that ask districts about the negotiated CBAs. The treatment of contracts as similar to surveys or tests requires some translation. First, contracts differ from surveys or tests because they are not constructed with a specified list of regulations and rules to cover. Tests include a set number of questions that test takers are expected to answer—every test taker answers the same set of questions. At first glance, union-management contracts do not have this feature. The "items" contained in a contract are generated through negotiations between the teachers' union and the school board. Because each district has different actors who choose to negotiate over different items, and exist in different contexts, each district contract may have markedly different regulations and clauses in it. Not every contract has a clause relating to every issue. Nonetheless, the absence of a given restriction within a contract implies a negative answer to the question "does the contract restrict the district with regard to X?" Given this, we can treat the contracts as a whole as having a defined set of items over which the two parties bargain. The entire list of possible negotiable items can then be seen as the complete set of questions—the survey or test. Put differently, each contract can be read as an answer sheet to some large survey or test with the contract regulations as the responses to the questions.

We developed the "survey" or test that might be "answered" by a district and union during negotiations organically from the contracts themselves. During the reading of the first 100 contracts, we recorded every possible aspect of a contract; anything that was written in the contract became a "question" to be included on the contract test. As we read more and more of these first 100 contracts, broad categories of clauses and specific areas of provisions within each of these broader clauses became evident, and we began to outline conditionally structured item strings. For example, district-union contracts address the issue of "Association" (teachers' union) rights. Among other things, this area of the contract addresses concerns regarding whether members of the Association can take leave for union business, the amount of time members can take, and the number of members who are allowed to take such leave time. When reading the section of the Association rights areas of contracts regarding Association leave, we developed a "survey" that first asked whether the contract promised specific Association leave. If it did (the contract "answered" in the affirmative), then five follow-up "questions" were asked of the survey, including: "Does the Association president receive additional "Association leave" time off?" If the "answer" again was in the affirmative, then the next follow-up questions could be asked, including, for example, "Does the contract specify the total number of days of Association release time the president receives per year?" Only if the answer again is "yes" will the contract be asked "Does the president receive more than 10 days each year?" and then "Does the president receive more than 20 days each year?" And so forth. If at any point, the answer to a question was "no," the following string of items would not be asked of the contract.

We refined the "survey" by excluding items that were already dictated by the California State Education Code, Government Code, or state regulations, which are broad and highly regulatory (Brewer & Smith, 2007). Much of the language included in district contracts simply reiterates laws over which the district administrations and local teachers' unions have no control. We only included items in the contract test if the district contracts included "extra" nonregulated language in addition to the standard or required language.[3]

After completing a thorough examination of the 100 preliminary and randomly selected contracts, we had developed a survey instrument consisting of a total of 639 "questions," or items asked of the contracts as a whole. This method of item selection avoids one of the fundamental problems with previous contract-based measures by ensuring that we do not choose our items based on subjective assessments of what we believe should indicate union strength or affect specific outcomes. We then further reduced the number of items included in the final instrument based on the ability of the items to discriminate among contracts. Because our IRT measure is built on a conditionally structured framework in which the conditional probabilities of responding "yes" to a given item provides the most information the closer they are to .5, we manipulated the items from within the larger instrument to create a tighter, more usable set of

334 binary items, which offer more meaningful responses and indicate whether a contract's "response" to a specific question sets that contract at or above a given threshold point. Teachers' union-district CBAs can be broadly classified into four areas: benefits, working conditions, evaluations and grievances, and Association rights. We include 140 items in the benefits area, 120 in working conditions, 46 in evaluations and grievances, and 28 items in the Association rights area.

All of the full survey questions were "asked" of each contract. The dichotomous response to these questions can be considered a yes/no response, similar to a respondent answering a survey or a test question. As an example, let us consider the possible questions asked of a contract regarding Association (union) rights.

1. Does the contract promise specific leave that union members may take for Association business (Association leave)?
2. How many union representatives can take Association leave?
3. How many days can each union representative have each year for Association leave?
4. How many total days each year can union representatives take for Association leave?
5. How much of Association leave costs does the district pay?
6. Does the union president receive (additional) Association leave time?
7. How much time off does the Association president receive?
8. How much of the costs associated with the president's Association leave does the district pay?
9. Does the contract specify that the union receives any particular rights above and beyond those specified by the California Education Code?
10. On which policy areas must the district administration consult with the Association?

Some of these questions elicit nonbinary responses. To apply the PIIR model, we convert these 10 questions into 28 binary items, each of which indicates whether a contract is at or above a specific threshold level for each question. We use thresholds generated by examining the marginal probabilities of a given response to each question and setting cutoff levels with the aim of ensuring that the conditional probability of a district answering in the affirmative rests between 0.20 and 0.80.[4] Because the amount of information provided by an item j is a function of the probability of being asked item j and the conditional probability of responding in the affirmative to item j, setting the threshold levels around these cut points allows for the most differentiation across the range of responses.

Table 1 outlines the 28 binary items regarding Association rights that were constructed from the 10 broader questions above and shows the observed item response frequencies from the sample of 466 district contracts. We use this method to construct binary items for all areas of the contracts. The data used

TABLE 1
Observed Association Rights Item Response Frequencies

	Item Description	Gate Items	Marginal Probability	Risk Set	# Yes	Conditional Probability
1	Do any union members receive specific leave that they may take for Association business (Association leave)?	None	.843	466	393	.843
2	Does the contract address the number of union members who may take Association leave?	1	.092	393	43	.109
3	Does the contract provide a specific number rather than rely on the "reasonable number" designation from Government Code § 3543.1?	1, 2	.067	43	31	.721
4	Does the contract address the number of days of Association leave that may be taken by each designated union representative?	1	.103	393	48	.122
5	Can each designated union representative take 5 or more days of Association leave per year?	1, 4	.079	48	37	.771
6	Does the contract specify the *maximum* number of days of Association leave all union members can collectively take each year?	1	.569	393	265	.674
7	Does the Association receive, in total, 10 or more days of Association leave each year?	1, 6	.461	265	215	.811
8	Does the Association receive, in total, 20 or more days of Association leave each year?	1, 6, 7	.292	215	136	.633
9	Does the Association receive, in total, 50 or more days of Association leave each year?	1, 6, 7, 8	.103	136	48	.353

(continued)

TABLE 1 *(continued)*

	Item Description	Gate Items	Marginal Probability	Risk Set	# Yes	Conditional Probability
10	Does the district pay for any of the costs for Association leave?	1	.425	393	198	.504
11	Does the district cover all costs related to Association leave?	1, 10	.099	198	46	.232
12	Does the union president receive (additional) Association leave time?	1	.513	393	239	.608
13	Does the district pay for any of the costs of the union president's Association leave?	1, 12	.378	239	176	.736
14	Does the district cover all costs related to the union president's Association leave?	1, 12, 13	.167	176	78	.443
15	Does the contract address the number of days of Association leave that the union president may take each year?	1, 12	.470	239	219	.916
16	Does the union president receive 10 or more days of Association leave each year?	1, 12, 15	.378	219	176	.804
17	Does the union president receive 20 or more days of Association leave each year?	1, 12, 15, 16	.288	176	134	.761
18	Does the union president receive 40 or more days of Association leave each year?	1, 12, 15, 16, 17	.195	134	91	.679
19	Does the union president receive full-time Association leave each year he or she is president?	1, 12, 15, 16, 17, 18	.109	91	51	.560

(continued)

TABLE 1 *(continued)*

	Item Description	Gate Items	Marginal Probability	Risk Set	# Yes	Conditional Probability
20	Does the union receive any particular rights beyond those identified by the California Education Code? (i.e., not including facilities, equipment, access to member mailboxes)	None	.212	466	99	.212
21	Must each school principal consult with the union school representative?	None	.058	466	27	.058
22	Must a union representative and a representative from the administration meet and consult at least once each month?	None	.077	466	36	.077
23	Must the administration consult with the union on more policies than those guaranteed in California Government Code § 3453.2 (e.g., curriculum and content, educational objectives)?	None	.161	466	75	.161
24	Must the administration consult with the union on budgetary and fiscal policies?	23	.058	75	27	.360
25	Must the administration consult with the union on the school calendar or schedule?	23	.054	75	25	.333
26	Must the administration consult with the union on employment and personnel policies?	23	.039	75	18	.240
27	Must the administration consult with the union on staff development policies?	23	.041	75	19	.253
28	Must the administration consult with the union on any policies that may affect union members?	23	.015	75	7	.093

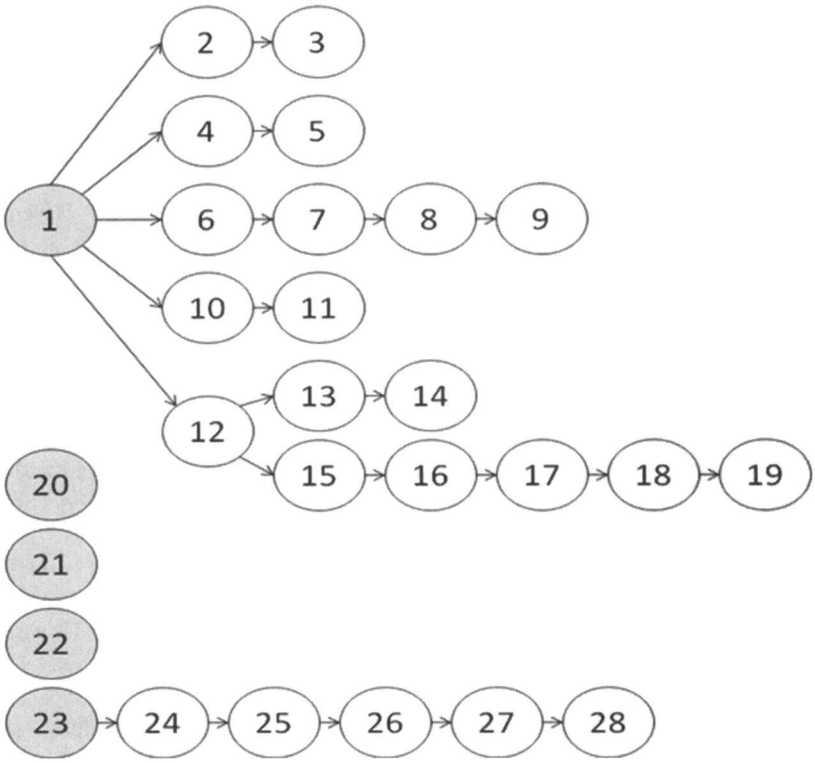

FIGURE 1. *Association rights provisions item structure.*

in this article are generated from the "responses" from each district to these sets of questions. Additional terms in Table 1 are explained in Section 2.2 below.

2.2 Applying the PIIR Model to California Teachers' Union Contracts

The PIIR Model is based on the assumption that there is some single underlying level of restrictiveness, denoted θ_i^*, inherent in all teachers' union contracts. This underlying level of restrictiveness affects the probability that a contract will contain any given contract provision. This is the same set of assumptions that underlies any simple IRT approach to test making. In this way, contracts are like tests. However, unlike a simple algebra test, contracts are conditionally structured. In an algebra test, each question has nothing to do with the other questions—a test taker can answer Question 9 right even if he or she answers Question 8 incorrectly. A contract, however, is often not structured in this way. There are some items that can exist only if the contract contains

a previous item; the existence of one item is conditional on the existence of another. It is this conditional item structure that the Reardon and Raudenbush (2006) PIIR model, by combining IRT with a hazard model, is designed to accommodate.

Figure 1 provides a schematic outline of the conditional item structure for the 28 binary items we constructed for the contracts' Association rights provisions. The questions associated with each item can be referenced to Table 1. Based on the conditional item structure for this and other sets of provisions, we define a gate matrix as described in Reardon and Raudenbush (2006). The gate matrix identifies the set of gate items for each contract item, where a "gate item" in a teachers' union contract is an item that must be answered in the affirmative (a regulation must exist) for some logically subsequent item to be asked of the contract.[5]

More detail on the conditional item structure of the Association leave items, as well as the marginal and conditional probabilities that each item is included in a contract, is described in Table 1. Eighty-four percent of teachers' union contracts specify that union members are allowed time off for Association leave (Item 1). If the only interesting question regarding Association leave was simply "does this contract address Association leave," the use of an expanded Rasch model with its independence assumption would be sufficient. However, there are additional interesting questions associated with Association leave. For example, "Does the contract specify the maximum number of days of Association leave? Are members allowed 10 or more days of Association leave each year? Twenty or more days? Fifty or more days?" (Items 6 through 9). The next questions are follow-ups to the first: A contract must first respond in the affirmative that it addresses Association leave at all to be asked whether the CBA specifies the maximum number of days union members receive for Association leave each year. In turn, to create a distinction that the contract specifies that union members receive at least 10 days each year, the contract must have responded in the affirmative that it specifies the maximum number of days for Association leave. This first question, then, "Does the contract address Association leave," is a gate question for the second, third, fourth, and fifth questions, and the second question, "Does the CBA specify the maximum number of days of Association leave each year?" is a gate question for the third through fifth questions, and so on.

In this case, the risk set—the set of districts with contracts that could feasibly have a regulation, given the existence of the gate regulation—are those districts that have contracts that do address Association leave. Table 1 indicates how many contracts are asked each item, given the gate structure, as well as which items in the Association leave provisions are gate items for subsequent items.

Model notation. The formal PIIR model is a simple random- or fixed-effects model, with items nested within contracts. Following Reardon and Raudenbush

(2006), we model the probability of contract i containing a given item k as a function of the latent level of contract i's restrictiveness, θ_i, and a vector of dummy variables indicating that a given response refers to a specific item in the contract. The coefficients on the dummy variables, γ_j, are interpreted as the conditional severity or restrictiveness of each contract item. We fit the contract-specific latent restrictiveness measure as randomly varying across districts; the conditional severity parameters of each item are assumed constant across contracts.[6]

Formally, let Y_{ik} denote the outcome (0,1) of each item k, and ϕ_{ik} equal the conditional probability of an affirmative response to item k for contract i, conditional on passing the gate item, h_{ik}, such that $h_{ik} = 1$.

$$\phi_{ik} = \Pr(Y_{ik} = 1 | h_{ik} = 1). \tag{1}$$

The structural model is

$$\log\left[\frac{\phi_{ik}}{1 - \phi_{ik}}\right] = \theta_i + \sum_{j=1}^{K} \gamma_j D_{ij}, \tag{2}$$

where θ_i is assumed normally distributed with a mean equal to zero and variance equal to τ.

Fitting the model with contract data. To fit the PIIR model described above, the data must be structured into a contract-item data set such that there exists one observation for each item k for each contract i where $h_{ik} = 1$. In other words, there is a line of data for each item asked of each contract, except for items whose nonexistence is logically determined by a negative response to a logically prior gate item. Every contract is at least asked the first-order gate items (those gate items that are not themselves reliant on an affirmative response to any item). In the 334 item contract, 116 (or approximately one third) are first-order gate items. For example, there are five first-order gate items in the Association rights example depicted in Table 1 and Figure 1 (shaded circles).

Following Reardon and Raudenbush (2006), we fit a random-effects logit model to these data using the EM algorithm with Laplace approximation to the likelihood (Raudenbush, Yang, & Yosef, 2000) using the software package HLM6 (Raudenbush, Bryk, Cheong, & Congdon, 2005).

Obtaining contract-specific estimates of latent union strength. We fit the random-effects model and obtain an empirical Bayes contract-specific posterior mean ($\theta^*_{i,334}$) and variance ($V^*_{i,334}$).[7] We label the estimates with a "334" to indicate the number of items included in the model. Figure 2 shows the empirical distribution of the contract restrictiveness measure, $\theta^*_{i,334}$. Overlaid on the distribution is the estimated distribution of contracts, normally distributed with variance $\hat{\tau}$, and the kernel density curve to compare the proximity of the distribution $\theta^*_{i,334}$ to a normal distribution. The estimated trait is fairly normally distributed

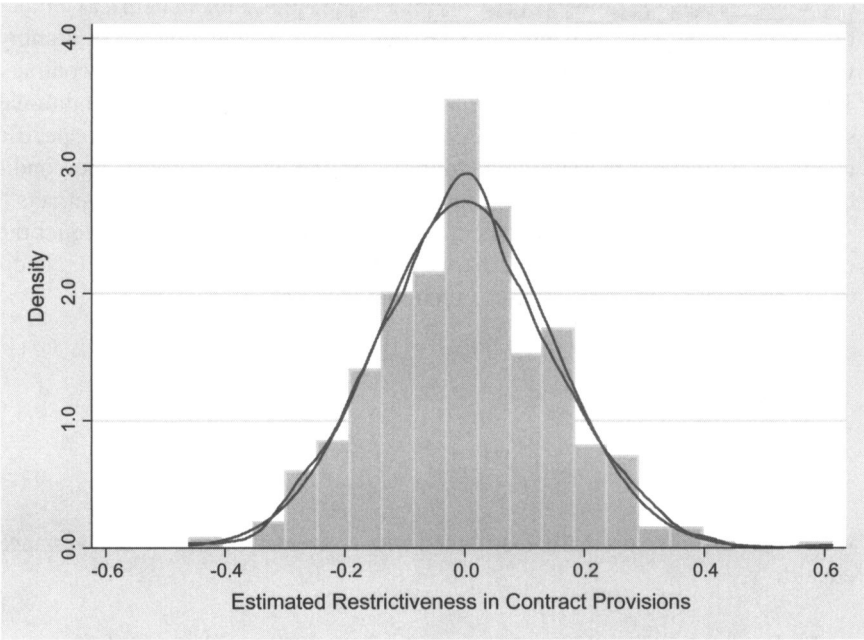

FIGURE 2. *Distribution of estimated contract restrictiveness, 334 item measure (empirical Bayes estimate).*

with a mean at 0.0003 (at zero by construction), indicating that there is a range of response patterns among the contracts. The measure of contract restrictiveness, $\theta^*_{i,334}$, has a median of −0.006, with a standard deviation of 0.146 and a range from −0.460 to 0.612. The empirical Bayes estimation strategy used in the modeling of $\theta^*_{i,334}$ produces a shrunken estimator, here shrunk to the grand mean. This inherently accounts for some of the measurement error in the constructed variable, allowing for the use of $\theta^*_{i,334}$ in the modeling of relationships between the estimated contract restrictiveness, other covariates of interest, and relevant outcomes—including, in the case of teachers' union contracts, district distribution of resources, and student outcomes.

Understanding the reliability of the measure of union strength. One of the most important advantages to using a Rasch model to determine the underlying level of union strength in contracts is that it allows us to evaluate the reliability of our measure. The reliability, λ, provides the estimated ratio of true signal (variance of θs) to signal-plus-noise (variance of θ plus measurement error) in the measure of union strength and indicates the level to which we are able to differentiate between individual districts according to their estimated union

strength "scores." The model fit with 334 items and 466 contracts/districts has a moderate reliability of $\lambda = .572$. This is a lower reliability than we might expect given a test of 334 items, in part because of the relatively low sample size of contracts (466) and in part because the 334 items do not each provide independent information; because of the "skip patterns" induced by the conditional structure of the items, not all contracts provide information on each of the items. It is also an indication, however, that our measure of union strength obtained from examining the 334 contract items included in our PIIR model captures a fair amount of noise.

It is worthwhile to note that all measures of contract restrictiveness or union strength, whether generated through the PIIR approach or another method, will have some level of measurement error. The benefit of the PIIR model is that we are able to make explicit the amount of measurement error in our measure of contract restrictiveness, making clear any problems or strengths that this measure has over others. Researchers can then do with it as they wish, as opposed to more simplistic measures of contract restrictiveness or union strength, which provide no information about the extent to which they capture the trait as opposed to noise. Although it is possible given this reliability to compare districts in areas of the range and to use $\theta^*_{i,334}$ in statistical models, we would hesitate to compare individual districts along the scale $\theta^*_{i,334}$, given the moderately low reliability of the measure.

3. Improving the Measure of Contract Restrictiveness

If we continue to consider the contract as similar to a test, with a number of items that are equivalent to test questions and with an underlying trait—in this case union strength—that can be assessed through the existence of specific items in the contract, we can then gauge the usefulness of the contract in generating a measure of union strength as we would judge the usefulness of an algebra test in generating a measure of algebra ability. There are three indicators of a good test. First and second, tests should produce valid and reliable measures of the latent trait they are intended to measure. This is also the case in our example—assessing the existence of specific items in district contracts should allow us to obtain a reliable and valid measure of union strength. The third and important goal of a test instrument is to be "usable"—the instrument should be easily administered and replicated (Linn & Miller, 2005). The measure of contract restrictiveness obtained from the contracts using the PIIR model meets the validity goal based on its high degree of content validity—it uses an objective and complete set of contract items to measure the degree of restrictiveness toward the district administration inherent in the contract. However, our measure has higher measurement error than is generally useful (as evidenced by the reliability of .572) and is not particularly "usable" due to its length. With 334 items, researchers will be required to read contracts front-to-back and code each of the items included, making our measure extremely difficult to

reproduce. In this section, we explore ways to improve upon the PIIR-generated measure of contract restrictiveness to retain its validity while improving upon its reliability and usability.

3.1. Assessing the Reliability of $\theta^*_{i,334}$

Given the moderate model reliability of .572, one may worry that the measure is capturing more than one latent trait. Although we have no test that can confirm the unidimensionality of the measure, we can test for specific forms of multidimensionality that are predicted by existing theory (see Reardon & Raudenbush, 2006); if we find no evidence of these types of multidimensionality, we may be more confident that the measure captures a single latent trait of interest. For example, we may look to the theory behind collective bargaining to identify and test for a form of multidimensionality in our measure. The earlier Association rights example illustrated how each item in a contract can be seen as falling along a single spectrum that ranges from flexible, or not restrictive to the district (or restrictive to the teachers' union) to restrictive to the district (flexible to the teachers' union). However, it is possible that there is more than just an underlying level of restrictiveness inherent in each contract but rather that there are two competing latent traits inherent in contracts. For example, districts may use their contracts as vehicles to attract teachers to and keep them in their schools. Although including items intended to recruit and retain teachers may restrict the district, such regulations may reflect some latent level of district need or willingness to compromise more than an underlying level of union strength. To test that we are measuring only one restrictiveness construct, we separate the contract into two parts: Those broad aspects of contracts that teachers are likely to consider and be made aware before taking a job with the district and those aspects of the contract that are likely never read by the average teacher in the district and are most likely considered less important to the district for reasons of recruitment and retention. The idea is that contracts may be more restrictive to the union and more flexible to the district in those less visible areas that teachers likely do not consider in their employment decisions, whereas the district administration might allow for increased restrictiveness to the district and flexibility to the union in those areas that teachers are more likely to consider. We perform a simple analysis to determine how highly and positively correlated these two subgroups of contract items are. A strong positive correlation would suggest that we are measuring a single latent trait. A negative correlation, however, might imply that in fact we are measuring two latent constructs—district need or willingness to compromise as well as union strength.

The first "visible" group of contract items includes those that have to do with compensation and benefits, transfer policies, Association rights, and class size policies. We assume that potential teachers may consider these very important district policies before joining a district, thus giving the district an incentive to

trade off restrictiveness for attractive working conditions for teachers. In the other group of "less visible" contract items are policies surrounding grievances, layoffs, leaves, miscellaneous working conditions (such as whether the district provides free on-site parking or a staff room for its teachers), preparation time, professional growth, evaluations, and school days and hours. It is likely that teachers pay less attention to these items at the outset when they are considering district employment.

We run a similar model to that outlined in Equation 2, except that now we have two random effects, $\theta^*_{i,\text{visible}}$ and $\theta^*_{i,\text{less visible}}$. By running this model, we get two estimates of variance, one for each of the random effects, and we can judge the true correlation of these random effects to determine whether there exist one or two latent traits within the contract data. We find that the two measures are highly correlated: the estimated correlation between the two random effects is .745, suggesting that both the "visible" and the "less visible" measures are in fact measuring the same underlying trait. Although this confirms the measure's unidimensionality with respect to this specific grouping of items, it does not discount other forms of multdimensionality. For instance, one can imagine that we may wish to test for greater dimensionality, as it is possible that district administrations and unions "trade off" restrictiveness in one part of the contract in exchange for leniency in another. However, there is little theory to guide us in a test of which areas of a contract districts and unions may wish to trade off—each different district context and history will likely lead to different negotiations and result in different provisions included in the CBAs. Moreover, as Reardon and Raudenbush (2006) note, tests such as these have relatively low power and so would require considerably larger samples than we have available here.

3.2. Improving Reliability and Usability: Item Analysis and Selection

The next step is to assess the 334 contract items to ensure that they are all contributing to the measurement of the underlying trait. When generating tests, test makers begin with a large item bank of relevant test items and winnow the number of items down to a reasonable set that produces a reliable measure of the underlying trait the test is intended to assess. By testing the instrument on test takers, they determine which items truly capture the latent level of ability and which do not. Identifying any misfitting items allows those items adding only noise to the analysis to be removed from the scale. The intent is to choose items from a long test or an item bank that maximizes Cronbach's α for the constructed test while maintaining the validity of the test (Armstrong, Jones, & Wang, 1998). Given that our measure of contract restrictiveness suffers from both a usability problem—the instrument is too long to be easily replicable—and a reliability problem—the reliability of the contract is a moderate .572—we follow classical test theorists and slim down

the number of items in our contract instrument to improve construct reliability as well as usability.

Most other researchers who generate measures of union strength based on contract items subjectively choose items that they believe should exhibit restrictiveness toward the district. They do not assess the internal consistency of their measures or the degree of measurement error in their constructs. Keeping consistent with our goal of generating an objective measure of relative union strength, we base our item selection on the unbiased statistical methods used in test construction. We run an exploratory Cronbach's α analysis on all 334 items included in our initial model. We examine the item-total correlations produced for each of the 334 items. A low item-total correlation statistic indicates that the individual item fails to measure that which the other items measure and should be discarded. We follow the generally accepted standard used by test makers and drop items with item-total correlations lower than .25 (J. Abedi, personal communication, January 15, 2009). This is an iterative process; after we drop items with low item-total correlations, we reassess and remove any further items that have item-total correlations below .25 based on the new scale with fewer items. After three iterations of this process, there are no more items with particularly low item-total correlations, and we are left with an instrument of 39 items that span the breadth of contract.[8]

We fit the PIIR model with this reduced number of items to generate a new measure of union strength. We find that the reliability of the measure improves substantially, yielding a new reliability of .677. It is especially interesting to note that the reliability of our measure is increased as we discard 88% of the contract items because reliability naturally increases with the number of items in a scale. This indicates that the 295 discarded contract items were in fact capturing more noise than underlying trait.

3.3. Assessing the Validity of the Union Strength Measure

It is important to remember that a measure's reliability and a test's usability are worth little if the measure is not valid. As such, it is worthwhile for us to question whether dropping 88% of the items from our instrument causes the loss of too much content, invalidating the measure and causing it to no longer measure the underlying construct in which we are interested, union strength. Table 2 outlines the items that remain in our model. It shows that there are still important items left in each of the four broad areas of the contract. An examination of these remaining items shows that they appear to capture contract restrictiveness and the relative strength of the union, exhibiting a high degree of content validity. For instance, the items concerning teacher compensation that remain in the model (Items 1 through 5 in Table 2) assess the extent to which districts must pay teachers for obtaining doctorate degrees and for remaining in the district for very long

TABLE 2
Contract Items Included in Final Measure of Contract Restrictiveness

	Contract Items Description	Gate Items
Benefits		
Compensation		
1	Do teachers receive a bonus for having a doctorate degree?	None
2	Is this bonus over $1,000 per year?	1
3	Is this bonus over $2,000 per year?	1, 2
4	Do teachers receive longevity bonuses (bonuses for experience in the district) of over $4,000 per year?	None
5	Do teachers receive longevity bonuses (bonuses for experience in the district) of over $7,000 per year?	4
Leaves		
6	Are teachers allowed more "bereavement" leave than is promised in the California Education Code?	None
7	Are teachers specifically allowed more leave time for bereavement leave for which travel is not required?	6
8	Do teachers receive more leave time for family illness/family care leave than they are entitled through the Federal Family and Med Leave Act of 1992 and the California Family Rights Act of 1991 and AB109?	None
9	Is this extra leave for family illness/family care fully paid for by the district?	8
10	Do teachers receive extra parenting/child rearing leave?	None
11	Do teachers receive sabbatical leave?	None
Working conditions		
Transfer and vacancies		
12	Does the contract prioritize transfer rights for specific teachers?	None

(continued)

TABLE 2 *(continued)*

	Contract Items Description	Gate Items
13	Do teachers who apply for a voluntary transfer *not* receive priority in their transfer choices over those who are to be involuntarily transferred?	12
14	Do teachers who will be involuntarily transferred receive priority in their transfer choices over those who have applied for voluntary transfers?	12
15	Do teachers who are involuntarily transferred for "non-cause" reasons retain the right to return to that site or position if a spot opens?	None
16	Is seniority a factor in deciding who is involuntarily transferred?	None
17	Is seniority the deciding factor in deciding who is involuntarily transferred?	16
18	Is there a limit on the frequency with which teachers can be involuntarily transferred?	None
19	Does the contract outline the ways in which involuntarily transferred teachers can establish their preferences to fill vacancies?	None
Class size		
20	Does the contract address class size?	None
21	Does the contract specify a given class size?	20
22	Must districts balance classroom sizes within a specific time period from the beginning of the school year or semester?	20
23	Must districts balance classroom sizes within 4 weeks of the beginning of the school year or semester?	20, 22
School day		
24	Does the contract specify the length of the school day in instructional minutes?	None
Evaluations and grievance procedures		
25	Can tenured teachers who receive a satisfactory evaluation use a nonstandard evaluation procedure on their next evaluation cycle? (e.g., less frequent evaluations, self-evaluations)	None

(continued)

TABLE 2 *(continued)*

	Contract Items Description	Gate Items
26	Does the school board not have final, binding decision-making rights in grievance hearings?	None
27	Is the school board not allowed to participate in the grievance process at all?	26
28	Do grievances go to arbitration?	None
29	Are decisions about grievances made in arbitration hearings binding on the district?	28
Association rights		
30	Do any union members receive specific leave that they may take for Association business (Association leave)?	None
31	Does the contract specify the maximum number of days of Association leave all union members can collectively take each year?	30
32	Does the district pay for any of the costs for Association leave?	30
33	Does the union president receive (additional) Association leave time off?	30
34	Does the district pay for any of the costs of the union president's Association leave?	30, 33
35	Does the contract address the number of days of Association leave that the union president may take each year?	30, 33
36	Does the union president receive 10 or more days of Association leave each year?	30, 33, 35
37	Does the union president receive 20 or more days of Association leave each year?	30, 33, 35, 36
38	Does the union president receive 40 or more days of Association leave each year?	30, 33, 35, 36, 37
39	Does the union president receive full-time Association leave each year he or she is president?	30, 33, 35, 36, 37, 38

periods of time. This reflects what we know about teachers' unions; they argue for increased levels of compensation based on education and experience whereas district administrators and school boards often wish to compensate teachers based on measurable indicators of teacher quality.

Unions also argue for longer leaves for their members, and Table 2 shows that stronger unions are those that successfully negotiate more and longer leaves into the contract. Also in accordance with what we suspect about union preferences in negotiations, Table 2 indicates that stronger unions negotiate for the rights of involuntarily transferred teachers (those who are reassigned to positions within the district against their will) and for teachers with the most seniority in the district. The remaining 39 items also cover restrictions on class sizes and the length of teachers' instructional school days, as well as important evaluation benefits for tenured teachers. Union theory would also posit that unions fight for more union-friendly procedures in the hearing of grievances, including preventing the school board from making the final and binding rulings over teachers' complaints (grievances) and ensuring that teachers' complaints are heard by an impartial arbitrator. Table 2 reveals that our measure includes items that address these issues. Finally, Table 2 shows that our measure includes important rights for the union (Association). Stronger unions with more restrictive contracts argue for Association leave for their members, and specifically for their presidents, and for longer leave allowances for the union presidents. In addition, more restrictive contracts require the district to pay for at least part of the Association leave time off.

We also test for construct-related evidence of validity. Recent research has shown that teachers' unions are often stronger in larger districts (Moe, 2009; Rose & Sonstelie, 2010). If our measure truly captures union strength, then it should be positively correlated with district size. We run a simple bivariate correlation and find that our measure of union strength is positively and significantly correlated with the natural log of district size, with a correlation of .52 (disattenutated correlation is .64) and $p < .01$.

3.4. Comparing the Union Strength Measures With 39 and 334 Items

This new measure of contract restrictiveness generated from 39 contract items meets the tests of validity, reliability, and usability. Table 3 compares the original measure ($\theta^*_{i,334}$) that used 334 items and the newer measure ($\theta^*_{i,39}$) that uses only 39 items. The new measure has a higher standard deviation and a greater range, indicating that the measure is spread across the range of values of union strength. Figure 3 shows that the estimated trait based on 39 items is fairly normally distributed, with a mean at −0.004 (zero by construction) and a median of 0.053. The new measure of union strength has a standard deviation of 0.508. We see that there is still a wide range of response patterns among contracts even with this smaller number of contract items. The

TABLE 3
Union Strength Measure Comparison

	Mean	Median	*SD*	Range	λ	Correlation (θ^*_i, District Size)	Correlation ($\theta^*_{i,334}$, $\theta^*_{i,39}$)
$\theta^*_{i,334}$	0.0003	−0.0059	0.1464	(−0.4601, 0.6121)	.572	.25	.718
$\theta^*_{i,39}$	−0.0037	0.0533	0.5079	(−1.6367, 1.6551)	.677	.52	

Note: All correlation statistics are statistically significant with $p < .01$.

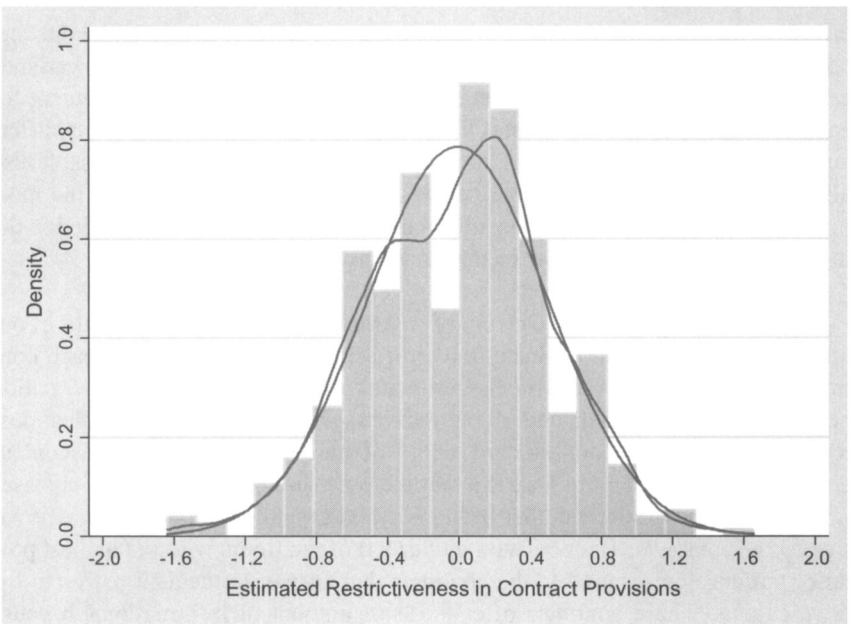

FIGURE 3. *Distribution of estimated contract restrictiveness, 39-item measure (empirical Bayes estimate).*

two measures are highly correlated, with corr($\theta^*_{i,334}$, $\theta^*_{i,39}$) = .718. Table 3 also shows that the second, more reliable measure of union strength, $\theta^*_{i,39}$, has greater construct validity as it has a higher correlation with district size than does the first measure of union strength.

4. Discussion and Conclusion

4.1. Distinctive Properties of the PIIR Measure of Union Strength

Using the PIIR model to generate a measure of contract restrictiveness provides us with a number of advantages over other measures of contract restrictiveness and union strength used in the literature. As discussed in Sections 2 and 3, our measure meets reliability, validity, and usability criteria used in the assessment of testing instruments. Not only can we generate a probabilistically based measure of relative union strength, but we are also able to assess the degree of measurement error, which falls at acceptable standards with a reliability of .677. In addition, our measure is a valid indicator of union strength, meeting both content- and construct-validity standards. The smaller set of 39 items also makes the measure easier to generate and replicate, meeting the important "usability" standard for construct creation. The PIIR approach provides two additional benefits over more traditional methods used to create measures of union strength and contract restrictiveness. First, we can use the empirical Bayes posterior variance estimates to determine the contract-specific standard errors of measurement for each of our contracts. This enables us to determine how precisely we may differentiate between contracts across the range of estimates of restrictiveness. It also allows us to determine the range of $\theta^*_{i,39}$ over which we can obtain the most information. Second, our method allows us to examine each item included in the model and rank them in order of their restrictiveness to the district.

Contract-specific standard errors of measurement. One of the strengths of our approach is that we can evaluate how precisely our model estimates each contracts' latent trait. The most precise estimates allow for the most differentiation between contracts in their underlying restrictiveness and have the smallest posterior variance (V^*_i) and standard errors of measurement (posterior standard deviation, $(V^*_i)^{1/2}$). Figure 4 shows that the precision of the estimates increases (V^*_i decreases) with the number of items asked of and answered in the affirmative by the contracts. The points on the far left of the figure with the highest posterior variance correspond to the contracts that answer in the affirmative to the fewest items. These contracts offer the least amount of information,[9] because they are asked relatively few questions and as a result are the least precise. Figure 5 indicates where along the distribution of $\theta^*_{i,39}$ the estimates are the most precise, with the lowest $(V^*_i)^{1/2}$. Figure 5 shows that the most precise estimates can be found in the middle of the distribution of estimated restrictiveness but appear fairly precise across the distribution of $\theta^*_{i,39}$.

Item severity: Determining the most and least restrictive contract items. Of particular interest to researchers who wish to understand the substance included in contracts, the PIIR model allows for the assessment of which of the items included in our measure are the most or least restrictive. Under PIIR model

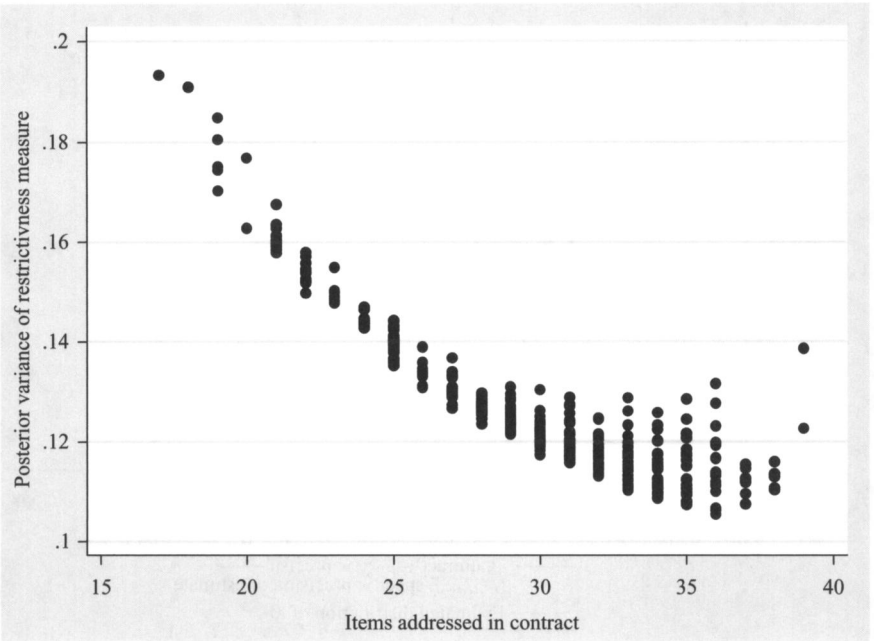

FIGURE 4. *Precision of estimated contract restrictiveness, 39-item measure, by number of items answered.*

assumptions, item severity (restrictiveness) is related to the frequency of the item being answered affirmatively. In the present case of district contracts, the most severe items are those items that have the fewest affirmative responses—when regulations are the rarest. For the most part, contract regulations fit these assumptions: If a regulation does not often occur in contracts in the California sample, it is likely because it is particularly restrictive.[10]

The conditional item severities for each item in the model are simply the negative of the estimated coefficients on the item indicator variables in each random effects model. The conditional severity of an item, item k, is the value of $\theta^*_{i,39}$ at which a contract would have a .5 probability of responding affirmatively to that item, or including the corresponding regulation, given that the contract is in the risk set for item k. Although these item severities are interesting in that they tell us the likelihood of a contract containing a given regulation conditional on it containing the gate regulation, of greater interest are the marginal severities associated with each item. The marginal severity of each item k is the value of $\theta^*_{i,39}$ at which a contract would have a .5 probability of "answering" item k affirmatively, given the conditional item severities and the gate matrix. In other words, these marginal item severities indicate the probability of a contract

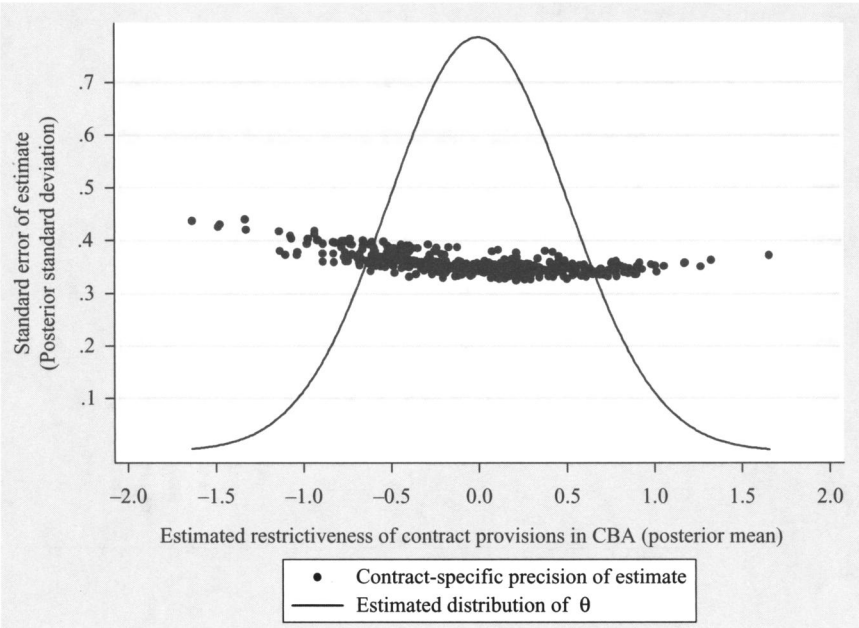

FIGURE 5. *Precision of estimated contract restrictiveness, 39-item measure.*

including a given regulation without conditioning on the containment of previous regulations (Reardon & Raudenbush, 2006). Table 4 reports the estimated conditional and marginal severities of each item included in the PIIR model.

The conditional item severities do not necessarily provide an ordered ranking of item restrictiveness because they depend on the item structure (the gate matrix) and so are often calculated conditional on different gate items. The marginal item severities provide a better approximation of an item severity ranking. We can estimate the "severity" or restrictiveness of each item as an approximation of how rare it is to find that item in a contract. As an example, we compare Items 8 and 14. Item 8 has an estimated marginal severity of 1.247, lower than the estimated marginal severity of Item 14, at 1.511. Comparing the marginal item severities tells us that requiring the district to allow involuntarily transferred teachers priority over voluntarily transferring teachers in their assignment choices is substantially more restrictive to the district than requiring the district to provide extra family illness/family care leave time.[11]

Figure 6 illustrates the fitted marginal probability curves for each contract item with the estimated distribution of $\theta^*_{i,39}$ overlaid. The order of the marginal probability curves provides an estimation of the severity ranking of the items in the contract restrictiveness measure. The most restrictive item, Item 3, is shown

TABLE 4
Estimated Conditional and Marginal Severities for Contract Items

Item	Description	Estimated Conditional Severity	Standard Error	Estimated Marginal Severity	Restrictiveness Rank
1	Do teachers receive a bonus for having a doctorate degree?	0.228	−0.109	0.228	23
2	Is this bonus over $1,000 per year?	0.115	−0.155	1.054	9
3	Is this bonus over $2,000 per year?	1.332	−0.236	2.031	1
4	Do teachers receive longevity bonuses (bonuses for experience in the district) of over $4,000 per year?	−0.007	−0.106	−0.007	26
5	Do teachers receive longevity bonuses (bonuses for experience in the district) of over $7,000 per year?	−0.606	−0.155	0.606	16
6	Are teachers allowed more "bereavement" leave than is promised in the California Education Code?	0.115	−0.107	0.115	25
7	Are teachers specifically allowed more leave time for bereavement leave for which travel is not required?	−0.459	−0.156	0.738	15
8	Do teachers receive more leave time for family illness/family care leave than they are entitled through the Federal Family and Med Leave Act of 1992 and the California Family Rights Act of 1991 and AB109?	1.247	−0.121	1.247	7
9	Is this extra leave for family illness/family care fully paid for by the district?	0.974	−0.223	1.999	2
10	Do teachers receive extra parenting/child rearing leave?	−0.819	−0.114	−0.819	31
11	Do teachers receive sabbatical leave?	−1.111	−0.119	−1.111	32
12	Does the contract prioritize transfer rights for specific teachers?	0.477	−0.108	0.477	17

(continued)

TABLE 4 *(continued)*

Item	Description	Estimated Conditional Severity	Standard Error	Estimated Marginal Severity	Restrictiveness Rank
13	Do teachers who apply for a voluntary transfer *not* receive priority in their transfer choices over those who are to be involuntarily transferred?	−0.249	−0.200	1.042	10
14	Do teachers who will be involuntarily transferred receive priority in their transfer choices over those who have applied for voluntary transfers?	0.767	−0.198	1.511	6
15	Do teachers who are involuntarily transferred for "non-cause" reasons retain the right to return to that site or position if a spot opens?	1.558	−0.136	1.558	5
16	Is seniority a factor in deciding who is involuntarily transferred?	−1.111	−0.118	−1.111	33
17	Is seniority the deciding factor in deciding who is involuntarily transferred?	−1.019	−0.136	−0.183	28
18	Is there a limit on the frequency with which teachers can be involuntarily transferred?	0.887	−0.116	0.887	12
19	Does the contract outline the ways in which involuntarily transferred teachers can establish their preferences to fill vacancies?	0.323	−0.107	0.323	21
20	Does the contract address class size?	−3.231	−0.244	−3.231	39
21	Does the contract specify a given class size?	−1.996	−0.156	−1.600	35
22	Must districts balance classroom sizes within a specific time period from the beginning of the school year or semester?	0.712	−0.112	0.750	14

(continued)

TABLE 4 *(continued)*

Item	Description	Estimated Conditional Severity	Standard Error	Estimated Marginal Severity	Restrictiveness Rank
23	Must districts balance classroom sizes within 4 weeks of the beginning of the school year or semester?	−1.09	−0.208	0.995	11
24	Does the contract specify the length of the school day in instructional minutes?	0.370	−0.106	0.370	20
25	Can tenured teachers who receive a satisfactory evaluation use a nonstandard evaluation procedure on their next evaluation cycle? (e.g., less frequent evaluations, self-evaluations)	1.618	−0.135	1.618	3
26	Does the school board not have final, binding decision-making rights in grievance hearings?	−1.812	−0.192	−1.812	36
27	Is the school board not allowed to participate in the grievance process at all?	−1.372	−0.14	−0.694	30
28	Do grievances go to arbitration?	−2.524	−0.207	−2.524	38
29	Are decisions about grievances made in arbitration hearings binding on the district?	−1.740	−0.178	−1.197	34
30	Do any union members receive specific leave that they may take for Association business (Association leave)?	−1.812	−0.144	−1.812	37
31	Does the contract specify the maximum number of days of Association leave all union members can collectively take each year?	−0.685	−0.138	−0.256	29
32	Does the district pay for any of the costs for Association leave?	0.082	−0.122	0.320	22

(continued)

TABLE 4 *(continued)*

Item	Description	Estimated Conditional Severity	Standard Error	Estimated Marginal Severity	Restrictiveness Rank
33	Does the union president receive (additional) Association leave time off?	−0.376	−0.208	−0.034	27
34	Does the district pay for any of the costs of the union president's Association leave?	−0.863	−0.163	0.462	19
35	Does the contract address the number of days of Association leave that the union president may take each year?	−0.150	−0.195	0.137	24
36	Does the union president receive 10 or more days of Association leave each year?	−1.247	−0.195	0.469	18
37	Does the union president receive 20 or more days of Association leave each year?	−0.896	−0.196	0.791	13
38	Does the union president receive 40 or more days of Association leave each year?	−0.363	−0.208	1.159	8
39	Does the union president receive full-time Association leave each year he or she is president?	0.266	−0.235	1.600	4

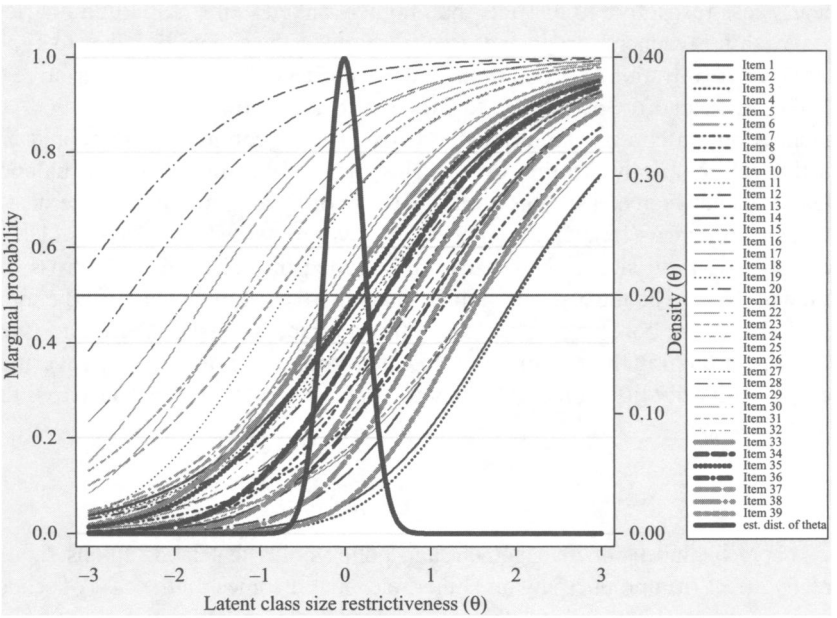

FIGURE 6. *Estimated marginal probabilities, by contract restrictiveness.*

by the black dotted line on the far right of the distribution of $\theta^*_{i,39}$ in Figure 6. The least restrictive item, Item 20, is shown by the dash-dot line to the far left of the distribution of $\theta^*_{i,39}$. The right-hand column of Table 4 ranks the 39 items included in the measure of contract restrictiveness from most (ranked 1) to least (ranked 39) restrictive.

A quick examination of Table 4 and Figure 6 shows that the items that are deemed most restrictive to the district make intuitive sense. For example, the contract item that most restricts the district, Item 3, requires the district to provide over $2,000 per year to teachers who hold a doctorate degree. As discussed earlier, teachers' unions want teacher compensation to be based on experience and educational credits, whereas districts more often wish to base compensation on expressed merit, in terms of either student achievement gains or career ladder advancement. Requiring districts to provide relatively large bonuses to teachers with a specific degree, regardless of the quality of the teacher, appears to restrict the district. The second most restrictive item of the 39, Item 9, requires districts to fully compensate teachers for extra leave time taken to care for their families. This means that the district will have to pay for a substitute teacher as well as for the regular teacher when a teacher takes more time off due to a family illness than is mandated by state and federal laws. Although the 2 least restrictive items of the 39, Items 28 and 20, still restrict district administrations, they are

clearly less restrictive to districts than higher ranked items. Although districts may wish for teachers' grievances to go straight to the school board or to avoid third-party arbitration, it is common, and therefore not particularly restrictive to a given district, for grievances to proceed to arbitration. Similarly, 96% of teachers' union contracts address class size. This is common practice and not particularly restrictive to the district. However, requiring districts to balance class sizes between classrooms within 4 weeks of the beginning of the semester or the school year (Item 23) is fairly restrictive to the district, ranking 11th in severity. Figure 6 shows that items are included that discriminate across the range of $\theta^*_{i,39}$. The most severe items only truly discriminate between the most restrictive of the contracts, as so few of the contracts get asked these questions and answer them in the affirmative. However, the spread of marginal severities across the distribution of indicates that the items provide information across the range of $\theta^*_{i,39}$.

4.2. Conclusion

Theory behind labor management and policy holds that labor unions should greatly affect firm productivity and other relevant outcomes. In the case of school districts and teachers' unions, theory suggests that teachers' unions will affect important district and student outcomes. Previous research on unions has for the most part focused on dichotomous indicators of the existence of unions or of unionization, which, although interesting in and of themselves, lack practical relevance when considering how unions with heterogenous levels of power or influence differentially affect specific outcomes. Studies that have attempted to estimate the relationship between union strength, a more relevant measure, and important outcomes face problems due to the difficulty in measuring union strength; proxies for union strength such as size serve only as loose and untested surrogate measures. Additional problems are introduced as researchers turn to the study of union contracts to give some indication of union strength due to difficulties that arise as researchers use simple index measures of union strength that provide no understanding of the amount of uncertainty contained in the measure, are based on subjective interpretations of strength or restrictiveness, and use only one or a few aspects of contracts to reflect the level of constraint inherent in the entire contract.

In response to the methodological troubles facing researchers who wish to explore questions of union strength, this article provides an empirical example of the PIIR model applied to the study of teachers' union contracts. Our method of item selection and use of the PIIR approach has a number of advantages over more traditional ways of measuring union strength. First, our method provides an objective and statistically sound approach to measuring union strength. We begin with a large set of contract items that are selected objectively to provide the maximum amount of information about the level of restrictiveness inherent in teachers' union contracts. We then select items based on the accepted statistically

based standards used in test construction to maximize the ratio of true signal to noise in our measure. We also ensure that our contract instrument used to generate the measure of union strength addresses the range of important content covered in union contracts.

Second, the use of the PIIR model allows for the creation of a transparent and probabilistically based interval scale along which individual contracts are placed according to their specific level of restrictiveness, as well as standard errors of measurement for each contract. We are also able to assess the reliability of our measure and judge that it captures the underlying trait well enough to use in statistical analyses of the relationships between union strength and covariates and outcomes of importance.

Third, and especially helpful to researchers who are substantively interested in the content of teachers' union contracts, our approach allows for the examination of contract content included in the generation of the measure of contract restrictiveness. This is a nice by-product of the PIIR approach to modeling contract restrictiveness; not only do we obtain a reliable, valid, and usable measure of union strength but we are also able to examine the item difficulties to determine which items in a contract are the least and the most restrictive to districts and to unions and to gain a better understanding of the bargaining process and outcomes.

Although using a latent trait measurement approach to model the underlying level of restrictiveness exhibited by an institutional document such as a contract may seem unorthodox, we perceive it to be a plausible and reproducible answer to many of the issues plaguing extant research on teachers' unions and their collectively bargained contracts. There is room to grow with the application of latent trait models to the measurement of organizational characteristics. For example, future endeavors might examine the possibility of using a two-parameter IRT model that captures items discriminatory power as well severity, allowing for another tool for objective item assessment. It is our hope that given the historical difficulty researchers have had in understanding the union-management negotiations process and measuring the results of this process, the approach exemplified in this study may provide a roadmap for further research using latent trait modeling approaches in union bargaining environments.

Notes

1. We are indebted to Prof. William Koski and to Eileen Horng for providing us with the unique data resource of California school district contracts, without which this article would not have been possible.

2. We use as the greater population of districts those with four or more schools for two main reasons. First, contract regulations may not matter in districts with fewer than four schools. For example, it may not matter which teachers are given preference in transfer decisions if there is are only three schools in the district, an elementary school, a middle school, and a high school. Second, contracts can

serve a markedly different purpose in very small districts with less need for negotiation and bureaucratic processes. Small districts with relatively few teachers and administrators may be more able to make informal policy decisions without the need for highly regulated contracts.

3. It is for this reason that measures generated from the length of the contract or the number of regulations in a contract likely fail to accurately measure union strength. For example, 5 of the 39 items included in Moe's (2009) contract restrictiveness measure are simply counts of the number of pages devoted to various areas of the contract (assignments, voluntary transfers, involuntary transfers and surplusing, teacher evaluation, and parent complaints). Because many aspects of these contract areas are covered in the California Education and Government Codes, not all districts include them in their contracts. Length measures more likely indicate district efficiency, the quality of the district or union's lawyers, or some level of distrust between the union and the district management than they do union strength and contract restrictiveness.

4. We do not always meet this goal of ensuring conditional probabilities between .20 and .80, because in some cases the conditional probabilities of gate items or of other important items fall outside of this range.

5. For a more detailed explanation of the generation and use of gate matrices and contract-item data sets from these question-item strings, see Reardon and Raudenbush (2006).

6. Although we could fit the model as a wholly fixed-effects model, we choose a random- and fixed-effects model, following Reardon and Raudenbush (2006), because it is more efficient, allows for θ_i to be modeled as a function of a set of district-level covariates and because it enables us to test for multidimensionality by estimating more than one latent trait and examining the tau correlations between the potential dimensions.

7. See Reardon and Raudenbush (2006) for the mathematical structure of these models.

8. It is important to note that this method of item selection may lead us to discard some good items along with many poor items. This is due to the correlated errors in the model that result from the conditional item structure. Nonetheless, our goal is to develop a reliable measure that uses a modest number of items, rather than to identify all possible well-fitting items. For this purpose, the Cronbach's α in the α-if-item-deleted statistics provides us with a very useful heuristic to winnow the items included in our measure down to a reliable and usable set. To ensure the appropriateness of our remaining 39 items, we also perform two other sets of item selection analyses. First, we remove items from the larger scale by dropping items with high α-if-item-deleted statistics. These statistics indicate the internal consistency of the scale (represented by Cronbach's α) without each item. If the internal consistency of the items would be higher without that item, we remove it from the scale. We then remove items with item-total correlations lower than 0.25. Our second

analysis removes all items with item-total correlations lower than .1, and when there are no longer any items in the scale with such low correlations, we begin removing items with item-total correlations lower than 0.25. All three analyses leave us with virtually the same set items.

9. Formally, we can compute the total contract information, analogous to the total test information of a test, as the inverse of the empirical Bayes posterior variance V^*_i of $\theta^*_{i,39}$, $\hat{V}_i^{*-1}$, given the data and the estimates of γ and τ (Reardon & Raudenbush, 2006).

10. If a quite rare regulation is *not* the most restrictive (severe) by some subjective standard but, for example, is the result of some historic district-specific problem, this item may not fit in the same dimension as the other items. Given that the Rasch model assumes only one dimension, it is not possible to determine these misfitting items. However, it is likely that we drop misfitting items through our item selection procedures based on our analyses of item-total correlations. We also attempt to accommodate these assumptions and ensure that we are capturing contract restrictiveness rather than historical remnants in contracts by removing outlier regulations, or those that only occur in one or two contracts, from the model.

11. Note that Reardon and Raudenbush (2006) ensure that PIIR model correctly fits their data by checking the observed versus model-predicted conditional and marginal probabilities. They find that the PIIR model is able to predict very similar marginal and conditional probabilities as are observed in the data. They find that the model may slightly underestimate the observed probabilities, with the greatest discrepancies occurring for the rarest items.

References

Armstrong, R. D., Jones, D. H., & Wang, Z. (1998). Optimization of classical reliability in test construction. *Journal of Educational and Behavioral Statistics*, *23*, 1–17.

Baugh, W. H., & Stone, J. A. (1982). Teachers, unions, and wages in the 1970s: Unionism now pays. *Industrial Labor Relations Review*, *35*, 368–376.

Benecki, S. (1987). Municipal expenditure levels and collective bargaining. *Industrial Relations*, *27*, 216–230.

Blau, F. R., & Kahn, L. M. (1996). International differences in male wage inequality: Institutions versus market forces. *Journal of Political Economy*, *104*, 791–837.

Brewer, D. J., & Smith, J. (2007). *Evaluating the 'crazy quilt': Educational governance in California*. Paper presented as part of the *Getting Down to Facts* research studies, Stanford, CA.

Chambers, J. G. (1977). The impact of collective bargaining for teachers on resource allocation in public school districts. *Journal of Urban Economics*, *4*, 324–339.

Cheong, Y. F., & Raudenbush, S. W. (2000). Measurement and structural models for children's problem behaviors. *Psychological Methods*, *5*, 477–495.

Clark, K. B. (1980). Unions and productivity: Micro-econometric evidence. *The Quarterly Journal of Economics*, *95*, 613–639.

Easton, T. (1988). Bargaining and the determinants of teacher salaries. *Industrial and Labor Relations Review, 41*, 263–278.

Eberts, R. W. (2007). Teachers unions and student performance: Help or hindrance? *The Future of Children, 17*, 175–200.

\Eberts, R. W., & Stone, J. A. (1987). Teachers unions and the productivity of public schools. *Industrial and Labor Relations Review, 40*, 354–363.

Ehrenberg, R. G., Sherman, D. R., & Schwarz, J. L. (1983). Unions and productivity in the public sector: A study of municipal libraries. *Industrial and Labor Relations Review, 36*, 199–213.

Freeman, R. B., & Medoff, J. L. (1984). *What do unions do?* New York, NY: Basic Books.

Freeman, R. B., & Valletta, R. G. (1988). The effects of public sector labor laws on labor market institutions and outcomes. In R. B. Freeman & C. Ichniowski (Eds), *When public sector workers unionize.* Cambridge, MA: National Bureau of Economic Research.

Goldhaber, D. (2006). Are teachers unions good for students? In J. Hannaway & A. J. Rotherham (Eds), *Collective bargaining in education: Negotiating change in today's schools* (pp. 141–57). Cambridge, MA: Harvard Education Press.

Hambelton, R. K., Swaminathan, J., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: SAGE.

Hess, F. M., & Kelly, A. P. (2006). Scapegoat, albatross, or what? In J. Hannaway & A. J. Rotherham (Eds), *Collective bargaining in education: Negotiating change in today's schools* (pp. 53–87). Harvard Education Press: Cambridge, MA.

Hess, F. M., & Loup, C. (2008). *The leadership limbo: Teacher labor agreements in America's fifty largest school districts.* The Thomas B. Fordham Institute: Washington, D.C.

Hoxby, C. M. (1996). How teachers unions affect education production. *The Quarterly Journal of Economics, 111*, 617–718.

Kleiner, M., & Petree, D. (1988). Unionism and licensing of public school teachers: Impact on wages and educational output. In R. Freeman & C. Ichniowski (Eds), *When public sector workers unionize.* Chicago, IL: University of Chicago Press.

Koski, W. S., & Horng, E. L. (2007). Curbing or facilitating inequality? Law, collective bargaining, and teacher assignment among schools in California. *Education Finance and Policy, 2*, 262–300.

Levin, J., Mulhern, J., & Schunck, J. (2005). *Unintended consequences: The case for reforming staffing rules in urban teachers union contracts.* New York, NY: The New Teacher Project.

Linn, R. L., & Miller, M. D. (2005). *Measurement and assessment in teaching* (9th ed.). Columbus, OH: Pearson-Merrill Prentice Hall.

Milkman, M. I. (1997). Teachers' unions, productivity and minority student achievement. *Journal of Labor Research, 18*, 247–250.

Moe, T. M. (2006). Union power and the education of children. In J. Hannaway & A. J. Rotherham (Eds), *Collective bargaining in education: Negotiating change in today's schools* (pp. 229–255). Cambridge, MA: Harvard Education Press.

Moe, T. M. (2009). Collective bargaining and the performance of the public schools. *American Journal of Political Science, 53*, 156–174.

Nash, J. (1953). Two-person cooperative games. *Econometrica*, *21*, 128–140.

Osgood, D. W., McMorris, J. B., & Potenza, M. T. (2002). Analyzing multiple-item measures of crime and deviance 1: Item response theory scaling. *Journal of Quantitative Criminology*, *18*, 267–196.

Pencavel, J. (1997). The distributional and efficiency effects of trade unions in Britain. *British Journal of Industrial Relations*, *15*, 137–156.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests.* Chicago, IL: University of Chicago Press.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. (2005). *Hierarchical linear and non-linear modeling, version 6: Users' guide and software program.* Chicago, IL: Scientific Software International.

Raudenbush, S. W., Johnson, C., & Sampson, R. J. (2003). A multivariate, multilevel Rasch model for self-reported criminal behavior. *Sociological Methodology*, *33*, 169–211.

Raudenbush, S. W., & Sampson, R. J. (1999). Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Sociological Methodology*, *29*, 1–41.

Raudenbush, S. W., Yang, M., & Yosef, M. (2000). Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation. *Journal of Computational and Graphical Statistics*, *9*, 141–157.

Reardon, S. F., & Raudenbush, S. W. (2006). A partial independence item response model for surveys with filter questions. *Journal of Sociological Methods*, *36*, 257–300.

Rose, H., & Sonstelie, J. (2010). School board politics, school district size, and the bargaining power of teachers' unions. *Journal of Urban Economics*, *67*(3), 438–450.

Steelman, L. C., Powell, B., & Carini, R. M. (2000). Do teacher unions hinder educational performance? *Harvard Educational Review*, *70*, 437–466.

Trejo, S. J. (1991). Public sector unions and municipal employment. *Industrial and Labor Relations Review*, *45*, 166–180.

United States Department of Labor, Bureau of Labor Statistics. (2005). *Current population survey: Union membership by industry, 2005.* Retrieved January 30, 2009, from http://www.bls.gov/opub/ted/2006/jan/wk4/art02.htm

Woodbury, S. (1985). The scope of bargaining and bargaining outcomes in public schools. *Industrial and Labor Relations Review*, *38*, 195–210.

Wright, B., & Masters, G. (1982). *Rating scale analysis: Rasch measurement.* Chicago, IL: MESA Press.

Wright, B., & Stone, M. (1979). *Best test design: Rasch measurement.* Chicago, IL: MESA Press.

Zax, J. S. (1989). Employment and local public sector unions. *Industrial Relations*, *28*, 21–31.

Authors

KATHARINE O. STRUNK is an assistant professor of education and policy in the Rossier School of Education and, by courtesy, in the School of Policy, Planning and Development at the University of Southern California, 901E Waite Philips Hall, 3470 Trousdale Parkway, Los Angeles, CA 90089; kstrunk@usc.edu. She is also an affiliate of the Center on Education Governance and of Policy Analysis for California Education. Her research focuses on teacher labor markets, education governance and education finance.

SEAN F. REARDON is an associate professor of education and (by courtesy) sociology at Stanford University, 520 Galvez Mall, #526, Stanford, CA 94305-3084; sean.reardon@stanford.edu. His research focuses on the measurement, causes, patterns, and consequences of social and educational inequality and on methods of statistical measurement and analysis

Manuscript received April 13, 2009
Revision received September 18, 2009
Accepted November 8, 2009