Assessing Prison Performance from the Inmates' View: Comparing Private and Public Prisons

Scott D. Camp, Ph.D.
Gerald G. Gaes, Ph.D.
Jody Klein-Saffran, Ph.D.
Dawn M. Daggett, M.S.
William G. Saylor, M.A.

August 6, 2001

Office Research and Evaluation
Federal Bureau of Prisons
320 First Street NW
400 Building, Room 3025
Washington, D.C. 20534

contact: scamp@bop.gov
202-307-3871 x. 148 (Camp)

Assessing Prison Performance from the Inmates' View: Comparing Private and Public Prisons

Assessing Prison Performance from the Inmates' View:
Comparing Private and Public Prisons

Abstract

The public sector needs to monitor the performance of the private prisons, and it is necessary to conduct the monitoring as objectively as possible. This paper demonstrates that an often overlooked source of data, surveys of inmates, can be used to differentiate prisons on areas such as gang activity, safety and security, sanitation, and food service delivery. Hierarchical linear models were used to generate the prison performance measures. Second, we show that inmates and staff largely agree in their assessments of conditions at the prison. Finally, we demonstrate that while there is considerable consistency for different measures within the topical areas examined, there is no necessary correspondence in performance across the different topical areas of gang management, safety and security, sanitation, and food service delivery. While surveys will and should never replace operational reviews and audits, we demonstrated that they can be effectively used to obtain information about operational differences between prisons.

Assessing Prison Performance from the Inmates' View:
Comparing Private and Public Prisons

The emergence of private adult prisons has generated an industry of proponents and opponents who articulate, respectively, the merits and dangers of privatization. Even when the arguments have turned to empirical matters, the results have been less than conclusive (see the reviews by Austin & Coventry, 2001; Gaes, Camp, & Saylor, 1998; General Accounting Office, 1996). Largely missing from discussions of prison privatization are the views of inmates themselves. Also, existing studies rarely examine whether the methods employed used to compare prisons are appropriate (for an exception on prison quality see Camp & Gaes, 2000; for an exception on cost comparisons see Nelson, 1999).

The question of how best to compare public and private prisons is relevant because both proponents and opponents of private prisons claim that public and private prisons operate differently. Obviously, proponents of private prisons claim that private prisons operate more efficiently with higher quality provided, and opponents argue the opposite or at least that quality is threatened by private operations. The existing quality comparisons have primarily relied upon reviews of operational data and/or audits (Archambeault & Deis, 1996; Bowery, 1997; Tennessee Select Oversight Committee on Corrections, 1995; Thomas, 1997). There are some issues that make audits problematic for comparing prisons on a continuing basis. First, audits are expensive and somewhat subjective, although the solution used in Tennessee to use members of both the public and private sectors as audit team members was creative. Another problem with audits is the difficulty of obtaining information on a sufficient number of facilities at approximately the same time. Finally, audits disrupt normal institutional operations at the respective prisons.

The collection of operational data is certainly less disruptive than audits as it is generally part of normal institutional operations, but the use of operational data presents its own challenges. The most serious challenge to operational data is the ease with which it can be manipulated by differences in reporting, correctional philosophy, and operational practice. The operational data that are most commonly used to compare prisons are misconduct data. The usual assumption is that differences in reported misconduct result from "real" differences in the underlying behaviors. While this assumption may be more adequate for the most serious types of misconduct in prison, such as murders or escapes from secure prisons, it is more tenuous for less serious misconduct. If a warden can convey a philosophy of informal resolution of minor disputes, and especially if the warden is fortunate enough to have a seasoned workforce, then many types of misconduct may never appear in official, operational records. As a result, the operational data are "contaminated" in the sense that they reflect differences other than differences in underlying misconduct rates. The contamination of data for less serious misconduct is most unfortunate for research purposes as the more serious misconduct is much less common, especially in lower security level prisons where private operators tend to be concentrated.

A less common source of data for comparing prisons is survey data collected from inmates and staff (Camp, Gaes, & Saylor, 2001; Logan, 1992; Sechrest & Shichor, 1996). Survey data have the normal advantages associated with survey data: the data are cheap to collect, they cause little disruption during collection, and they can be administered to many institutions in a relatively short time span. Of course, survey data have the typical disadvantages, the one of apparent concern to prison administrators is the ability of respondents to "exaggerate" if they are so inclined. To be blunt, prison administrators often fear that respondents lie to make them look bad. However, an analysis of inmate survey data demonstrated that inmate responses systematically

fluctuated at different U.S. federal prisons, suggesting that the surveys provided something more than a uniform “gripe” forum for the inmates (Camp, 1999). Likewise, researchers in the Office of Research and Evaluation at the Federal Bureau of Prisons have demonstrated that staff responses to certain survey questions are influenced by differences between prisons (Camp, Saylor, & Harer, 1997; Camp, Saylor, & Wright, 1999).

This paper focuses upon how inmate attitudes are influenced by the conditions of the prisons in which the inmates are incarcerated. This is a use of the data not usually undertaken, with some exceptions such as Camp (1999). Most analysts of inmate survey data typically use the data to describe the needs, program experiences (Carlson, 1997), characteristics, or other experiences of inmates, such as victimization in prisons (Maitland, 1996; McCorkle, 1993) or the extent and nature of victimization and violence within prisons (O'Donnell, 1998). Data collected for these surveys often consist of questions which asked inmates about their personal experiences of victimization within the last month of their incarceration (Edgar, 1998; O'Donnell, 1998).

A different use of survey data collected from inmates is provided by the Correctional Service of Canada (CSC). In 1994, CSC contracted with Price Waterhouse to conduct a National Inmate Survey of approximately 4,000 inmates across the five regions of the CSC. The questionnaire was thorough and broken down into five sections: background information, institutional environment, security, staff and programs, and health (Robinson, 1996). The final report examined each question and presented a table which compared the survey responses by region and security level (Price Waterhouse, 1996). However, no further controls were used to adjust the comparisons for individual differences among the raters, controls such as those used by Camp (1999). Also, the measures were not examined to determine whether they actually provided

desirable measurement properties at the levels of institution and region to which the data were aggregated as discussed below.

Given the focus of previous studies using inmate data, analysts have rarely asked extremely important and fundamental questions: How were the inmate responses conditioned by the institutions in which the inmates are incarcerated? Were the responses impervious to the experiences of the inmates in their respective prisons? Or were the responses affected by operational and other differences between the prisons in which the inmates are incarcerated? And, finally, do inmates provide the same information as staff when asked about prison conditions? Typically, analysts have made the mistake of either assuming that responses are necessarily influenced by the prisons in which inmates are incarcerated (see the critique of Logan in Camp et al., 2001), or they simply never address the issue at all.

Several methodological tasks were addressed to answer the overriding question of this analysis: Do inmates provide information about prison conditions that is different from the information solicited from staff. Clearly we could and did ask inmates questions that were not relevant to staff, and we report the results of selected questions here. Beyond this obvious difference in the information provided by inmates and staff, the question is whether inmates have their own perspective on prison operations that differs from that of staff. Common sense could certainly lead us to this presumption. Alternatively, it could also be argued that the same institutional features that influence staff perceptions at an institution also have an impact upon inmates. While inmates and staff are unlikely to have identical opinions about their institutions, their respective opinions may rank the same when compared respectively to inmates and staff at other institutions.

Given the above objective, there were four steps that were followed. First, we investigated the organizational properties of the data collected from inmates and staff. As has been demonstrated elsewhere, it is not sufficient to compare prisons on the proportion of inmates and staff who provide favorable evaluations to survey questions (Camp et al., 2001). Instead, it is necessary to use multilevel modeling techniques to control for pertinent factors when constructing appropriates measures for comparisons at the organizational level (Bryk & Raudenbush, 1992). Second, once the appropriate measures were defined, it was necessary to examine how well inmate and staff evaluations correlated across institutions. That is, if inmates rated institution X as being high on a given measure in comparison to their peers at other prisons, did staff at the institution provide a corresponding evaluation? Both bar graphs and correlations (parametric and nonparametric) were used to assess this component. Third, we examined the coherence of measures within an institution to answer whether inmates at a particular institution tended to rate all aspects of prison operations similarly. A unique graphing package, *Parallel Coordinates Display*, that was designed to implement the data presentation ideas developed by analysts such as Hartigan (1975), Inselburg (1990), and Wegman (1990), was used to generate the graphs for this portion of the study (Howell, 2001). Finally, we used all of the preceding information to compare the operations at one private prison with operations at selected BOP prisons.

Data and Methods

The Federal Bureau of Prisons routinely administers a survey of staff, known as the Prison Social Climate Survey (PSCS). The PSCS was first administered in 1988, and it has been administered annually since that time. In 1999, the year analyzed here, 86.9 percent of the 10,710 staff surveyed returned useable surveys. The complete PSCS questionnaire is divided into topic areas on the socio-demographics of the respondent and their work history, the work environment,

personal safety and security, personal well-being, and quality of life. The sections are mixed across four versions so that individual respondents answer only a subset of the entire instrument. Some questions are asked of all respondents (see Camp et al., 2001).

In March and April of 2000, about six months after the 1999 staff administration, an inmate version of the PSCS was given to inmates at ten low-security prisons. The ten institutions were deliberately chosen to be representative of all low-security prisons in the BOP. Surveys of inmates occur on an as-needed basis, and this survey administration was predicated on the need to obtain additional information about prison operations at Taft Correctional Institution, a private prison operated by Wackenhut Corrections Corporation for the Federal Bureau of Prisons. A little over 100 inmates were surveyed at each institution, or 1,080 inmates overall. Completed surveys were obtained from 950 inmates, for a response rate of 88.0 percent.

As can be seen in Appendix 1, the inmates who completed the survey looked different from the inmates selected for the survey with the exception of age where the intended and actual sample data agree. Even with the discrepancies noted in Appendix 1, we are confident that there was no systematic bias in the types of respondents who completed the survey because of the control that was exercised in administering the survey. The survey was completed by calling inmates out to an area of the prison where the surveys were administered. We are confident that the inmates on our survey lists were the inmates who showed up. For the most part, almost all inmates who were available for callout completed the survey. Most of the non-completions were the result of inmates not being available or even present at the prison. Plus, there were reporting differences reflected in Appendix 1. Hispanic inmates who classified their own race, as opposed to the official data where race was coded with staff assistance, were likely to choose their race as other (88 inmates) or leave the race question blank (231 inmates). Clearly, many Hispanic inmates

viewed their *race* as being Hispanic, where the survey forced this to be a choice for *ethnicity*. Nonetheless, it does appear from the appendix that more inmates view themselves as Hispanics than were recorded as such by the BOP.

There are two other caveats about the data worth noting. First, in the random selection of inmates into the sample, inmates with an IHP status were disproportionately sampled in five of the ten prisons.[1] Taft, at the time, held a large number of IHP inmates, and we wanted to collect information from other prisons to test whether IHP status influenced inmate evaluations of prison conditions. Originally, we had planned to match survey responses back to the respective individual-level operational data with a matching program to re-identify the IHP inmates. However, the socio-demographic information collected in the surveys was not rich enough to obtain satisfactory matches. Since we could not identify the IHP inmates with the matching technique, and because we did not ask the respondents directly about their IHP status, we included a variable in all of the models analyzed here for the propensity, or predicted probability, that the respondents were IHP inmates. The coefficients for predicting the propensity scores were derived from the operational data on the intended sample, where IHP status was known, and the coefficients were then used on the survey responses to calculate the propensity score. In almost all cases, the propensity score did not significantly enter into equations of the outcome variables examined here, leading us to conclude that the sample was not biased in those institutions where IHP inmates were oversampled. The second caveat in the data is the time separation between the inmate and staff surveys. There were no major events in the BOP, such as Bureau-wide inmate

[1]As used here, IHP inmate indicates an inmate who will be turned over to the Immigration and Naturalization Service upon release from Federal prison. Generally, these inmates are deported back to their home countries.

disturbances, that would have had an undue, systematic impact upon altering the later-collected inmate views. Nonetheless, it would have been desirable to collect the data contemporaneously.

Missing data in the inmate and staff databases were handled with the data augmentation procedures described by Schafer (1997) and implemented in the experimental SAS Version 8.1 procedure PROC MI. Since PROC MI requires that categorical variables be measured on an ordinal scale, some variables in the inmate and staff databases were recoded into dummy variables that were appropriate. In particular, race in both the inmate and staff databases was recoded into a dummy variable comparing non-minorities (whites) and minorities. Additionally, for the staff database, an eleven-category variable measuring job category was collapsed into correctional officers, non-correctional officers. For each data source, three imputed databases were produced for the analyses that follow.

Step 1. The inmate and staff data were modeled with the HLM package developed by Raudenbush, Bryk, Cheong, and Congdon (2000) as described in Bryk and Raudenbush (1992). While HLM models are estimated in one step, the models can be thought of as existing at different levels. At level-1, the outcomes observed for each individual are modeled as a function of the covariate pattern for that individual. The following level-1 equation looks very much like an OLS model with some important exceptions. First, error and associated variance has been separated into two components, only one of which is shown in the level-1 equation. The other component of variance, represented by u_{0j} in the level-2 equation, is associated with the clustering unit, in this case each prison. It is used to measure the systematic differences between prisons and their impact upon individual-level evaluations. The error component shown at level-1, r_{ij}, is the typical error associated with each individual, stripped of the influence of the clustering unit. The *i* subscript references the *i* individuals with the *j* units.

Level 1: $Y_{ij} = \gamma_{00} + \Sigma\beta X + r_{ij.}$

The second level of an HLM model specifies how the level-1 random coefficients are modeled as a function of the characteristics of the clustering unit. In the models examined here, the only level-1 coefficient that is treated as random is the intercept. As such, the intercept for each institution is modeled by the following level-2 equation. The estimate of u_{0j} is important for this analysis as it is the estimate of the amount by which the "average" response for the outcome variable is raised or lowered by the respondent being located at institution *j*.

Level 2: $\beta_{0j} = \gamma_{00} + \Sigma\gamma W + u_{0j.}$

The equations presented above were used where the outcome variables were treated as quasi-continuous. While none of the outcome measures we used were truly continuous, previous experience has demonstrated that Likert items with at least seven categories work well when treated as continuous measures. Most of our items met this criteria (see Table 1). Several of the outcome measures, though, were dichotomies where respondents indicated their agreement or not with a survey item. For models of these items, a nonlinear specification was used for the model. Conceptually, the models still existed at two levels, but level-1 of the model estimated a logit, or the log odds of agreement with the statement. Since the estimate of the variance of a probability or proportion used to compute a logit is a function of the probability/proportion, there is no independent estimate of variance available for the individual-level error term, the r_{ij}. This means that a measure of the intra-class correlation (ICC) could not be computed. The ICC was computed for the quasi-continuous measures and can be thought of in either of two ways. First, ICC can be viewed as the amount of variance that exists in the dependent measure at the level of

the organization. ICC can also be thought of as the correlation between responses associated with the nesting of respondents within the respective institutions, e.g., inter-rater reliability.

The matrix of individual-level covariates, **X**, differed in the respective analyses for inmates and staff. Different information was collected for each group. For inmates, the individual-level controls included dummy variables indicating Hispanic ethnicity, race (minorities coded 1), and whether the survey was taken in Spanish. Sex was not included as all inmates surveyed were male. The continuous variables used for inmates consisted of age, time in prison, time left to serve on sentence, time at current prison, number of federal prisons, number of state prisons, time spent in current housing unit, and the propensity score for being an IHP inmate. For staff, the individual-level controls included dummy variables for sex (females coded 1), Hispanic ethnicity, race (minorities coded 1), supervisory status (supervisors coded 1), college degree, being a correctional officer, working on the day shift, and not working at the main facility (such as at a prison camp). The only continuously measured variables for staff were age and years of BOP tenure. For parsimony, the individual-level controls were treated as nuisance variables impacting upon the u_{0j} estimates of institutional differences and are not presented for discussion.

The matrix of organizational-level covariates, **W**, also differed for inmates and staff. Because information on inmates was only collected at ten prisons, there were not sufficient degrees of freedom to model the random $\$_{0j}$ coefficients in the inmate models. As such, the level-2 model for inmates was simply $\beta_{0j} = \gamma_{00} + u_{0j}$. For staff evaluations, surveys were collected at 98 institutions, and the **W** matrix included controls for the security-level of the prison as well as region of the BOP. Region of the BOP corresponds roughly to U.S. geographical regions. With the exception of the effect of region, though, the inmate and staff models did not differ significantly because all of the inmate prisons surveyed were low-security prisons.

Because of the inability to model the random $\$_{0j}$ for inmate responses, partial results for staff data are presented where no level-2 covariates were included. Further, the limited models for staff data were restricted to the ten institutions for which inmate data were collected. This allows for at least an intuition about how inmates results would look if the surveys were distributed to more institutions.

Step 2. Once measures were identified with desirable organizational properties, those measures were examined to see if there was correspondence between the organizational measures developed from inmate and staff data, the u_{0j} terms from the HLM models. The u_{0j} terms that were generated from staff data for the ten prisons examined in this step were taken from models of all 98 prisons. First, correlations were computed for the respective u_{0j} terms. For example, there were two measures of problems of sanitation in the dining hall (DIRTDINE) for each prison, one calculated from inmate data and one from staff data. The correlations assessed the congruence between measures developed from the different sources. The actual metric of the u_{0j} terms were compared with a Pearson correlation coefficient, and the rankings of the institutions produced by the u_{0j} terms were assessed with the nonparametric Spearman's rho coefficient. In addition to examining the Pearson and Spearman correlation coefficients, bar graphs were generated as an aid in visualizing the correspondence between measures of organizational performance that were developed from inmate and staff data.

Step 3. William Saylor and Schoeneck Howell of the Office of Research and Evaluation have worked over the past decade to develop better means of presenting information on measures of prison performance. They have consistently tried to simplify the presentation of information so that more data can be grasped at one time by BOP managers. One graphical method they have developed, building upon the work of Hartigan (1975), Inselburg (1990), and Wegman (1990),

for such a display of information is the technique called Parallel Coordinates Display. The idea is rather straightforward. For a given number of institutions defined by the user, create horizontal axes for each measure selected. Then, plot the values recorded for each prison on the axes and join the values for each prison on adjacent axes with a vertical line. The result is a plot that easily demonstrates whether the prisons as a group or separately are ranked consistently by the measures. If the lines on the plot are generally parallel to one another, it is an indication that the prisons are consistently ranked on the measures in question. On the other hand, if the lines crisscross one another, this is an indication of inconsistency. The software developed to realize this notion, PCDisWin, was used to examine the u_{0j} values produced by the HLM runs in step 1 (Howell, 2001).

Step 4. The information gleaned from the previous steps was used to compare the operations at one private prison, Taft Correctional Institution (TCI), with the operations of the 9 BOP prisons at which inmate surveys were gathered. In particular, TCI was compared with three other BOP prisons, FCI Elkton, FCI Forrest City, and FCI Yazoo City, that were built upon identical architectural footprints and activated at about the same time as TCI (cite). Primarily, these comparisons were generated from information produced in steps 2 and 3 above.

Results

The results of the analyses of the organizational properties of the different outcome variables, step 1 of the analysis, are presented in Table 2. For all measures, the reliability of the ranking of the institutions that was based on the deviations from the intercept or the typical response, the u_{0j}, is presented. While there is no technical cut-off value, most analysts consider as acceptable a measure with a reliability of at least .70. For those outcome variables that were measured with more than two categories (quasi-continuous variables), it was also possible to

calculate the ICC. Obviously, the most desirable measures for making institution comparisons are those with high values for ICC and reliability.

Step 1. Identify Organizational Properties of Measures

The results presented in Table 2 strongly suggest that there is merit to using inmates to evaluate conditions at prisons. A striking finding is that if we had relied only upon staff working at the ten prisons where inmate surveys were given, then we would have not been able to conclude that any of the measures had desirable organizational properties. None of the reliability scores for measures generated with staff data from the limited number of prisons exceeded 0.7. In fact, almost all of the reliability scores were well below the cutoff. We also encountered difficulty in generating estimates for several of the nonlinear models. At least partially due to the low reliability in measuring the random component associated with the intercept, the u_{0j}, many of the nonlinear models for the limited staff data failed to converge in 3,000 iterations for at least one of the multiply imputed databases. Since the results from the other databases clearly suggested low levels of reliability for the measures, and because similarly low levels of reliability were noted for the measures when using staff data from all institutions, the problematic models were not "fixed" in any sense and the results were reported in Table 2 for those models where there was convergence to a solution.

Likewise, the reliability values for the measures computed from staff data for all 98 prisons were generally low. There were only two measures, the measure of sanitation in the dining halls (DIRTDINE) and the measure of sanitation in the housing unit (DIRTUNIT), that had sufficient levels of reliability to be used as organizational-level measure. Even for these two measures, the reliability values for the measures derived from inmate data were higher, as were the ICC values. Nonetheless, since these two measures could be reliably used when generated from either inmate

or staff data, they were used in the next step of the analysis (presented below) to assess whether there was general congruence between inmate and staff evaluation of sanitation.

The reliability values for measures computed from inmate survey data were typically acceptable. Of the twenty-four outcome measures examined here, a reliability of at least .7 was noted for sixteen of the measures. In fact, most of these measures had reliability values in excess of .8, with some going as high as .97. Given the fact that there were only ten institutions, this finding is remarkable, especially in comparison to the general failure of any of the measures computed with staff data to exceed a reliability of .7 for these same ten prisons. All of the inmate measures that did not meet the general cutoff for reliability were items that asked about security and safety conditions at the prison. With perfect hindsight, it would be easy to speculate about why these measures were not reliable, but we resist the temptation. Instead, we note that there were five security and safety condition questions that were reliable.

Step 2. Congruence between Inmate and Staff Evaluations

Table 3 and Figures 1 and 2 turn attention to the next issue addressed by this analysis, whether inmates and staff provide congruent responses when evaluating prison conditions. The obvious caveat to this portion of the analysis was that we only found two questions with which to compare measures generated from inmate and staff data. Nonetheless, the results presented in Table 3 and Figures 1 and 2 suggest that inmates and staff were highly congruent in their evaluations. From Table 3, we see that the correlations, both the parametric, Pearson and the non-parametric, Spearman correlation coefficients suggested a close match between inmate and staff evaluations. For the sanitation in the dining hall question, the respective coefficients were .741 amd .661. Both coefficients were statistically significant even with scores on only ten prisons. For

the item about sanitation in the housing unit, the correlations were even stronger, .900 and .855. Again, both correlations were statistically significant.

The relationship between the organizational measures generated from inmate and staff data are illustrated nicely in Figures 1 and 2. As can be seen there, it was almost universally the case that if inmate evaluations were effected in a positive direction that staff values were as well. There were two exceptions noted in the graphs, but both of these exceptions were for prisons where both inmates and staff provided evaluations that were near the expected value, see FOR (FCI Forrest City) in Figure 1 and YAZ (FCI Yazoo City) in Figure 2. An interesting feature of these graphs is that they reflect why the reliability coefficients were greater measures generated from inmate data for these measures. As can be easily seen, the inmate deviations from the overall expected value are much greater than the staff deviations. It is important to keep in mind that a positive deviation from the expected is not a desirable feature in these graphs. A positive value indicates that the evaluations at that institution were pushed to where inmates and/or staff were more inclined to agree that sanitation was a problem. In Figure 1, for example, conditions at BAS led inmates to indicate a problem with sanitation in the dining hall, controlling for all individual-level characteristics of the inmates included in the model. At WAS, on the other, conditions led to more favorable evaluations as evaluations of dining hall sanitation were systematically lower than expected. In Figure 2, we see that the institutions with systematically higher indications of sanitation problems in the housing units were ASH and BAS. ALF, COL, and WAS stood out as having lower than expected evaluations.

Step 3. Assessing Congruence of Subject Area Measures

The evaluation of whether the different measures were congruent for the topical areas noted in Tables 1 and 2 was accomplished with Figure 3 through 6. Generally speaking, the

different measures did tend to hold together. Figure 3 presents the results for the three measures of gang activity. With the exception of several of the institutions which were in the center of the graph, the plot demonstrates that prisons high on one measure tended to be high on the others and vice versa. Plus, the institutions at the extremes tended to hold their rank in comparison to the other prisons. One thing to keep in mind when looking at the plots is that the dichotomous and quasi-continuous variables are both plotted on the same graph. They are easy to distinguish as the dichotomies always range between 0 and 1, demonstrating what proportion of the inmates at the given facilities answered affirmatively to the measure under examination. For example, almost 48 percent of the "typical" inmates (.48 multiplied by 100) at TAF (Taft) claimed that there was gang activity at the facility.[2] The quasi-continuous measures can take on both positive and negative values. In the case of these measures, the value is the amount above or below the "mean" value the institution fell. For example, the institution represented by the line furthest to the right in Figure 3 is approximately .85 units above the expected value for rating danger to gang members (GANGSAFE). This is a substantial amount of distance above the expected value for a measure that has only seven values.

Figure 4 shows the same general consistency for the security and safety measures. The graph has a funny bend between SAFEHIT and ENUFSTFD, but that is really a function of the change from a quasi-continuous metric to a proportion. Figure 5 seems to show the least consistency of the subject area measures examined so far. But this first perception is somewhat misleading. It appears that there are two factors underlying the measures plotted here. The first two measures

[2]"Typical" is defined by the specification of the model, but it is the proportion of those respondents who were coded 0 on all of the independent variables considered in the analysis. In this analysis, that would have meant a non-Hispanic, non-minority, respondent who took the survey in English. The person would have been of average value for age, time in prison, time left to serve on sentence, time at current prison, number of federal prisons, number of state prisons, time spent in current housing unit, and the propensity score for being an IHP inmate. All of these variables were centered on the mean for all respondents.

are clearly sanitation measures, and the institutions hold their respective rankings well. The other factor appears to be noise at the institution. If we look at the lines connecting EVENOISE and SLPNOISE, it can be seen that the measures are fairly congruent. There does appear to be incongruity between sanitation and noise at an institution. Figure 6 shows that there is stability in the rankings of food service. Not only is there stability, there is greater separation (distance) between highly and lowly ranked prisons. The clear separation of these institutions was also captured in the reliability scores. The reliability values for the group of food service measures were the highest noted in this analysis.

Step 4. Comparing Taft and the BOP Prisons

The graphs produced for this analysis make it easy to compare the different prisons. In Figures 3 through 6, Taft and three of the BOP prisons were highlighted. These three prisons have been identified by the BOP as comparisons for Taft. In Figure 3, it is easy to identify that all of the comparison prisons and Taft fall in the middle of the pack when ranking gang-related aspects of prison operations. Taft is generally lower than FOR, ELK, and YAZ. The only exception to this pattern was that ELK rated slightly better than TAF on the danger to inmates.

Figure 4 shows that TAF ranked near the middle of all of the prisons with respect to safety and security concerns. The three BOP comparison prisons, on the other hand, generally had the worst ratings for these measures. Inmates at these facilities rated the institutions as higher than the other prisons both in terms of their own perceived safety from being hit or assaulted, and these same comparisons facilities tended to have the worst ratings for sufficient numbers of staff being available at the institution to protect inmates.

Figure 5 demonstrates that Taft experienced some problems with sanitation, at least according to the inmate evaluations. Taft was higher than almost all BOP prisons on the measures

of sanitation in the dining hall and housing units. The three BOP comparison prisons fell in the middle of the distribution of ten prisons. For noise at the institution, on the other hand, Taft ranked very well with the lowest rating for noise during the evening hours and the next to lowest rating for noise during sleeping hours. The BOP prisons produced mixed results, but noise was clearly more of an issue at these prisons. In fact, ELK had the highest score for noise being a problem during sleeping hours. YAZ was not far behind and FOR fell in the middle of the prisons for these measures.

Figure 6 demonstrates that Taft clearly had problems with food services. The rankings produced with the inmate data demonstrate that Taft rated less favorably than the BOP prisons as they had the highest ranking of problems for three of the four items, namely lack of quality for the food, lack of variety of food, and not enough food being served. For the fourth item, the lack of appeal or the appearance of the food, Taft had the second highest score, again indicating a problem. The BOP comparison prisons scored better than Taft, but they tended to be on the higher side of the scores or in the middle of the rankings of the prisons.

Discussion and Conclusion

Measures derived from inmate data deserve better treatment than the distrust they often receive from prison administrators. As demonstrated here, inmate data can be used to generate reliable measures that differentiate performance at prisons. Not only were the inmate measures much more reliable than measures developed from staff data, the inmate-derived measures correlated well with staff evaluations for the two measures of sanitation where comparisons were appropriate. The correspondence between inmate and staff evaluations on the sanitation items and the non-correspondence for the other survey questions clearly demonstrated two facts. First, where items of common interest to staff and inmates are used, inmates and staff appear to be

affected by similar situational factors in providing evaluations of institutional operations. Second, it is important to identify and develop survey items that are clearly relevant to and appropriate for the intended sample.

The parallel coordinates graphs demonstrated the ease by which multiple measures can be examined together. Even though the models used to construct the graphs were probably beyond the understanding of the typical correctional administrator, the results of the models are easily explained to nontechnical audiences. By dividing the 24 measures into subject areas, and limiting the graphs to subject areas, we were able to quickly examine how Taft compared to the other nine BOP prisons, but especially how Taft compared to the three comparison institutions that were very similar with respect to physical plant. The findings suggested that Taft was neither better nor worse than BOP prisons in a consistent manner. In several areas, Taft performed about average in comparison to all BOP prisons, most notably on the gang measures and the safety and security concern measures. In other areas, Taft had some problematic findings. There were problems noted for the measures of sanitation in both the dining hall and housing units, and there were more serious problems noted for the food service measures where Taft had the most problematic ranking for three of the four measures. For the fourth measure of food services, Taft had the next to worst rating. While we would have liked to provide a simple solution to the question of whether private prisons perform better or worse than public prisons, our results suggest that the answer is more complex. We would argue that private prisons need the same level of monitoring as public prisons to insure adequate performance.

The BOP comparison prisons also had some problematic findings that may have been related to the competition between them and Taft. In particular, the finding that the institutions generally rated worst in terms of providing an environment in which inmates felt safe from being

hit or assaulted was noteworthy, especially since these institutions also rated lowest in terms of inmates believing that there were enough staff during the different shifts to provide for their safety. This is an interesting finding because Nelson has documented that the BOP comparison facilities operate with fewer staff than Taft (Nelson, 1999). The BOP comparison prisons also tended to have bad ratings for the food service measures, although clearly Taft was more problematic in this area. These findings are significant as the money spent on food is more easily manipulated than other parts of a prison budget such as wages for staff, even though as part of the overall prison budget food service spending is small. In estimates of BOP spending at comparison prisons in 1999, Nelson found that the BOP spent $2.63 on food for each inmate per day out of a total per diem cost of $37.14.

Taft was not different from the BOP comparison prisons in a generally consistent manner. Taft did exhibit some troubling outcomes, as did the BOP comparison prisons. Without complete consensus among all measures used, we could not conclude that on the whole Taft did better or worse than the BOP comparison prisons. While the prisons clearly differed on a number of measures, we did not develop an overall rating of the institutions with some combination of the different aspects of prison operations examined here. In part, the failure to provide an overall judgement of performance of Taft and the BOP comparison prisons was methodological, and part of the failure was substantive. From the substantive side, subject matter experts need to determine which factors of prison operations are more important than others for ranking prisons. Food service, where Taft ranked poorly, is generally accepted as a key area of institutional operations, but so is safety and security where the BOP comparison prisons fared poorly. Methodologically, more work is necessary to document the correspondence between subjective measures of prison performance and operational outcomes, such as inmate misconduct or the success of prison

operations and programs. With a better substantive understanding and related methodological work, we think it will be possible to develop more systematic overall evaluations.

We would never argue for sole reliance upon survey data, whether collected from inmates or staff, to assess performance at prisons. But given the relative ease and low expense associated with surveys, we do argue for using them as part of a more comprehensive strategy for comparing prisons.

REFERENCES

Archambeault, W. G., & Deis, D. R. (1996). *Cost Effectiveness Comparisons of Private Versus Public Prisons in Louisiana: A Comprehensive Analysis of Allen, Avoyelles, and Winn Correctional Centers*. Baton Rouge: Louisiana State University.

Austin, J., & Coventry, G. (2001). *Emerging Issues on Privatized Prisons*. Washington, D.C.: Bureau of Justice Assistance.

Bowery, M. (1997). *Private Prisons in NSW: Junee - Year Three* (38). Sydney, Australia: NSW Department of Corrective Services.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage.

Camp, S. D. (1999). Do Inmate Survey Data Reflect Prison Conditions? Using Surveys to Assess Prison Conditions of Confinement. *The Prison Journal, 79*(2), 250-268.

Camp, S. D., & Gaes, G. G. (2000). Private Adult Prisons: What Do We Really Know and Why We Don't Know More? In D. Shichor & M. Gilbert (Eds.), *Privatization of Criminal Justice: Past, Present and Future* (pp. 283-298). Cincinnati, OH: Anderson.

Camp, S. D., Gaes, G. G., & Saylor, W. G. (2001). Quality of Prison Operations in the U.S. Federal Sector: A Comparison with a Private Prison. *Punishment & Society, Forthcoming*.

Camp, S. D., Saylor, W. G., & Harer, M. D. (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*(4), 739-761.

Camp, S. D., Saylor, W. G., & Wright, K. N. (1999). Creating Performance Measures from Survey Data: A Practical Discussion. *Corrections Management Quarterly, 3*(1), 71-80.

Carlson, J. R. (1997). Differential Perceptions of Women Inmates, Correctional Staff, and Correctional Officers Regarding Rehabilitation Program Modalities. *Journal of Offender Rehabilitation, 25*(3/4), 159-170.

Edgar, I. O. D. a. K. (1998). Routine Victimization in Prisons. *The Howard Journal, 37*(3), 266-279.

Gaes, G. G., Camp, S. D., & Saylor, W. G. (1998). Appendix 2: Comparing the Quality of Publicly and Privately Operated Prisons: A Review. In D. McDonald & E. Fournier & M. Russell-Einhorn & S. Crawford (Eds.), *Private Prisons in the United States: An Assessment of Current Practice* (pp. 1-38). Boston, MA: Abt Associates Inc.

General Accounting Office. (1996). *Private and Public Prisons: Studies Comparing Operational Costs and/or Quality of Service*. Washington, D.C.: U.S. General Accounting Office.

Hartigan, J. A. (1975). *Clustering Algorithms*. New York: John Wiley & Sons.

Howell, S. (2001). Parallel Coordinates Display (Version 2.0). Washington, D.C.: Federal Bureau of Prisons.

Inselburg, A., & Dimsdale, B. (1990). *Parallel Coordinates: A Tool for Visualizing Multi-Dimensional Geometry.* Paper presented at the First IEEE Conference on Visualization.

Logan, C. H. (1992). Well Kept: Comparing Quality of Confinement in Private and Public Prisons. *The Journal of Criminal Law and Criminology, 83*(3), 577-613.

Maitland, A. S. a. S., R.D. (1996). Victimization in Prisons: A Study of Factors Related to the General Well-Being of Youthful Offenders. *Federal Probation, 60*(2), 24-31.

McCorkle, R. C. (1993). Fear of Victimization and Symptoms of Psychopathology among Prison Inmates. *Journal of Offender Rehabilitaion, 19*(1/2), 27-41.

Nelson, J. (1999). *Taft Prison Facility: Cost Scenarios*. Washington, D.C.: National Institute of Corrections. http://www.nicic.org/pubs/1999/015712.pdf

O'Donnell, K. E. a. I. (1998). Assault in Prison: The Victim's Contribution. *British Journal of Criminology, 38*(4 Autumn), 635-650.

Price Waterhouse. (1996). *1995 National Inmate Sruvey: Final Report.* Ottawa: Price Waterhouse.

Raudenbush, S., Bryk, A., Cheong, Y. F., & Congdon, R. (2000). HLM 5: Hierarchical Linear and Nonlinear Modeling (Version 5.03). Lincolnwood, IL: Scientific Software International, Inc.

Robinson, M. L. a. D. (1996). *Attitudes of Correctional Officers Towards Offenders* (Executive Summary). Ontario: Price Waterhouse.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. New York: Chapman & Hall.

Sechrest, D. K., & Shichor, D. (1996). Comparing Public and Private Correctional Facilities in California: An Exploratory Study. In G. L. Mays & T. Gray (Eds.), *Privatization and the Provision of Correctional Services.*

Tennessee Select Oversight Committee on Corrections. (1995). *Comparative Evaluation of Privately-Managed Corrections Corporation of America Prison (South Central Correctional Center) and State-Managed Prototypical Prisons (Northeast Correctional Center, Northwest Correctional Center)*. Nashville, TN: Tennessee Select Oversight Committee on Corrections.

Thomas, C. W. (1997). *Comparing the Cost and Performance of Public and Private Prisons in Arizona*. Gainesville, FL: Center for Studies in Criminology & Law, University of Florida.

Wegman, E. J. (1990). Hyperdimensional Data Analysis Using Parallel Coordinates. *Journal of the American Statistical Association, 85*(411), 664-675.

Figure 1. Dining Hall Sanitation Problems

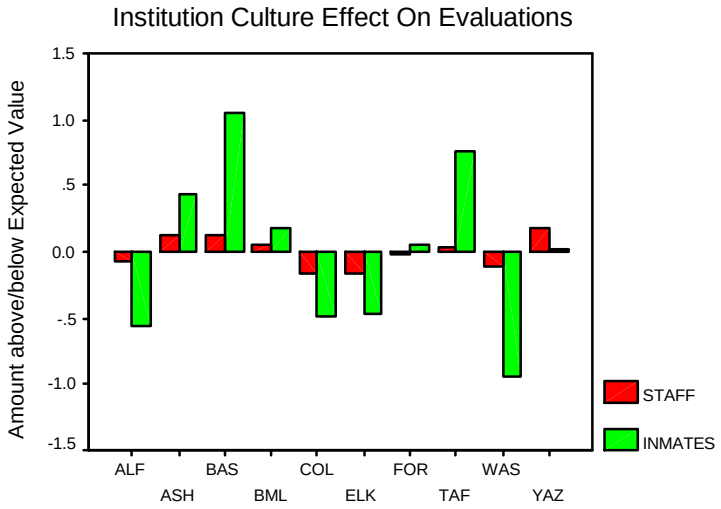

Figure 2. Housing Unit Sanitation Problems

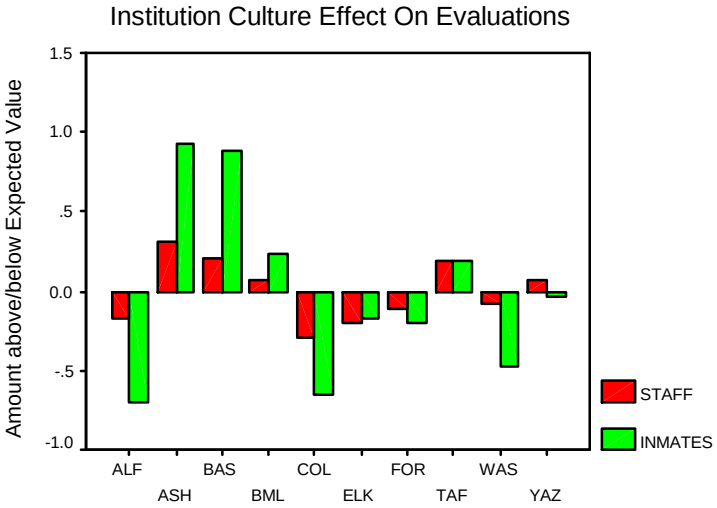

Figure 3. Gang-Related Activity

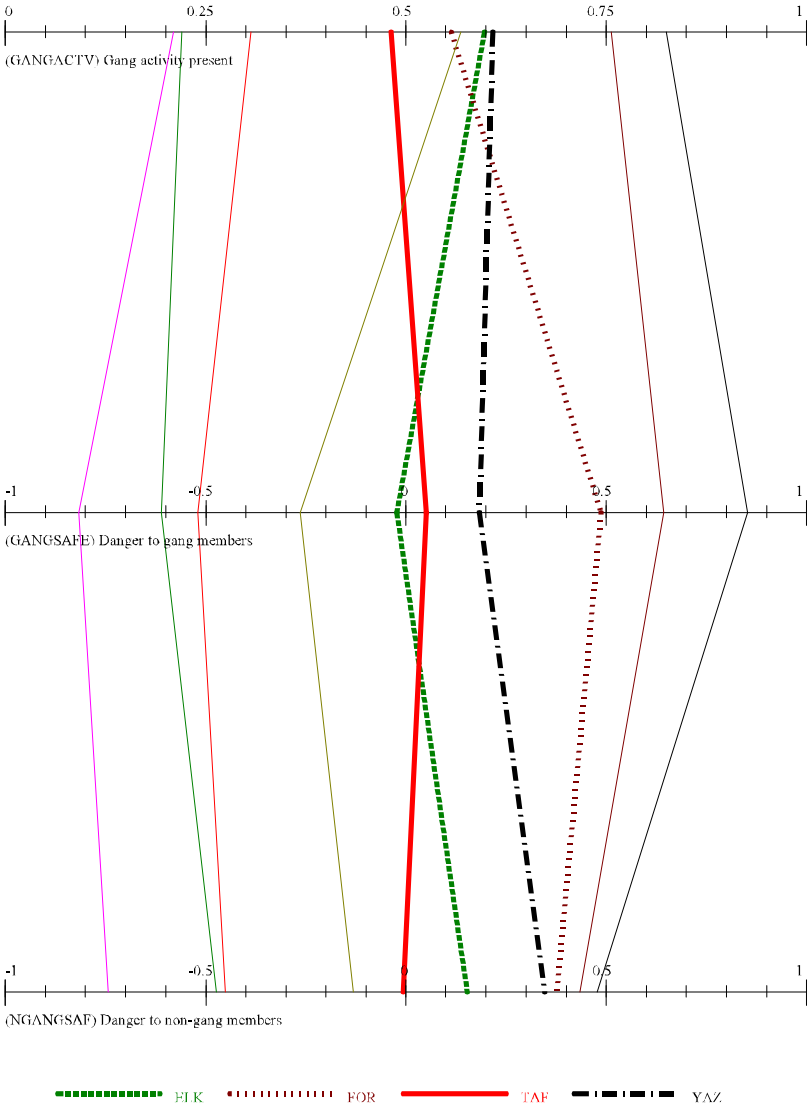

Figure 4. Security/Safety Concerns

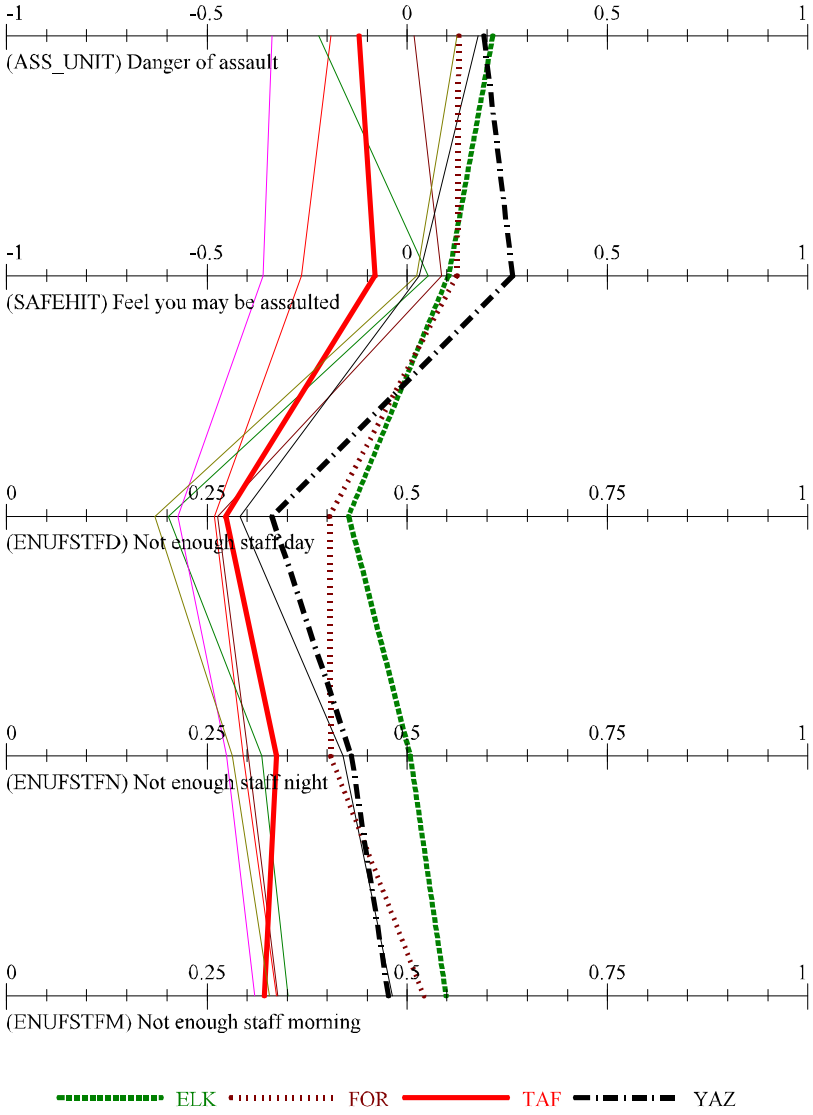

Figure 5. Overall Sanitation/Conditions

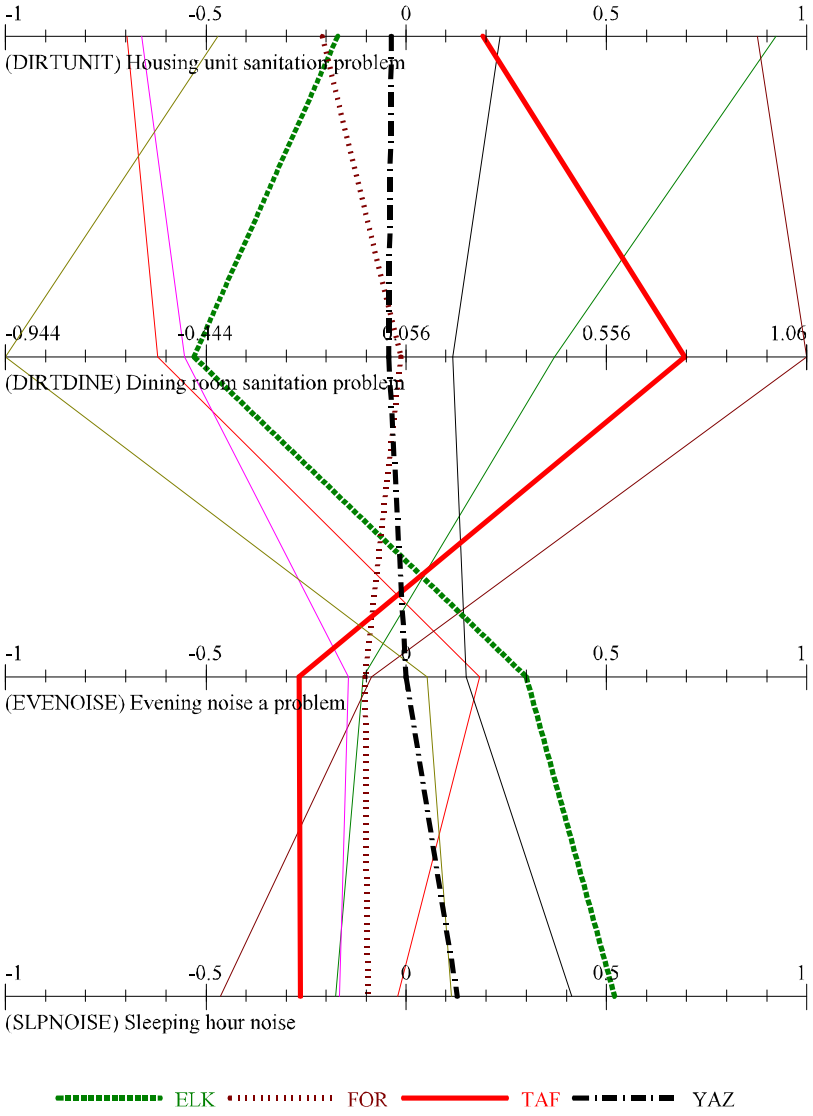

Figure 6. Food Service Delivery

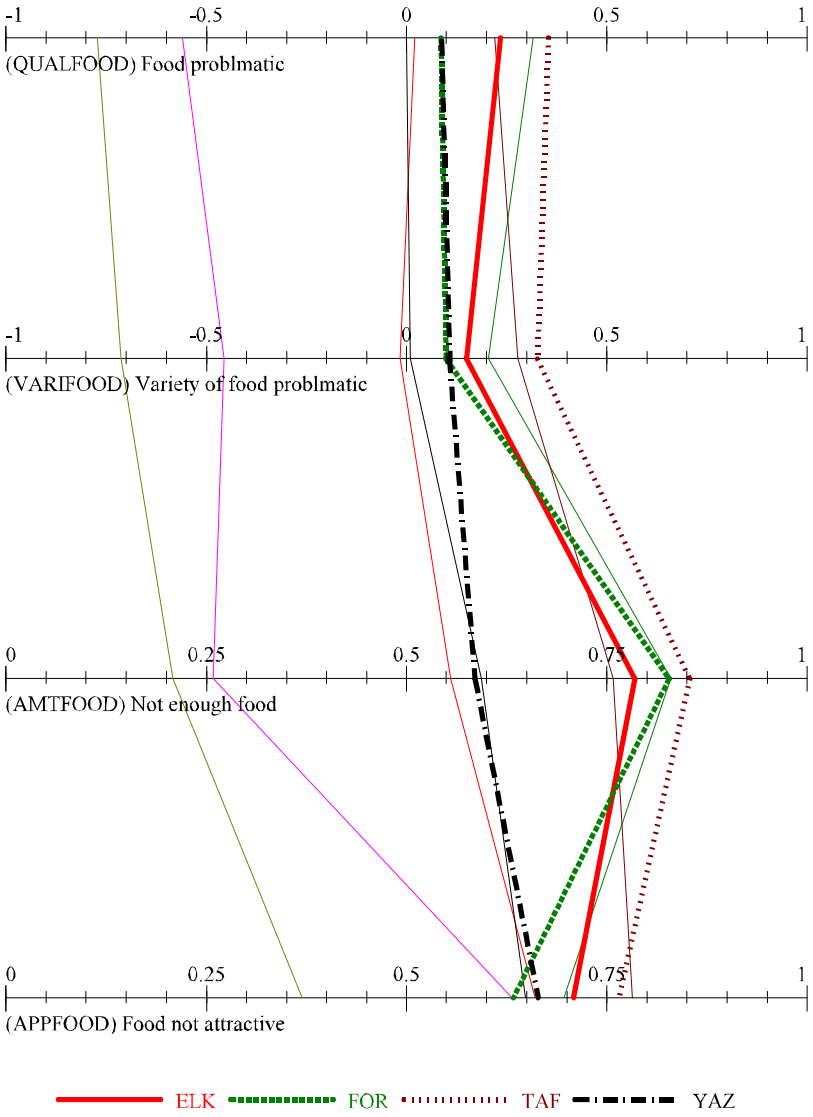

Table 1. Questions Analyzed

Variable Name	Wording of Question	Staff Version	Differences	Cate-gories*
Gang-Related Issues				
GANGACTV	Has there been any gang activity in this prison during the past 6 months?	Yes	No Knowledge Choice for Staff	2/3
GANGSAFE	How safe or dangerous do you think it has been in this prison for inmates who ***are*** members of a gang?	Yes	None	6
NGANGSAF	How safe or dangerous do you think it has been in this prison for inmates who ***are not*** members of a gang?	Yes	None	6
Security/Safety Concerns				
ASS_UNIT	How likely is it that an inmate would be assaulted in his/her living unit?	Yes	Slight Wording	4
SAFE_HIT	How safe do you feel from being hit, punched, or assaulted by other inmates?	No		5
SAFE_PRP	How safe do you feel your property has been?	No		5
YUASSLTD	Have you been physically assaulted by an ***inmate*** within the last 6 months?	No		2
STOPSEX	Staff have prevented forced sex among inmates.	No		6
ENUFSTFD	Do you think there have been enough staff here to provide for the safety and security of ***inmates***: During the *day* (8 am - 4 pm) shift?	Yes	No Knowledge Choice for Staff	2/3
ENUFSTFN	Do you think there have been enough staff here to provide for the safety and security of ***inmates***: During the *night* (4 pm - midnight) shift?	Yes	No Knowledge Choice for Staff	2/3
ENUFSTFM	Do you think there have been enough staff here to provide for the safety and security of ***inmates***: During the *morning* (midnight - 8:00 am) shift?	Yes	No Knowledge Choice for Staff	2/3
SAFSTFMA	How safe do you think it is for ***male*** staff members who have frequent contact with inmates in this prison?	Yes	Slight Wording	6
SAFSTFFE	How safe do you think it is for ***female*** staff members who have frequent contact with inmates in this prison?	Yes	Slight Wording	6

Table 1. Questions Analyzed, Continued

Variable Name	Wording of Question	Staff Version	Differences	Categories*
SEARCHES	How often have there been shakedowns in this institution during the past 6 months?	Yes	Slight Wording, No Knowledge Choice for Staff	7/8
STRPSRCH	How often have you been strip or pat searched at this institution during the past 6 months (*not* including those required for visits)?	No		7
SEARFREQ	Are the shakedowns done frequently enough?	Yes	None	2
Overall Sanitation/Conditions				
DIRTUNIT	How often have insects, rodents, dirt, or litter been a problem in the housing units?	Yes	No Knowledge Choice for Staff	7/8
DIRTDINE	How often have insects, rodents, dirt, or litter been a problem in the dining hall?	Yes	No Knowledge Choice for Staff	7/8
EVENOISE	How noisy has it been in your housing unit during the evening hours?	Yes	Wording, question and choices	5
SLPNOISE	How noisy has it been in your housing unit during sleeping hours?	Yes	Wording, question and choices	5
Food Service Delivery				
QUALFOOD	The quality of food at this prison has been: (poor, fair, good)	No		3
VARIFOOD	The variety of food at this prison has been: (poor, fair, good)	No		3
AMTFOOD	The amount of food served for main courses has been (Not enough, Enough)	No		2
APPFOOD	The appearance of the food at this prison has been: (Unappealing, Appealing)	No		2

Table 2. Results for Inmate and Staff Administrations of the PSCS in 2000

	Inmate PSCS[1]		Staff PSCS (Limited)[2]		Staff PSCS (All, Full)[3]	
Variable Name	Reliability	ICC	Reliability	ICC	Reliability	ICC
Gang-Related Issues						
GANGACTV	.939	ND[4]	.614[5]	ND	.583	ND
GANGSAFE	.936	.134	.018	.000	.299	.005
NGANGSAF	.895	.084	.052	.001	.284	.004
Security/Safety Concerns						
ASS_UNIT	.848	.056	.125	.002	.411	.008
SAFE_HIT	.706	.025	NA[7]	NA	NA	NA
SAFE_PRP	.409	.008	NA	NA	NA	NA
YUASSLTD	.443	ND	NA	NA	NA	NA
STOPSEX	.672	.023	NA	NA	NA	NA
ENUFSTFD	.764	ND	.363[6]	ND	.157	ND
ENUFSTFN	.721	ND	.116[6]	ND	.437	ND
ENUFSTFM	.812	ND	.357[5]	ND	.344	ND
SAFSTFMA	.676	.022	.011	.000	.367	.006
SAFSTFFE	.618	.017	.015	.000	.361	.006
SEARCHES	.662	.021	.026	.000	.233	.003
STRPSRCH	.596	.016	NA	NA	NA	NA
SEARFREQ	.089[5]	ND	.259	ND	.270	ND
Overall Sanitation/Conditions						
DIRTUNIT	.919	.107	.593	.017	.802	.042
DIRTDINE	.930	.123	.314	.006	.803	.042
EVENOISE	.761	.033	.540	.013	.648	.020
SLPNOISE	.852	.057	.659	.024	.567	.014
Food Service Delivery						
QUALFOOD	.978	.316	NA	NA	NA	NA
VARIFOOD	.971	.262	NA	NA	NA	NA
AMTFOOD	.952	ND	NA	NA	NA	NA
APPFOOD	.831	ND	NA	NA	NA	NA

TABLE 2 NOTES:

1. N=950
2. N=916
3. N=9,491
4. Not defined. ICC cannot be computed for binomial models where there is no estimate of individual error, r_{ij}.
5. Estimate based on results for only one of the three multiply imputed databases. After 3,000 iterations, HLM was unable to derive a solution for the other two databases. This result should be treated with caution.
6. Estimate based on results for two of the three multiply imputed databases. After 3,000 iterations, HLM was unable to derive a solution for the third database.
7. Not available. The question was not asked of staff respondents.

Table 3. Correlations between Inmate- and Staff- Based Evaluation of Prison Operations
Where both Inmate and Staff Data Produce Reliable Measures

Variable Name	Pearson (metric of u_{0j})		Spearman rho (rankings of u_{0j})	
DIRTINE	.741	p=.014	.661	p=.038
DIRTUNIT	.900	p=.000	.855	p=.002