

Creating Performance Measures from Survey Data:  
A Practical Discussion

Scott D. Camp  
William G. Saylor

Office of Research and Evaluation  
Federal Bureau of Prisons  
320 First Street NW  
Washington, DC 20534  
202-724-3121

Kevin N. Wright

University of Binghamton  
Binghamton, NY

January 30, 1998

The views expressed in this paper represent those of the authors and not necessarily those of either the Federal Bureau of Prisons or the Department of Justice.

We would like to acknowledge the useful feedback provided by Phyllis Newton and four reviewers for *Corrections Management Quarterly (CMQ)*. We are responsible, though, for any remaining inadequacies. This paper appears in *CMQ* 3: 1: 71-80.

## Creating Performance Measures from Survey Data: A Practical Discussion

### Abstract

Survey data collected from individuals are often used in applied settings to create summary measures of organizational characteristics by calculating simple averages. While this has generated debate in the past in the field of organizational research, there has not been much recent published literature to provide guidance on the applied uses of summary measures. Creating summary measures is often desirable if the measures can be thought of as indicators of management performance.

This paper examines the issues involved in using summary measures of individual-level data as performance indicators. These points are made in the concrete context of examining two scales created by the Federal Bureau of Prisons from annual staff surveys. The scales examined, scales of job satisfaction and institutional commitment, nicely illustrate the necessity of examining issues surrounding the purpose, conceptualization, measurement, and presentation of summary measures. As is demonstrated, a good measure at the individual level of analysis is not necessarily a desirable or good measure at a summary level of analysis, and summary measures often have to be adjusted for factors that are unrelated to performance.

## Creating Performance Measures from Survey Data:

### A Practical Discussion

Applied organizations researchers routinely conduct surveys of staff members to obtain information about individual behaviors and attitudes. For example, measures of organizational commitment are extremely common. Current theories of high performance work organizations (Applebaum and Batt 1994) emphasize that a committed work force is necessary to implement and operate contemporary, flexible forms of work organization. Surveys provide a convenient and economical means of obtaining feedback regarding the attitudinal and behavioral characteristics of individuals working in organizations.

Survey data are often put to a secondary use in applied settings, although these uses are not often reported in the published literature. Survey data are routinely used to create summary or group-level measures. For example, both the Federal Bureau of Prisons (BOP) and Correctional Services of Canada create summary averages of individual-level survey data to characterize regional and institutional differences in the perceptions of staff regarding the attitudes or behaviors in question. Either explicitly or implicitly, the simple averages are often seen as performance indicators for the respective management of the organizational unit.

For example, the BOP reports in the Key Indicators/Strategic Support System, an executive information system distributed monthly on CD-ROM (Gilman 1991; Saylor 1988), on the percentages of staff holding favorable evaluations<sup>1</sup> of institutional commitment. The data are

---

<sup>1</sup>The items used to construct the scale of institutional commitment are measured with a 7 point Likert scale. A respondent's evaluation is judged to be favorable if, on average, the responses are either somewhat agree, agree, or strongly agree. The actual scale items are presented later in the section on "Conceptualization of Summary Measures."

summarized in various ways including by institution, region, and security level. Summaries for survey data collected in 1995 suggest that 65% of staff at high security BOP institutions held favorable perceptions of institutional commitment. This compares well to the BOP overall as 56% of all BOP staff working at field locations in 1995 provided favorable evaluations of commitment to their institution. This comparison suggests that there is something about working at high security institutions, perhaps different managerial practices at high security institutions, that fosters higher levels of institutional commitment. Such a suggestion, however, may be unwarranted as we will examine later.

There are four issues raised by using survey data in the summary manner described above that are discussed in this paper. There are issues associated with 1) purpose: the intended uses of the summary data; 2) conceptualization: the theoretical understanding of the measures (Klein, Dansereau, and Hall 1994; Lazarsfeld and Menzel 1961; Lincoln and Zeitz 1980); 3) measurement: the proper identification of what causes the measures to differ (Camp, Saylor, and Harer 1997; Mossholder and Bedeian 1983; Wright, Camp, Gilman, and Saylor 1996); and 4) presentation: the choice of the appropriate measure to display to managers (Fitz-Gibbon 1991; Saylor 1996). There is little published research in the organizational literature guiding practitioners regarding the *uses* of summary measures with respect to these four issue.<sup>2</sup> We demonstrate how the four areas of concern can be addressed in an applied organizational setting

---

<sup>2</sup>For an overview of existing literature on the multilevel nature of organizational data, see the conceptual piece by Klein and her colleagues (Klein, Danserau and Hall 1994). Fortunately, there has been work done, especially in applied research on performance and achievement in education, that can be used to provide guidance for applied uses in corrections (Bryk and Raudenbush 1992; Fitz-Gibbon 1991).

by looking at the actual uses of two scales created from survey data at the Federal Bureau of Prisons.

*Purpose*

Examining simple or unadjusted summary measures is fine for some purposes and in some situations. For example, if we simply want to know which group has the highest percentage of some characteristic, say voluntary turnover, without trying to understand the reasons for why the group has the characteristic, then an examination of simple or unadjusted measures is fine. For example, if we are tasked with allocating training budgets to institutions, then one piece of information that is useful is the turnover rate. Institutions with high turnover rates need more resources for training new staff, regardless of the reason(s) for the high turnover rates.

Likewise, if we know that individuals are randomly assigned to different groups, then unadjusted measures are also fine. In evaluation studies, we sometimes randomly assign individuals to groups to control for differences individuals bring to our studies. However, we do not assign workers to institutions on a random basis. Similarly, we are usually concerned with why some group has a given characteristic. In the turnover example, most of us would be led to the question of why the different institutions have different turnover rates. Is the turnover rate affected by management practices? Can we rank institutional management on their effects on turnover rates? To answer these types of questions, where we are explicitly concerned with ranking the relative performance of management, it is absolutely essential to control for the individual- and group-level factors that influence individual evaluations or individual actions. It is this latter, more typical case, where performance is at issue, that forms the basis of the following discussion.

### *Conceptualization of Summary Measures*

The paramount theoretical issue associated with creating summary measures is this: does the summary measure capture a group characteristic, or does the measure summarize individual responses? The answer to this question determines whether or not the summary measure can be used as an indicator of management performance. If the summary measure is only dependent upon the characteristics of the individuals answering the question, then it is not fair or appropriate to hold management accountable for a summary created from the individual responses. If, on the other hand, the responses provided by individuals are influenced by working at a particular correctional institution (or other group-level unit), then it is appropriate to hold management accountable for that portion of the response that is influenced by working at that institution.

Let us put this in practical terms. If you want to create a simple summary measure of the dangerousness of inmates that is based on staff perceptions, then you first ask a random sample of staff at each facility how dangerous they think inmates are. From this, you create a simple average of the responses provided by the staff at the respective institutions. You can then rank the institutions based on these averages. But what does this tell us? Are there factors other than the dangerousness of inmates that affect the responses staff provide? Sure there are. For example, staff with low amounts of tenure working in corrections will evaluate inmates differently than staff who have worked in corrections for longer periods of time. If the average tenure of staff differs from one institution to another, then this average difference in the characteristics of the individuals providing evaluations will systematically affect the average evaluations of inmate dangerousness.

The trick, then, is to present an average evaluation of dangerousness that controls for the average characteristics of the staff providing the evaluations. Another way of saying this is that

we need to present a summary measure that captures only that part of the dangerousness of inmates that is due to differences between institutions with respect to the characteristics of their inmates. The method we use to construct such measures is technically complex, but from the point of view of data consumers, it is not necessary to understand the technical details to effectively utilize the information.

As we demonstrate below, if correctional administrators continue to rely upon simple averages to evaluate institutional or management performance, then they are going to perpetuate the use of measures that confuse true institutional differences with differences caused by having different combinations of evaluators/subjects. For example, all existing evaluations of prison privatization have relied upon simple averages to compare privately and publicly managed prisons. Such comparisons, though, do not allow for a proper understanding of whether the private and public institutions vary because of performance differences or because the institutions are staffed by and incarcerate different types of individuals (Federal Bureau of Prisons. Office of Research and Evaluation 1998). Without methods to statistically disentangle these differences, we really have no clear idea of what the summary measures mean, and this can seriously affect our ability to assess performance or conduct meaningful evaluations.

We examine two scales currently used by the BOP to illustrate the points made above. We examine the characteristics, both individual- and group-level, of the scales measuring institutional commitment and job satisfaction. These scales are of continuing interest to organizational researchers, and prior work has been performed on these scales from data collected at the BOP (Camp, Saylor and Harer 1997; Wright, Camp, Gilman and Saylor 1996).

The institutional commitment scale is calculated from the following three items.

1. This facility is the best in the whole BOP.
2. I would rather be stationed at this facility than any other I know about.
3. I would like to continue to work at this facility.

The respondents have seven choices for each item ranging from strongly disagree to strongly agree. While the first item appears to ask respondents to rate or evaluate the operations of the institution, the other two items ask respondents about their intentions to remain working at the institution. In the literature on organizational commitment, intent to stay is a major component of organizational commitment. Despite having only three items, this scale has high reliability at the individual level as measured by the Cronbach's alpha of .83. Generally speaking, the reliability of a measure refers to how consistent the measure is from one use to another. A perfectly reliable measure has a reliability value of 1 (Vogt 1993: 195). Generally, any value for reliability less than 0.7 is considered suspect. In the specific case of Cronbach's alpha, reliability means that there is general agreement among the respondents about the three items. If a respondent rates one item favorably, then it is extremely likely that the respondent will also rate the other items favorably.

The items that comprise the job satisfaction scale ask respondents about their evaluations of satisfaction with their jobs.<sup>3</sup>

1. I would be more satisfied with some other job at this facility than I am with my present job.
2. My BOP job is usually interesting to me.
3. My BOP job suits me very well.
4. My BOP job is usually worthwhile.
5. If I have a chance, I will change to a job at the same rate of pay at this facility.
6. I am currently looking for or considering another job outside the BOP.

Clearly, the scale asks people about their individual feelings of job satisfaction and nothing directly about the correctional institution in which they are working. As an individual-level measure, the job satisfaction scale is quite satisfactory as the Cronbach alpha for the scale is .82.

---

<sup>3</sup>Items 1, 5, and 6 are reverse coded in creating the scale.



In sum, the two scales display much agreement by individuals when they rate the items. *Individuals* are consistent in giving either positive or negative evaluations. There are few cases where a respondent provides a favorable evaluation of one item and an unfavorable evaluation of the other items. This does not mean, however, that we can use this information to justify creating simple averages of individual responses. Creating summary measures is justified when there is agreement *among* individuals at an institution that institutional commitment is high (or low). Testing for this type of agreement *among* raters is the subject matter of the following section.

#### *Measurement of Summary Properties*

Differences in average evaluations by members of an organization may arise from differences in individual evaluators, from differences in organizational influences, or from both sources. Prior work (Camp, Saylor and Harer 1997; Wright, Camp, Gilman and Saylor 1996) suggests that the two individual-level scales examined in this treatment nicely illustrate this point. The interested reader can turn to the discussions in these papers for technical details about the hierarchical linear models (HLM) used to correct the summary measures, or the reader can consult Bryk and Raudenbush (1992) for a more complete overview. In this paper, we simply focus on the substantive results that can be drawn from the respective HLM models.

We start with an examination of job satisfaction. We examine whether it makes sense to create a summary measure for this scale that can be used as a description of an organizational property or performance. We look at three statistical models for job satisfaction (and later for institutional commitment). The models test, respectively, whether there are meaningful differences between institutions in evaluations of job satisfaction (simple model), whether the differences remain after controlling for individual differences in the average respondents at the respective

institutions (base model), and whether the differences in job satisfaction remain after controlling for average individual differences of the evaluators and organizational differences between the institutions in which the evaluators work (full model). If differences in evaluations of job satisfaction remain after controlling for the individual and institutional effects included in the full model, we argue that the differences are influenced by management practices. Although we do not report the results here, the individual-level variables control for supervisory status, Hispanic ethnicity, sex, whether respondent is a line staff correctional officer, age, race, tenure with the BOP, tenure at the institution, education, functional department in which employed, whether the respondent has ever worked at another BOP facility, and whether the respondent works at a satellite camp facility.<sup>4</sup> The full model incorporates group-level measures as well as the individual-level measures into the HLM equations. The variables control for mean sentence length of the inmates, percentage of minority staff members, the percentage of staff with tenure greater than or equal to five years, objective inmate crowding, security level, percentage of black inmates, and percentage of Hispanic inmates.

The scale for job satisfaction used by the BOP does not appear to be influenced by the different institutional contexts that exist across the BOP. As can be seen in the simple model results for job satisfaction presented in Table 1, the intra-class correlation (ICC) is only .00522. The ICC can be thought of as the amount (proportion) of the difference in evaluations of job satisfaction that is caused by the varying institutional contexts in which workers are located. The percentage of variance (the amount of difference) is computed by multiplying the ICC by 100.

---

<sup>4</sup>Complete results are available from the authors. The individual-level and group-level control variables were chosen because previous empirical and theoretical research has suggested their importance. We do not feel that it is appropriate to review this literature here.

When developing group-level performance indicators, we want the ICC to be as high as possible. The low value of ICC for the simple model of job satisfaction suggests that before adding any other variables to the model, less than 1% (0.522%) of the variance in job satisfaction is attributable to differences between institutions. Almost all of the variance, more than 99% of the variance, in job satisfaction evaluations exists within the individual evaluators of job satisfaction.

Another way of stating this is that less than 1% of the differences in job satisfaction evaluations is caused by organizational-level factors, such as management practices. As such, management in the BOP exerts little if any influence over individual perceptions of job satisfaction, at least on an institution by institution basis. The base model for job satisfaction demonstrates the same results. The base model also shows that less than 1% of the differences in job satisfaction evaluations are caused by institutional differences. The poor properties of job satisfaction as an organizational performance indicator of the ability of institutional management to manipulate the work environment to raise or lower perceptions of job satisfaction are confirmed by reliability estimates. The reliability values are well below the general rule of thumb cutoff value of .70 for both the simple (0.257) and base models (0.302) of job satisfaction. Because of the poor performance of job satisfaction as a group-level measure capturing differences between institutions, there was no need to develop a full model.

Institutional commitment, on the other hand, appears to be influenced by unmeasured differences across institutions, what we are assuming to be differences in performance of institutional management in raising or lowering individual-level evaluations of institutional commitment. For example, the simple model suggests that 12.7% (the ICC is 0.12735) of the difference in evaluations of institutional commitment is caused by differences between institutions.

The base model, where the individual-level measures are entered into the model, also suggests that almost 12% (ICC is 0.11708) of the difference exists at the group level or between institutions. Even after entering the group-level measures that influence evaluations of institutional commitment, there is still over 7% of the difference in evaluations of institutional commitment residing between institutions. Additionally, the reliability scores for the different models of institutional commitment suggest that the summary scores are quite reliable. For the full model, the one we will use to create an adjusted summary measure of institutional commitment as outlined in the next section, the reliability is well above the rule of thumb cutoff of 0.7. The reliability for the full model is 0.828. All of this empirical evidence suggests that institutional practices by management influence individual-level evaluations of institutional commitment. The adjusted summary measure presented below is an indicator of the ability of institutional management/practices to influence individual-level evaluations of institutional commitment.

#### *Presentation of Summary Measures*

One method of providing adjusted measures, adjusted both for individual and group level sources of variation, is to use the HLM methods discussed previously to produce what is called an empirical Bayes (EB) residual estimator (Bryk and Raudenbush 1992; Fitz-Gibbon 1991). While the statistic sounds intimidating, it has a straight-forward interpretation. The EB estimator we use here is the difference between the “average” we actually observe for a group and the “average” that is expected for the group after controlling for the individual and group factors that influence the average.<sup>5</sup> If a residual is positive, it means that the observed average for that group (in our

---

<sup>5</sup>In the models examined here, independent variables that are continuous are centered around the mean value for the parameter. This is done to create greater stability in the estimation of the intercepts. Categorical variables, such as sex and race, are coded with dummy (for

case the group is a correctional institution) is greater than would be expected. Generally, it would be thought that the institution is performing better than expected. On the other hand, if the residual is negative, then the group (institution) is performing less favorably than expected. In essence, the EB residual is that part of the evaluation of the variable at hand (institutional commitment in the present analysis) that we believe to be due to management practices.

It clearly makes a difference whether unadjusted or adjusted summary measures are used as performance indicators. Figure 1 shows the ranking of all 80 BOP institutions on the institutional commitment scale in 1995. The institutions are not identified by name because of confidentiality considerations. It is clear from the figure that i01 has the lowest unadjusted summary score for institutional commitment, and the other institutions are labeled in the order from low to high in which they appear in Figure 1.

Figure 2 shows a ranking of the same institutions on the institutional commitment scale, again for 1995, but now the rankings are determined by the difference between the observed summary value of institutional commitment and the expected value. The negative scores indicate institutions that are performing worse than would be expected given the individual characteristics of the staff at the institution and the group-level factors. Positive values indicate institutions that are performing better than expected. It is easy to identify institutions performing worse than expected as the bars all extend down from 0 and vice versa for the institutions performing above expectations. After accounting for known factors that influence evaluations of institutional commitment, factors unrelated to the performance of the institution in generating levels of

---

dichotomies) or effects vector schemes (where there are three or more categories). In essence, the expected values are “adjusted” averages for a “common” type of BOP employee at each of the institutions.

institutional commitment, we see a quite different ranking of the 80 institutions. Let us look at the bottom 10 and top 10 performing institutions as we are generally most interested in performance at the extremes.

As can be seen in Figure 1, based on the unadjusted summary scores, institutions one through ten (i01-i10) were the worst performers in terms of motivating staff to high levels of institutional commitment. However, we see in Figure 2 that when we account for the individual and group factors that influence evaluations of institutional commitment, only five of the institutions identified with the unadjusted scores remain among the bottom ten performers (i01, i10, i07, i02, and i06). The other five institutions that are low performers in generating levels of institutional commitment, according to the adjusted scores presented in Figure 2, come from the lower portion of the ranked unadjusted ratings presented in Figure 1 (i24, i20, i26, i23, and i31).

Without adjusting the summary scores, it is not possible to identify five of the institutions that may be experiencing substantial problems that lead staff to provide much lower evaluations of institutional commitment than would be expected. Of course, it is *extremely* important to note that the rankings provided by the EB residuals are simply *relational*. That is, at an absolute level, all of the 80 institutions may provide conditions that foster levels of institutional commitment that are sufficient for running good correctional institutions. The methods presented here do not address this issue. Instead, they simply provide a better means of ranking the relative performance of institutions. Nonetheless, if analysts or practitioners rely upon the unadjusted scores to produce relative rankings, they run the very real risk of not identifying half of the institutions in this example where performance is extremely poor. Additionally, if analysts or practitioners rely upon unadjusted score, they could also identify five institutions as extremely poor performers that do

not deserve this designation. The unadjusted scores are partly influenced by factors that are outside of the control or concern of institutional management, at least in terms of measuring management practices or institutional performance.

The results in Figure 2 also demonstrate misclassification with respect to superior performance. Only five of the top ten performing institutions are identified correctly with unadjusted scores. Most dramatically, institution 21 is identified with unadjusted scores as being a relatively low performing institution. After accounting for the appropriate factors, institution 21 is identified as a top performing institution. This is a dramatic reversal from the classification of institution 21 based on the unadjusted scores where it was identified as a weak to mediocre performing institution. The other additions to the top ten are somewhat less dramatic (i64 which turns out to be the best performing institution, i60, i68, and i61), but the changes in ranking are nonetheless important (probably most important for the respective wardens). If analysts or practitioners turn to the ranking of institutions identified as top performers by the unadjusted scores, they will be looking at some of the wrong institutions for guidance about best management practices and other aspects of organizational and managerial recognition.

A final chart, Figure 3, is presented to reinforce the points made about the different rankings produced by the unadjusted and adjusted summary scores. All of the information in the Figure 3 is contained in Figures 1 and 2, but the slightly modified presentation helps drive home the point about a lack of correspondence between rankings produced by using adjusted and unadjusted scores. In Figure 3, the top portion of the panel contains an ordered presentation of the empirical Bayes residuals for institutional commitment identical to that presented in Figure 2. The bottom panel of Figure 3 presents the unadjusted summary scores of institutional

commitment presented in Figure 1, although the presentation is slightly modified. Specifically, the mean value of institutional commitment for each institution has the grand mean (the mean for all BOP institutions) subtracted from it. Subtracting a constant, the grand mean, does not change the relative order of the rankings. The subtraction of the BOP mean for each measure was done to allow the bottom panel to resemble the nature of the graph in the top panel of Figure 3. But we see that the bottom panel does not have the smooth ordering noted for the top panel of Figure 3. The difference results from the institutions in the bottom panel being presented in the order of the empirical Bayes residuals, that is, the same order used in the top panel of Figure 3. The first bar in each panel represents the same institution, the second bar in each panel represents the same institution, etc. If the unadjusted scores produced a similar ranking to that produced by ranking the institutions by empirical Bayes residuals, the bottom panel would resemble the top panel in Figure 3 with the exception that the lines in the bottom panel are generally longer.<sup>6</sup> Since we can visually see in Figure 3 that there is a lack of correspondence between the adjusted (top panel) and unadjusted (bottom panel) summary measures of institutional commitment, we have further information about the consequences of using adjusted or unadjusted measures of organizational performance.

### *Concluding Remarks*

We noted early in this discussion that unadjusted measures of institutional commitment suggest that employees at high security institutions are more committed than the average BOP employee. We stated, though, that we should hold judgement about this bit of knowledge until we

---

<sup>6</sup>Empirical Bayes residuals are smaller because the individual- and group-level effects are modeled, and there is shrinkage of the residuals created when they are adjusted for reliability.



had developed appropriate models of institutional commitment at the individual and group levels. One of the group-level variables we included in the full model of institutional commitment was security level. Holding constant the other individual- and group-level measures considered in the full model, it turns out that employees at high security institutions are no more committed to their institutions than other BOP employees.<sup>7</sup> The bivariate relationship between institutional commitment and security level goes away in multivariate analysis.

We also have demonstrated how the empirical Bayes residuals from the HLM models can be used as more meaningful summary measures of institutional commitment. We contend that institutional commitment has fairly good group properties, so it is possible to summarize this measure and interpret it as a measure of organizational or management performance. We believe that an organizational measure of institutional commitment makes sense as an indirect measure of management's ability to motivate workers. Recall the case of institution 21. The unadjusted measure of institutional commitment gives no indication of the superior performance of the institutional management in raising individual evaluations of institutional commitment well above what would otherwise be expected (given the individual-level characteristics of the raters and the organizational features of the institution). Without the knowledge provided by the adjusted measures, we would most likely fail to recognize the superior performance and practice of the management at institution 21.

In summary, we hope this discussion has pointed to the potential problems inherent in the unquestioning use of unadjusted summary measures. We understand that it is very easy to create

---

<sup>7</sup>The results for the variables included in the HLM models are not presented here as they are not the primary focus of this discussion. However, as noted previously, the results are available from the authors.

and make use of unadjusted summary measures created from survey data. However, the measures simply may not have the meaning that practitioners believe. As such, we recommend that practitioners and analysts spend some time examining the purpose and measurement properties of any measures that are aggregated (including aggregate measures of individual behaviors such as misconduct). Some measures simply do not have any meaning as measures of institutional performance or management effectiveness. And measures that do have the potential to be used as performance measures are usually influenced by other factors unrelated to performance. For these measures, it is necessary to create adjusted scores that are then presented in an easily interpreted fashion. With this type of methodological foundation, it is possible to create appropriate measures that can be explicitly used as performance indicators.

## References

- Applebaum, Eileen and Rosemary Batt. 1994. *The New American Workplace: Transforming Work Systems in the United States*. Ithaca, NY: ILR Press.
- Bryk, Anthony S. and Stephen W. Raudenbush. 1992. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage.
- Camp, Scott D., William G. Saylor, and Miles D. Harer. 1997. "Aggregating Individual-Level Evaluations of the Organizational Social Climate: A Multilevel Investigation of the Work Environment at the Federal Bureau of Prisons." *Justice Quarterly* 14:739-761.
- Federal Bureau of Prisons. Office of Research and Evaluation. 1998. "Proposal to Evaluate the Performance of a Privately Operated Bureau of Prisons Facility." Federal Bureau of Prisons, Washington, D.C.
- Fitz-Gibbon, Carol T. 1991. "Multilevel Modelling in an Indicator System." Pp. 67-83 in *Schools, Classrooms, and Pupils: International Studies of Schooling from a Multilevel Perspective*, edited by S. W. Raudenbush and J. D. Willms. San Diego: Academic Press.
- Gilman, Evan. 1991. "Implementing Key Indicators." *Federal Prisons Journal* 2:48-56.
- Klein, Katherine J., Fred Dansereau, and Rosalie J. Hall. 1994. "Levels Issues in Theory Development, Data Collection, and Analysis." *Academy of Management Review* 19:195-229.
- Lazarsfeld, Paul F. and Herbert Menzel. 1961. "On the Relation Between Individual and Collective Properties." Pp. 422-440 in *Complex Organizations: A Sociological Reader*, edited by A. Etzioni. Glencoe, IL: Free Press.

- Lincoln, James R. and Gerald Zeitz. 1980. "Organizational Properties from Aggregate Data: Separating Individual and Structural Effects." *American Sociological Review* 45:391-408.
- Mossholder, K. W. and A. G. Bedeian. 1983. "Cross-Level Inference and Organizational Research: Perspectives on Interpretation and Application." *Academy of Management Review* 8:547-558.
- Saylor, William G. 1988. "Developing a Strategic Support System: Putting Social Science Research into Practice to Improve Prison Management." Federal Bureau of Prisons, Washington, D.C.
- Saylor, William G. 1996. "Modeling and Graphing Organizational Processes in Pursuit of Performance Benchmarks: Methods for Establishing and Evaluating Performance Measures." Federal Bureau of Prisons, Washington, D.C.
- Vogt, W. Paul. 1993. *Dictionary of Statistics and Methodology*. Newbury Park, CA: Sage.
- Wright, Kevin N., Scott D. Camp, Evan B. Gilman, and William G. Saylor. 1996. "Hierarchical Linear Models of Individual and Organizational Effects of Participatory Management Practices in a Prison System." Paper presented at the Annual Meeting of the American Society of Criminology, Chicago.

Table 1  
 Variance Components of Job Satisfaction and Institutional Operations

	Job Satisfaction			Institutional Commitment		
	Simple	Base	Full	Simple	Base	Full
Within	1.61712	1.42254	--	1.91865	1.82299	1.82310
Between	0.00848	0.00939	--	0.27999	0.24175	0.14174
ICC	0.00522	0.00656	--	0.12735	0.11708	0.07214
Reliability	0.257	0.302	--	0.900	0.891	0.828

Figure 1

# Ranking by Institutional Commitment Based on Unadjusted Aggregate Scores

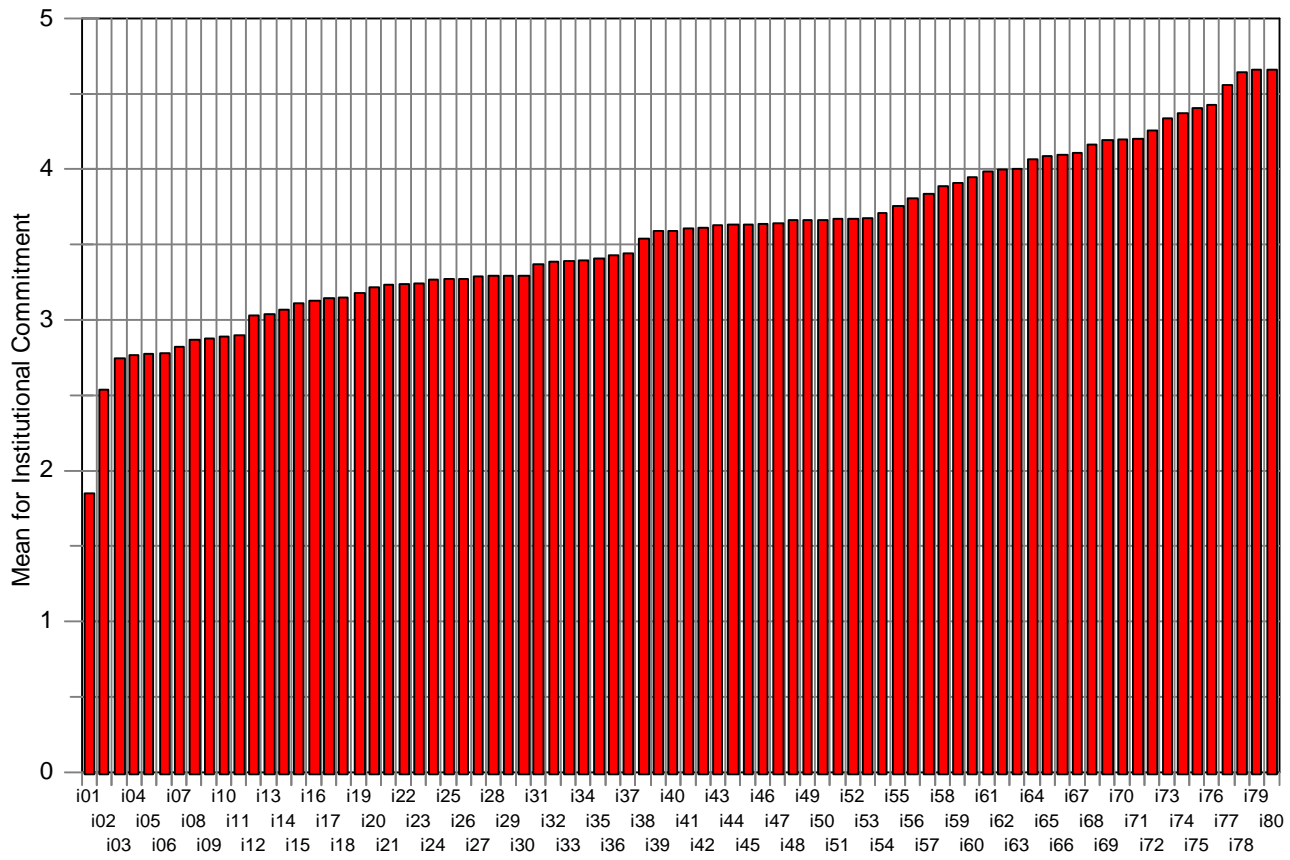


Figure 2

# Ranking By Institutional Commitment Based on Empirical Bayes Residuals

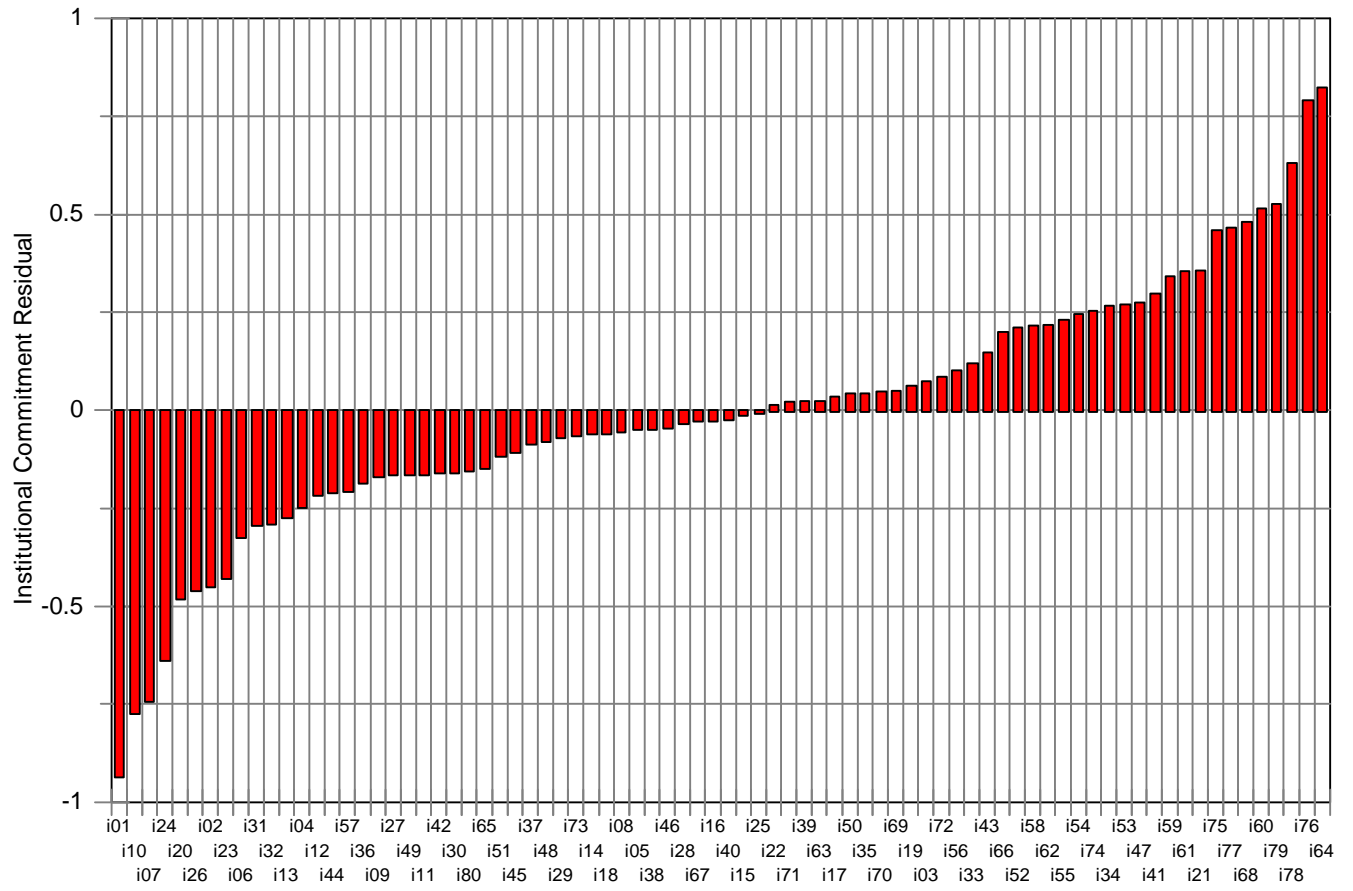
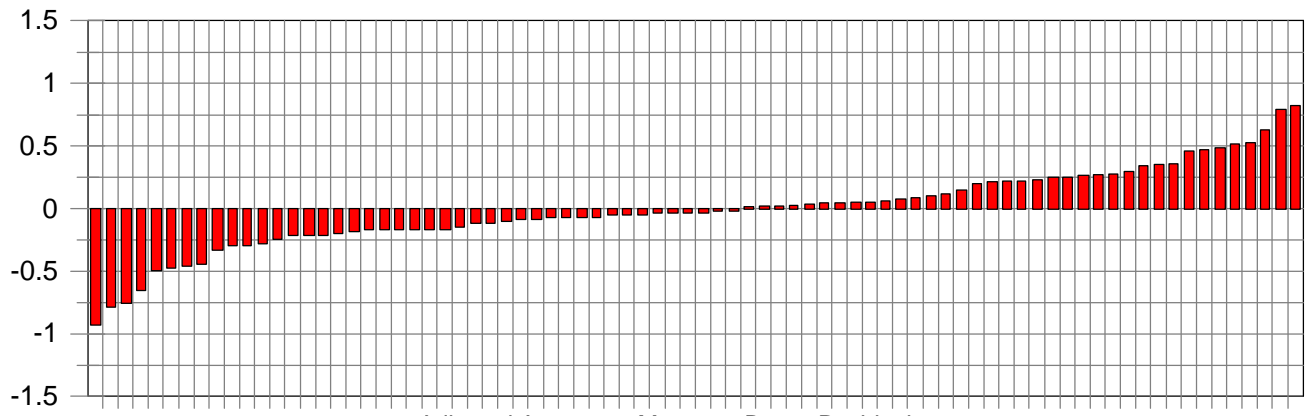
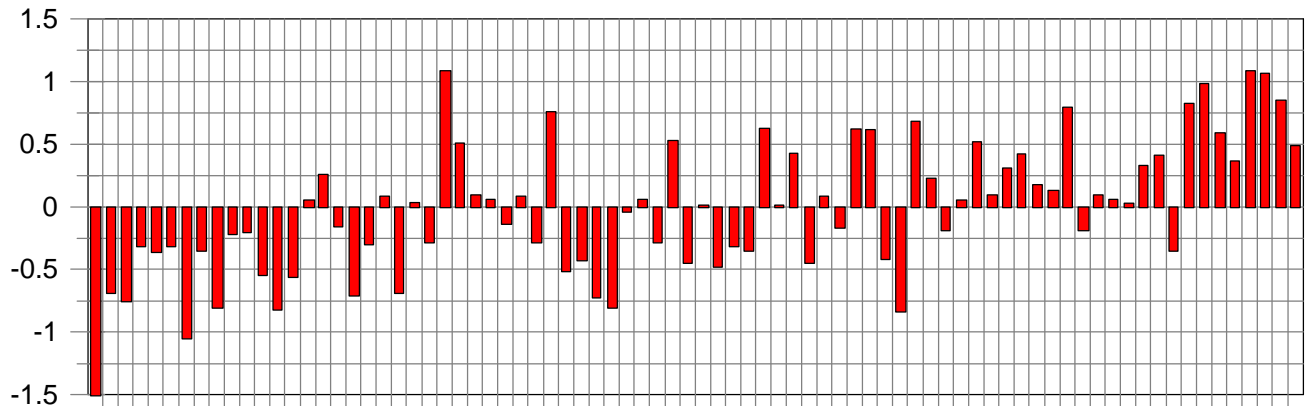


Figure 3

### Corresponding Performance Indicators Institutional Commitment



Adjusted Aggregate Measure, Bayes Residuals



Unadjusted Aggregate Measure, Deviations from BOP Mean