Measurement and Analysis: What Can and Does Go Wrong?*

Maureen Brown
University of North Carolina
brown@unc.edu

Dennis Goldenson
Software Engineering Institute
dg@sei.cmu.edu

Abstract

By now, one can point to many instances where measurement has been used effectively to inform management and technical decisions in support of the development and maintenance of software and software intensive systems. Yet, measurement is not very well integrated into software or systems engineering education or practice, and measurement remains challenging for all too many organizations. For this paper, we have analyzed 1350 findings drawn from 663 Software CMM® appraisals that were conducted between 1987 and 2002 inclusive. The results are augmented by questions from a survey of CIO's from state and local governments and the private sector. Our analyses suggest several areas where both managers and engineers would benefit from better guidance about the proper use of measurement and analysis. Future work may include lexical analyses based on natural language processing as well as studies of appraiser understanding of the measurement content in CMMI® models

1. Introduction

A disciplined approach to measurement and data analysis can be crucial for the success of a software or systems engineering enterprise [1] [2]. By now, one can point to many instances where measurement has been used effectively to inform management and technical decisions based on facts and objective evidence. By definition, high maturity organizations incorporate measurement into their day to day and strategic activities. Six sigma programs are becoming more and more common. The incorporation of Measurement and Analysis as a distinct process area in Capability Maturity Model® Integration (CMMI) models provides the management visibility and focus that organizations have needed to guide the use of measurement in their process improvement efforts. Still, measurement is not very well integrated into software or systems engineering education or practice, and measurement remains challenging for all too many organizations.

For this paper, we have analyzed 1350 finding drawn from 663 SW-CMM appraisals that were conducted between 1987 and 2002 inclusive. The findings are classified into several categories of common weakness and opportunities for improvement. These results are augmented by questions from a survey of CIO's from state and local governments and the private sector.

Our analyses suggest several areas where both managers and engineers would benefit from better guidance about the proper use of measurement and analysis. Future work may include further lexical analyses based on natural language processing as well as studies of appraiser understanding of Measurement and Analysis.

2. Background and method

2.1. CMM models and appraisal methods

Before describing the appraisal findings data that we have analyzed, a brief review of CMM models and appraisal methods may be instructive for some readers. The findings we analyzed are from appraisals of how well organizational processes correspond with the practices of the Capability Maturity Model for Software (SW-CMM) [3]. The SW-CMM is the first of several best practices models that have recently been incorporated into a family of models based on the Capability Maturity Model Integration (CMMI) framework [4][5].

We have limited our analysis to SW-CMM appraisal findings. The results of many more SW-CMM than CMMI appraisals had been reported to the Software Engineering Institute (SEI℠) at the time of

* Capability Maturity Model, Capability Maturity Modeling, Carnegie Mellon, CMM, and CMMI are registered in the U.S. Patent and Trademark Office by Carnegie Mellon University. CMM Integration, SCAMPI, SCAMPI Lead Appraiser, and SEI are service marks of Carnegie Mellon University. The authors thank Bob Ferguson, Shannon Schelin, Kenny Smith, Mike Zuccher, and Dave Zubrow for their respective contributions.

this writing. Moreover, the treatment of measurement is more explicit in the CMMI models [1].

The SW-CMM organizes software best practices into five maturity levels, each of which represents an evolutionary step that lays a foundation for continuous process improvement to the next successive maturity level [Table 1]. The model covers key practices for planning, engineering, and managing software development and maintenance. More mature software organizations are expected to be better able to meet their cost, schedule, functionality, product quality, and other performance objectives [3].

Table 1. Maturity levels and their key process areas (Source, Paulk [6])

Level	Focus	Key Process Areas
Level 5 Optimizing	Continuous process improvement	- Defect Prevention - Technology Change Management - Process Change Management
Level 4 Managed	Product & process quality	- Quantitative Process Management - Software Quality Management
Level 3 Defined	Engineering processes & organizational support	- Organization Process Focus - Organization Process Definition - Training Program - Integrated Software Management - Software Product Engineering - Intergroup Coordination - Peer Review
Level 2 Repeatable	Project management processes	- Requirements Management - Software Project Planning - Software Project Tracking & Oversight - Software Subcontract Management - Software Quality Assurance - Software Configuration Management
Level 1 Initial	Competent people (and heroics)	

CMM appraisals are perhaps most widely known for their quantitative benchmarks of maturity level, Key Process Area (KPA) and goal satisfaction profiles; however, most appraisals also present textual findings that provide additional qualitative context and clarification. The findings are meant to aid in understanding the ratings and to guide subsequent process improvement. The findings are presented verbally in formal presentations to the sponsors and other appraisal participants. The verbatim findings typically are short enough to fit in overhead slide presentations; however, further clarification commonly is provided in the verbal presentations. A few characteristic examples of reported findings can be seen in Table 2.

Table 2. Typical measurement related findings from SW-CMM appraisals

Lack of a consistent approach for capturing quality and productivity measurement data and comparing actuals with forecasts.
There is no common understanding, definition and measurement of Quality Assurance
Test coverage data is inconsistently measured and recorded
Measurements of the effectiveness and efficiency of project management activities are seldom made.

2.2. Appraisal findings

The appraisal finings described here are all from either CMM-Based Appraisals for Internal Process Improvement (CBA IPI) assessments or the earlier Software Process Assessments (SPA) that were replaced by CBA IPI in 1996 [7]. The data that we analyzed are drawn from an SEI repository called the Process Appraisal Information System (PAIS). The PAIS database is a unique data source. It includes all appraisal results that are submitted in confidence to the SEI as part of its authorized lead appraiser program.

The findings that we describe are drawn from 2910 CBA IPI and SPA appraisals of the SW-CMM that were conducted from 19 February 1987 through 28 June 2003; the appraisals included 36316 findings that were recorded as "weaknesses" or "opportunities for improvement." Of those, 663 appraisals included 1350 weaknesses and opportunities for improvement that included the root word "measure."

We initially classified the measurement related findings into 48 categories. These include the measurement categories that make up the Practical Software and Systems Measurement (PSM) performance model [8], along with a few additional categories that accommodate findings that relate more directly to the structure of the SW-CMM.

Some of the findings include content that can be classified properly into more than one of the 48 categories; hence we have a total of 1,549 coded findings. The analyses reported here are based on the larger number of coded findings. Additional analyses, including breakdowns by appraised organization, model structure, and non model content as well as tests of inter-coder reliability are ongoing as of the time of this writing.

2.3. The CIO survey

We have augmented the appraisal findings with analogous data from a recent survey of Chief Information Officers. The survey was done as part of a doctoral thesis at the University of North Carolina. At our request, the survey included a short series of questions about difficulties encountered by the CIOs' organizations in implementing their software measurement activities.

The survey was administered to 174 public and private sector CIO's in January of 2004. The public sector sample was drawn from the membership roles of the National Association of State Chief Information Officers (NASCIO) [9], the International City/County Management Association (ICMA) [10] and the US CIO Council [11]. A total of 83 public sector CIO's completed the survey, resulting in a response rate of 40 percent. In addition, a random sample of 200 private sector Chief Information Officers was drawn from the Leadership Library database [12]. The Leadership Library contains approximately 856 CIO listings from a variety of organizations, with differing sizes, functional areas, and organizational designs. Fifteen surveys were returned with no forwarding address. A total of 95 responses were obtained for a response rate of 51 percent of the valid addresses.

Our questions first asked the CIO's about the extent to which their organizations relied on measurement to guide their system development and maintenance efforts. The respondents were then asked to indicate the level of difficulty they encountered in establishing and using a series of classes of measures that they are required to address by the Clinger-Cohen Act [13]. The respondents were asked to characterize their answers on a scale of 1 to 10, where 10 indicated highest reliance or difficulty respectively. As described in section 3.2 below, we also compared the respondents' replies by sector.

3. Results

3.1. Appraisal findings

The appraised organizations come from a variety of sectors Figure 1. Almost half are commercial firms that sell products including various kinds of software and software intensive systems. The SW-CMM maturity levels of the appraised organizations closely mirror the full PAIS database for the period from which our findings are drawn Figure 2.

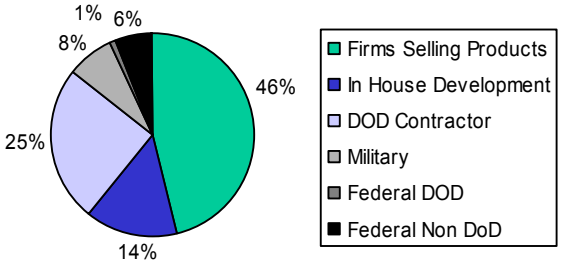

Figure 1. Appraised organizational sectors

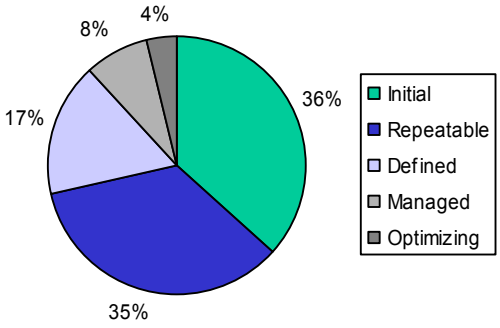

Figure 2. Maturity levels of the appraised organizations

We grouped the initial 48 categories into four composite sets of findings as shown in Figure 3. Over one third of the coded findings describe difficulties experienced by the appraised organizations in applying measurement to their management processes. Close to a third of the findings identify problems with the organizations' measurement processes themselves. A little over 20 percent of the findings describe problems faced by the organizations in using measurement to understand and improve their existing processes. The remaining findings describe problems in using measurement to characterize product quality and technical effectiveness.

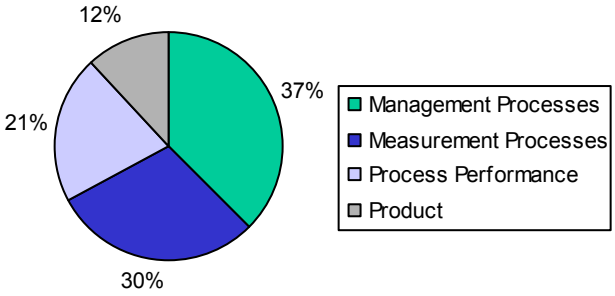

Figure 3. Grouped appraisal findings

Appraisal findings typically are arranged by KPA or other CMM model content. Not surprisingly, the largest of the four groups includes findings that address difficulties with, or lack of use, of measurement for management purposes.

As seen in Figure 4, 22 percent of the 582 weaknesses and opportunities for improvement that we classified under "management processes" focus on problems with measurement and analysis of quality assurance. These are followed closely by problems related to using measurement in project planning and estimation and to difficulties with measures of schedule and progress. Noticeable numbers of findings also address problems with the use of measurement in managing the appraised organizations' training and configuration management activities. The "other" category includes references to project management without further elaboration (38 instances), resources and cost (28 instances), policies (14 instances), risk (7 instances), and ROI concerns (2 instances).

As noted above, the findings in all six categories are closely coupled to the structure and content of the Software CMM. The five more homogeneous categories all map directly to the model's KPA structure. With the possible exception of the two references to measuring ROI, the findings in the other category map either to the KPA's or to the model's institutionalization common features.

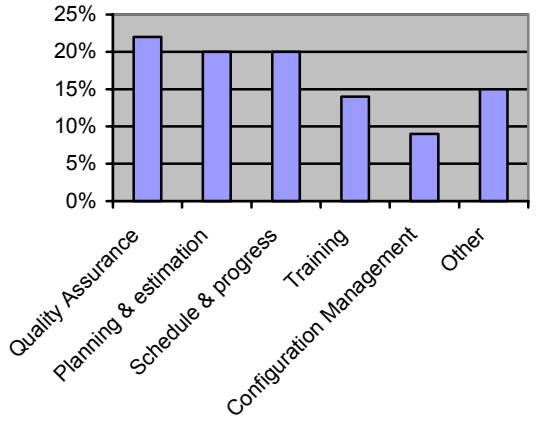

Figure 4. Weaknesses and improvement opportunities: Management processes

Of keen interest to those of us who specialize in measurement and analysis is the fact that the next largest category of findings includes weakness and opportunities for improvement in the measurement process itself. That result is particularly noteworthy since appraisers tend to focus on the structure and content of the model against which the organization is being appraised. While the SW-CMM does include a Measurement and Analysis common feature that applies to all KPA's, its measurement related content is considerably less explicit and complete than what now is included in the CMMI models [1][4][5].

As seen in Figure 5, 26 percent of the 461 coded findings of weaknesses and opportunities for improvement in the measurement process are statements about existing measures being inadequate for their intended purposes. As mentioned in Section 2.1, the findings statements tend to be rather short and terse; however, many or most of them seem to be identifying instances where measurement is poorly aligned with the business and technical needs of the appraised organizations. Problems of that kind speak directly to the practices that are addressed in the first goal of the CMMI Measurement and Analysis process area [1].

Similarly, 25 percent of the coded findings state that missing and incomplete data are a recurring problem in the appraised organizations; 15 percent complain about inconsistent use of the data that are collected, and 14 percent state that the measurement results are not used at all. The "other" category includes references to improvement of measurement processes (43 instances), inter group activities related to measurement (34 instances), measurements that are misunderstood or not understood (12 instances), and leadership in the organization (3 instances).

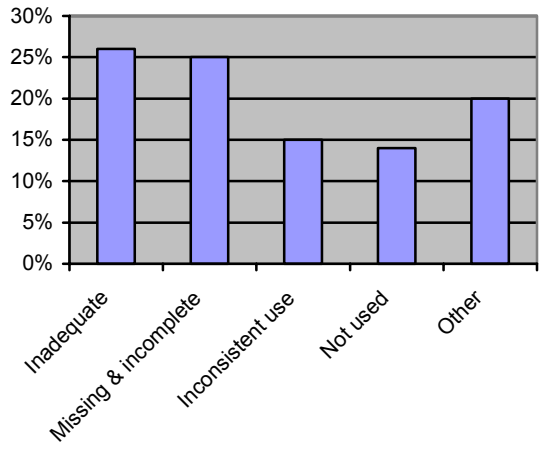

Figure 5. Weaknesses and improvement opportunities: Measurement processes

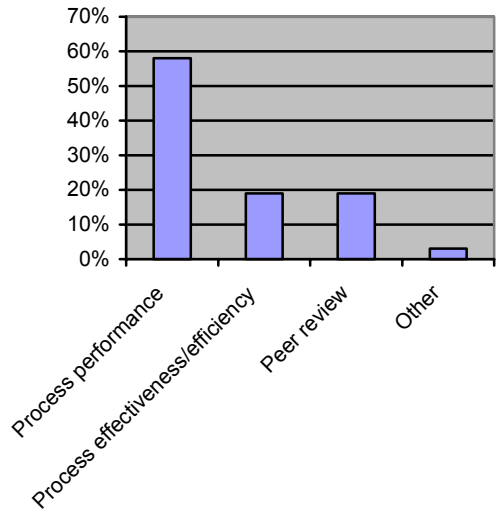

Figure 6. Weaknesses and improvement opportunities: Process performance

Process performance constitutes the third most common set of measurement related weaknesses and opportunities for improvement in the appraised organizations These findings describe problems faced by the organizations in using measurement to understand and improve their existing processes. As seen in Figure 6, well over half of the 319 coded comments we grouped here mentioned difficulties with measuring process performance in explicit terms. In addition, 19 percent of the coded findings make explicit reference to problems with measurement and analysis of process effectiveness or efficiency. Combined, these two sets of findings account for over three fourths of the total.

References to problems with using measurement to support peer reviews constitute another 19 percent of the coded findings that we grouped under process performance. The "other" category includes references to process compliance (5 instances), tool shortage (4 instances), incremental capability (1 instance), and personnel (1 instance).

These results are particularly noteworthy since the measurement of process performance is too often associated only with high maturity practices. As noted later in this section, however, it is clear that at least some organizations and appraisal teams are seriously addressing issues of process performance earlier on.

Of course, Peer Review is a maturity level 3 KPA in the SW-CMM, but concerns about process performance are often addressed by peer reviews. Similar issues also are raised in lower maturity organizations, often in the context of Software Project Tracking and Oversight or the institutionalization common features, particularly Measurement and Analysis.

The final group of 187 coded findings identifies weaknesses and opportunities for improvement related to the measurement of product quality and the extent of technical effectiveness that supports product development and maintenance. As seen in Figure 7, references directly to product quality account for 43 percent of these findings. Another 40 percent identify difficulties in the measurement of functional correctness, and 10 percent cite problems with measuring product size and stability. The other category includes references to customer satisfaction (4 instances), technical effectiveness (4 instances), reliability (3 instances), security (1 instance), supportability (1 instance), usability (1 instance), and technical volatility (1 instance).

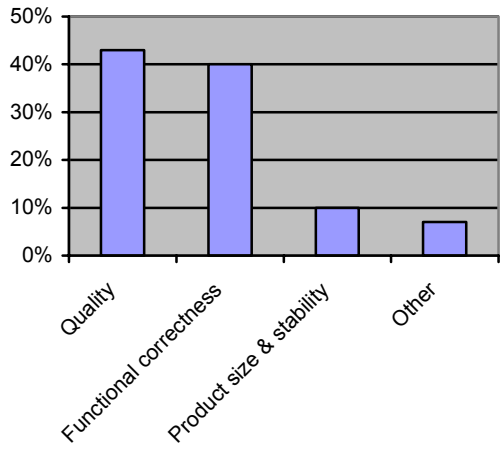

Figure 7. Weaknesses and improvement opportunities: Product

While these finding are only 12 percent of the total that we analyzed, they too are noteworthy, especially in a sample that includes so many lower maturity organizations. Of course a major reason for process improvement is the prospect that product quality and project performance will improve concomitantly; however the emphasis in CMM models and appraisals is on process and process adherence.

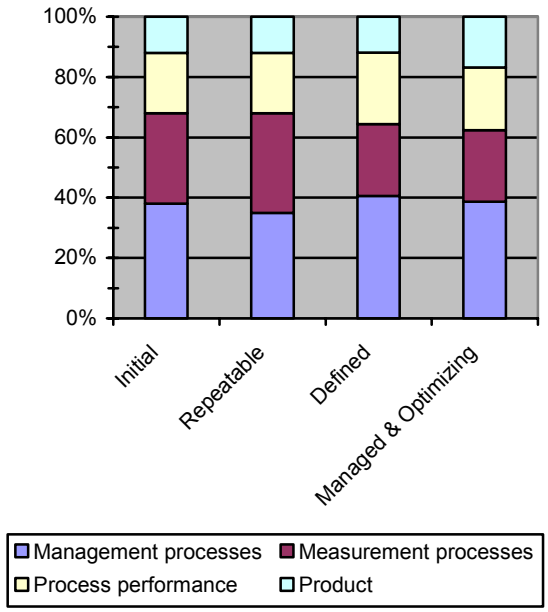

Figure 8. Weaknesses and improvement opportunities by maturity level

As mentioned earlier in this section, we also compared the categorizations of our coded findings by maturity level of the appraised organizations. Notice in Figure 8 that all four groups of findings remain problematic across the maturity levels. That is true with respect to difficulties experienced by the appraised organizations in applying measurement to their management processes. The same is so for measurement itself. While the nature of the difficulties they experience may differ, the proper enactment and institutionalization of measurement processes remains a problem for higher maturity organizations. A similar pattern exists for process performance. While the use of measurement to support process performance and improvement is particularly pertinent at maturity levels 4 and 5, noticeable proportions of the weaknesses and opportunities found in lower maturity organizations also address similar issues. Finally, problems with measuring product characteristics also are experienced by organizations at all levels of maturity.

3.2. The CIO survey

The CIO's differ in the in the extent to which their organization rely on measurement to guide their system development and maintenance efforts. One fourth of them can be characterized as having a high degree of reliance on measurement (scores of 8 through 10 on the10 point scale). Of the remainder, 39 percent report medium reliance (scores of 4 through 7), and 36 fall into the low reliance category (scores of 1 through 3). This variation in reliance on measurement is interesting in its own right. It also enables comparisons with respect to the main questions series about which classes of measures give them the most difficulty.

The CIO's replies differed considerably when they were asked to indicate the level of difficulty they encountered in establishing and using the series of measures. As seen in Figure 9, measures for tracking buy-in, risk, customer satisfaction, organizational readiness, leadership commitment, and process performance all posed sizeable challenges for the survey respondents.

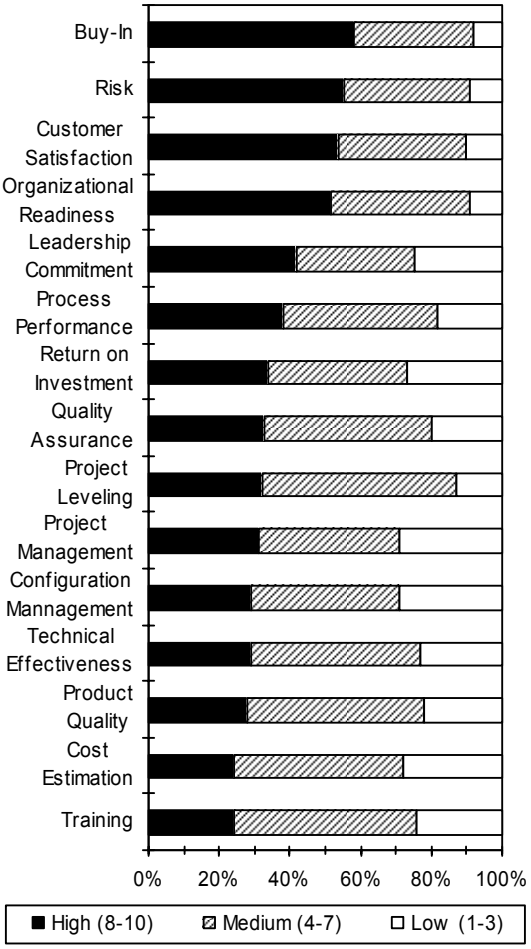

Figure 9. Difficulty establishing and using measurement

Of course, the classes of measures about which we asked the CIO's differ from the categories into which we classified the appraisal findings. In particular, many of the survey categories refer to the product or technology being developed, as well as or instead of

project or organizational processes *per se*, e.g., buy-in, organizational readiness, and leadership commitment. "Project leveling" refers to the existence of technologies and/or processes that are shared across projects.

There are both parallels and differences across the results from our two data sets. With the exception of process performance, those areas that give the CIO's most difficulty are not mirrored by the appraisal findings; however, that does not necessarily mean that the two sets of organizations differ in the difficulties that they face. It may also, or instead, be that the survey asks explicitly about topics that are not comparably covered by CMM based process appraisals.

It is possible as well that some of the reported difficulty levels from the survey data would appear more similar to the appraisal findings results in derived measures that combine the responses from similar questions comparably to the way we combined the appraisal findings. For example, one derived measure could look at tuples of the survey replies about training, cost estimation, configuration management, project management, and quality assurance. Those analyses are as yet incomplete at the time of this writing.

We also compared the results by sector. Perhaps not surprisingly, the public sector CIO's reported lower reliance on measurement than did those from the private sector ($p < .01$). Interestingly enough, the public sector CIO's also said they had greater difficulty in establishing measures for cost estimation, quality assurance, project management, product quality, and technical effectiveness than did their private sector counterparts ($p < .01$).

4. Discussion

4.1. Summary and conclusions

First of all, our analysis of the SW-CMM appraisal findings suggests several areas where better guidance about measurement and analysis might benefit appraisers and appraisal teams as well as software and systems engineering organizations as they work to improve their work processes and the products and services that they deliver. The existence of so many findings that identify inadequacies in the measurement processes of the appraised organizations is particularly noteworthy. Such a result emphasizes the importance of early attention to measurement in helping to clarify organizational goals and objectives as well as to inform management and technical decisions.

The importance of measurement and the difficulties in implementing measurement effectively are evident to measurement specialists; however, measurement is not comparably emphasized in either the SW-CMM model or its appraisal methods. The problems with the measurement of product characteristics that appear in the appraisal findings are also noteworthy for similar reasons.

The extent of findings relating to the use of measurement in support of management processes comes as no surprise given the structure and content of the Software CMM; however, the fact that findings of that category are the most common ones we see lends some credibility and face validity to our results. Finally, the relative similarities in appraisal findings across the maturity levels suggest further ways to improve the kinds of guidance that we offer to both practitioners and appraisers. That may be particularly so with respect to weaknesses in using measurement to monitor and improve process performance.

Our survey of CIO's provides complementary results. While the difficulties reported by the CIO's in using measurement in their organizations differ at first glance from the appraisal findings, they also highlight the fact that the results are dependent at least to some extent on the context in which the questions are asked. The survey did find a noticeable amount of difficulty in implementing measurement in all of the areas about which it queried, including areas similar to the appraisal findings; however, the survey also identified problem areas that typically are not emphasized in appraisals of the SW-CMM.

4.2. What's next?

Additional analyses, including breakdowns by appraised organization, model structure, and non model content as well as tests of inter-coder reliability are ongoing as of the time of this writing. We may also experiment with possible recoding of the appraisal findings according to different categories, perhaps more tightly coupled with the CMMI Measurement and Analysis process area. In addition, all PAIS findings are tagged by measurement domain, typically by KPA but also by common feature and other general issues. As noted in Section 3.2, we are also continuing to analyze the existing CIO survey data.

Future work may include further lexical analyses based on natural language processing as well as additional studies of appraiser and practitioner understanding of measurement and analysis. Of particular interest would be further analyses of appraisal findings that include CMMI appraisal results

as well as synonyms in addition to "measurement." Similarly, lexical analyses could be done on qualitative survey data gathered from both practitioners and appraisers [14]. Finally, analyses of appraisal findings of organizational strengths are also needed. While many of them appear at first glance to be boilerplate restatements of model content, the fact remains that there are slightly more findings of strengths than weaknesses of all kinds, including those related to measurement.

Our results also raise questions about the use of appraisal findings data. Would it be possible or desirable, for example, to change PAIS reporting procedures to capture fuller information about finding content and context? Any such work should begin as a research activity rather than as standard reporting procedure; and, proper expectations and incentives would have to be set with the appraiser corps. Similarly, additional research could be done on the use of appraisals findings to guide process improvement.

The fact that inadequate measurement processes and product quality are found so relatively often implies that appraisers often have a good appreciation about what can go wrong in the way appraised organizations handle, or don't handle, measurement. Still, the experience of those of us who specialize in measurement suggests that the problem is more widespread [2].

The Measurement and Analysis process area and the treatment of measurement elsewhere in CMMI models appear to be beneficial additions for organizations that are working to improve their measurement processes; however, more guidance clearly is necessary for them as well as for the appraisal corps. Such guidance can come in many forms. Interpretive documents to augment the CMMI product suite, revisions for the next model release, tutorials, and courses all are worth considering. It is our hope that this article will contribute to the dialog.

5. References

[1] D.R. Goldenson, J. Jarzombek, and T. Rout, "Measurement and Analysis in Capability Maturity Model Integration Models and Software Process Improvement," *CrossTalk: The Journal of Defense Software Engineering*, July 2003, pp. 20-24.

[2] D.R. Goldenson, A. Gopal, T. Mukhopadhyay, "Determinants of Success in Software Measurement Programs: Initial Results, *Proceedings of IEEE Metrics 1999*, Boca Raton, Florida, 1999.

[3] M.C. Paulk, B. Curtis, Bill, M.B. Chrissis, and C. Weber, *Capability Maturity Model for Software*, Version 1.1, Software Engineering Institute, CMU/SEI-93-TR-24, February 1993.

[4] M.B. Chrissis, M. Konrad, and S. Shrum, CMMI: *Guidelines for Process Integration and Product Improvement*, Addison-Wesley, Boston, 2003.

[5] D. Ahern, A. Clouse, and R. Turner, *CMMI Distilled: A Practical Introduction to Integrated Process Improvement*, Second Edition, Addison-Wesley, Boston, 2004.

[6] M. Paulk, "Analyzing the Conceptual Relationship Between ISO/IEC 15504 (Software Process Assessment) and the Capability Maturity Model for Software," in *Proceedings, Ninth International Conference on Software Quality*, Cambridge, MA, Oct. 4-6, 1999, American Society for Quality Control, Milwaukee, WI, 1999, pp. 293-303.

[7] D.K. Dunaway and S. Masters, *CMM®-Based Appraisal for Internal Process Improvement (CBA IPI)*, Version 1.2 Method Description, Software Engineering Institute, CMU/SEI-2001-TR-033, November 2001.

[8] J. McGarry, D. Card, C. Jones, B. Layman, E. Bailey,J. Dean, J., and F. Hall, *Practical Software Measurement: Objective Information for Decision Makers,* Boston, MA, Addison-Wesley, 2001.

[9] http://www.nascio.org

[10] http://www.icma.org

[11] http://www.cio.gov

[12] http://www.leadershipdirectories.com

[13] http://www.whitehouse.gov/omb/circulars/a130/a130trans4.html

[14] D. Dunaway, D. Goldenson, I. Monarch, and D. White: "How Well is CBA IPI Working? User Feedback," in *Proceedings of the 1998 Software Engineering Process Group Conference*, Chicago, IL, 1998.