A CASE STUDY IN THE EFFECTIVE USE AND QUANTITATIVE VALIDATION OF PERFORMANCE ADVISERS

Margo Thomas
The World Bank Group, Washington, DC, USA
E-mail: mthomas@ifc.org

Denise A. D. Bedford
Kent State University, USA
E-mail: Dbedfor3@kent.edu

Abstract

In 2011, a large organization, with a global mission and traditional structure set out to realign its mission in the context of the rapidly evolving global economy. This created space for a division within the institution to embark on an innovative change agenda with a renewed focus on talent management as an integral part of the change. The essential element of the talent management innovation related to performance evaluation and feedback with the goal of improving performance and results by providing richer feedback to staff. This novel approach involved leveraging the knowledge of neutral, third party Performance Advisers (PA) who gathered, reviewed and synthesized information about performance for the purpose of coaching evaluees. Based on follow-up surveys and focus groups, the pilot was broadly considered a success by most participants. A second evaluation approach was applied to assess the richness of the feedback, and to determine whether the coaching experiment produced objective, quantifiable, and verifiable changes in feedback from one performance cycle to the next. The second evaluation leveraged semantic analysis profiles and technologies. The results demonstrate observable and verifiable improvements in the feedback. This research provides important lessons for large, diverse organizations seeking to make changes in traditional performance management systems.
Key words: *performance evaluation, performance feedback, performance practices, semantic analysis methods, semantic technologies, strategic talent management.*

Introduction

It was in 1997 that consultants from McKinsey first spoke of a new war for talent (Michaels Handfield-Jones and Axelrod, 2001; Axelrod Handfield-Jones and Michaels, 2002). In the 21st century talent war large and small organizations compete to hire and retain the best human capital (Fishmnan, 1998; Trevor Gerhart & Boudreau, 1997). This perspective represents a paradigm shift from more traditional human resource approach of managing positions and salary budgets to one of identifying and leveraging management and staff knowledge as the organization's human capital (Collins and Mellahi, 2009). For the purpose of this research we adopt Collings and Mellahi's definition of strategic talent management as:

> "Activities and processes that involve the systematic identification of key positions which differentially contribute to the organization's sustainable competitive advantage, the development of a talent pool of high potential and high performing incumbents to fill these roles, and the development of a differentiated human resource architecture to facilitate filling these positions with

competent incumbents and to ensure their continued commitment to the organization." (Collings and Mellahi, 2009, p. 305)

The focus of talent management (Boudreau and Ramstad, 2005; Boudreau and Ramstad, 2007; Huselid et al, 2005) includes: (a) a shift from a vacancy led recruitment strategy toward recruiting ahead of the curve (Sparrow, 2007); (b) proactive identification of incumbents with the potential to fill key positions which may become available in the future or whose potential is such that developing it will increase the intellectual capital and the overall knowledge value of the organization; (c) systematic identification of future business needs in terms of knowledge, skills and capabilities that will be required in the future but are not currently available in house; and (d)recruiting the best people, finding positions for them and then nurturing and encouraging their performance and development (Stahl et al, 2007).

What does a strategic talent management strategy look like in practice? Lewis and Heckman (2006) propose a high-level hierarchy of the components of a strategic talent management practice. Their hierarchy is a conceptual framework that includes: (1) strategy and sustainable competitive advantage; (2) strategy and implications for talent; (3) talent pool strategy; (4) talent management systems; and (5) talent practices. Organizations that have embraced the war for talent invest in all five components of Lewis and Heckman's model. Organizations that have developed talent management strategies and align practice with those strategies understand the talent market, identify future opportunities, and see the future talent pool as a driving force for the market. They consider how to position their talent pools, and prepare competency architectures. They also critically review and adapt their traditional human resource management practices to grow their intellectual capital.

Turning strategy into practice is not a small challenge, though. It means shifting every day human resource management methods to 21st century talent practices. While there is much talk of talent management in the literature and at the strategic level, changing talent practice is more challenging. Changing talent practice means revisiting how we do performance management. Performance management is a continuous process of identifying mission and setting goals, planning performance, executing performance, assessing performance, reviewing performance and renewing performance (Aguinis Gottfreddson and Joo, 2012).

This case study describes how one organization experimented with innovative performance management methods to support its strategic talent management framework. The experimental approach involved integrating third-party coaching into the performance assessment stage. This case study provides a novel approach to coaching. Coaching has been defined as a process of helping employees recognize opportunities to improve their own performance and capabilities (Fournier, 1987; Orth et al, 1987; Popper and Lipshitz, 1992), as a way of empowering employees to move to higher levels of maturity (Burdett, 1998; Evered and Selman, 1989; Hargrove, 1995) and as a process of providing guidance, encouragement and support to the individual being coached (Redshaw, 2000; Ellinger, 1997; Ellinger and Bostrom, 1999; Mink et al, 1993; Ellinger Watkins and Bostrom, 1999). In this case, coaching by high performers was perceived as valued advice for career development and empowerment.

Case Study – The PA Initiative

The case study is drawn from a large multinational organization for which talent in and across many sectors is the comparative advantage. Drawing in the top talent from around the world is just the beginning, particularly as the organization is increasingly being forced to compete for the best. The organization strives to leverage this talent to deliver results for an increasingly discerning and demanding clientele in a rapidly evolving global environment. These changes have created space for a division within the institution to embark on an innovative

change agenda with a renewed focus on talent management as an integral part of the change. The essential element of the talent management innovation related to performance evaluation and feedback with the goal of improving performance and results by providing richer feedback to staff.

The talent management framework was comprehensive in its focus, including: (1) strategic staffing; (2) performance management; and (3) career management. This framework was designed to help recruit, retain and inspire high-quality talent to support the implementation of a new approach to delivering specialized knowledge, services and resources to clients. In 2011, the performance evaluation pilot was launched to test, in a contained and voluntary environment, a fundamentally different approach to providing timely and objective feedback. A talent management working group, comprised of staff and managers, human resource services staff, and legal department representatives, designed and implemented the pilot following a broad consultation process which involved focus groups and one-on-ones with staff and managers.

Three units participated in the Mid-year performance management pilot; a large decentralized regional unit; a small, highly specialized anchor, and a cross-cutting service unit. In total 48 staff and 19 PA advisors (including roughly 25% of staff and 5% of the PAs in the field) participated in the pilot PA

The PA was a neutral, objective third-party staff person, external to the unit, who was tasked with collecting, synthesizing and delivering in-depth and detailed performance feedback to staff and their respective supervisors/managers. It is important to note what the PA was not. The PA was not an investigator, decision-maker or an advocate for the management or staff.

Feedback is an important part of the performance management cycle – when effectively given and received it can have a significant impact on an individual's performance and attitudes. This Pilot encouraged three kinds of feedback, including feedback that addressed the six core behavioral areas (i.e., Business Judgment, Client Results, General Behavioral Concepts, Leadership, Learning and Knowledge, and Teamwork), provided concrete examples and practice advice, and managed the amount of abstraction and generalizations.

The responsibilities of the PAs were defined in detailed terms of reference. These terms of reference clearly delineated the PA's accountabilities vis a vis those of the unit manager. The PA received from the evaluees: (1) their account of accomplishments based the performance targets and deliverables agreed with managers at the beginning of the performance cycle; and (2) a list of internal and external feedback providers approved by their manager. The PA was responsible for contacting feedback providers and the manager to collect in-depth written and/or oral qualitative feedback on technical accomplishments; behaviors exemplified in achieving results; and, strengths and areas for development. The PA then provided a written synthesis of the performance feedback simultaneously to both manager and staff.

The pilot established clear guidelines on how the information that was collected would be used, provided guarantees about confidentiality, and assured a rigorous and transparent evaluation of the results and consultation on the recommendations. The PAs were required to sign a confidentiality agreement and to undertake a mandatory three-hour training session (which focused on giving and receiving feedback and understanding the PA process). Periodic meetings were organized with the PAs by a core organizing team. An anonymous feedback channel was also made available to PA's and staff participating in the pilot.

A list of potential PAs was prepared and vetted by managers and the human resources team. Those selected were invited to participate on a voluntary basis. The selected PAs were sufficiently experienced professionally and familiar with the work of the division to understand the context and basic substance of the work of the evaluees, in order to competently collect, synthesize and give feedback on performance to both the staff member and the manager.

68

Assessing the Results of the Pilot

A two phase evaluation was designed to assess the results of the PA Imitative. Phase 1 involved a staff survey and focus group interviews. Phase 2 involved a semantic analysis, which compared the PA's feedback, the pre-Initiative Traditional Performance Evaluation feedback and the post-Initiative Traditional Performance Evaluation feedback. Phase 1 results provided direct input on the process from participants. Phase 2 provided an objective and machine-based analysis of the actual changes in the feedback that resulted from the Initiative. The results of both phases are described below.

Phase 1 Evaluation - Survey Responses and Focus Interviews

Of the original cohort of roughly 50 staff and 20 PAs in the pilot, 5 staff and 3 PAs dropped out mid-way through the process for various reasons, including dissatisfaction with the pilot and work overload. The analysis of the results and outcomes of the pilot are based on a survey of staff and PA's and eight focus groups (four for PA's and four for staff), in which 34 staff and PA's participated. The response rate on the survey was 60% for staff and 68% for PA's. Some PA's and staff who did not either take/complete the survey preferred to provide write in comments or spoke to the team, and/or attended focus groups.

The feedback was largely positive on the role of the PA and the quality of feedback: 72% of the survey respondents said they were satisfied with the quality of performance feedback, and more than 66% reported that the quality of feedback was better than what they had received in previous Traditional Performance Evaluations (Traditional Performance Evaluation). In addition, PAs reported that the majority of staff interactions (93%) were positive and constructive. Most staff (89%) felt that their PA was open and forthcoming in discussions. Most PA's noted very positive experience and interactions with staff and supervisors. Regarding the role of the PA, most focus group participants valued the "objectivity" that the PA brought to the performance feedback process. The PA's reported that it took approximately one day per staff to collect, synthesize and provide feedback. Responses also suggested that staff valued competencies for grade-level benchmarking (71%) and for clarity on strengths and areas for development (85%).

In addition, a number of issues and concerns were raised through the survey and focus groups. First, the quality of the feedback on the technical accomplishments varied among PA's depending on their degree of familiarity and understanding of the technical aspects of the work. Further, a few staff noted that their PA provided feedback which was insufficiently actionable. Second, the pilot revealed that there is substantial variation across units on the extent to which feedback from external providers is solicited. This depended largely on the manager's discretion. The total number of feedback providers also varied between the Regional and Headquarters-based units. Third, pilot unit staff and PA staff noted the importance of maintaining PA quality and training on a sustainable basis, along with the importance of a clear articulation of the incentives of people who volunteer to serve as PA's. And, finally, in reflecting on this experience, several staff noted the need for a mentoring role, linking this idea to the career-management discussions between staff and managers.

The PA Initiative provides a good practice example of how to design a talent practice to support a talent management strategy. The initiative's feedback was overwhelmingly positive regarding the quality and value in terms of staff feedback, information and perspectives. The experience of PAs was also very positive. The feedback points to areas for improvement in terms of form design, automated processing of inputs and time requirements.

Margo THOMAS, Denise A. D. BEDFORD. A Case Study in the Effective Use and Quantitative Validation of Performance Advisers

PROBLEMS
OF MANAGEMENT
IN THE 21ˢᵗ CENTURY
Vol. 9, No. 1, 2014

69

Phase 2 Evaluation - Quantitative Analysis and Verification of Imitative Results

The second phase of the evaluation was a quantitative analysis of the semantics of the feedback pre- and post-coaching. The evaluation was conducted by an external team of faculty and students from Kent State University. The quantitative analysis was grounded on semantic modeling and processing of the pre- and post-coaching performance appraisal feedback.

The analysis was designed to answer four questions in a quantitatively verifiable way, including:

1. Did the coaching initiative contribute to a richer performance dialogue?

2. Did the coaching initiative increase the awareness and discussion of all six of the core behavioral competencies? If so, how did it speak to them?

3. Did the coaching initiative increase the practical and concrete language of feedback?

4. Did the coaching initiative lead to a more limited use of generalizations and abstractions in the feedback?

Question 1: Richness of the performance appraisal dialogue

The first indication of improvement expected was a simple increase in the richness of the dialogue between coach and staff, and then between manager and staff. The quantitative analysis was based on a single word count in the coaching feedback, the pre- and post-coaching Traditional Performance Evaluations.

Question 2 Core Behavioral Competency Indicators

We identified 15 indicators that were pertinent to the quality and relevance of feedback (Table 1). Six of the indicators pertained to the treatment and coverage of Core Behavioral Competencies. In the performance management context, Core Behavioral Competencies comprise the benchmarks against which performance is assessed and reviewed. Coaching feedback which addressed the core behavioral competencies was considered more valuable than feedback which did not. The organization's six Core Behavioral Competencies included: (1) business judgment; (2) client results; (3) general behavioral feedback; (4) leadership; (5) learning and knowledge sharing; and (6) teamwork.

70

Table 1. Indicators of Core Behavioral Competencies.

Indicator #	Indicator	Description
1	Business Judgment	Analyzes facts and data to support sound, logical decisions regarding own and others' work.
2	Client Results	Takes personal responsibility and accountability for timely response to client queries, requests or needs, working to remove obstacles that may impede execution or overall success.
3	General Behavioral Feedback	Communicates effectively, shows confidence in approaching tasks, engages others, gives credit, encourages others, shows initiative, sets clear goals, reaches out, shows concern for others
4	Leadership	Is able to direct and motivate team and individual staff members to deliver high quality results on time. Is able to resolve problems staying focused and providing leadership to the team through difficult times. Enables and supports growth opportunities for the team members, encourages them to stretch beyond their current experience or comfort zone. Provides ongoing feedback and mentoring.
5	Learning and Knowledge Sharing	Open to learning, shares own knowledge; looks to create knowledge products; communicate in a manner that allows them to hear others and to be heard
6	Teamwork	Collaborates with other team members and contributes productively to the team's work and output, demonstrating respect for different points of view.

Questions 3 and 4: Use of Extensional or Intentional Language

Four criteria focused on the nature of language used in the feedback process, specifically the use of intentional and extensional language. Generally speaking, linguists characterize intentional language as more abstract and conceptual, whereas extensional language is more concrete and enumerative of real examples and properties. Linguists suggest that concrete language is comparative in nature (i.e., compares one instance or example with another), references conditions, and also uses quantitative language. Linguists also suggest that language that is more abstract includes use of "allness" terms and superlatives, projects ideas or scenarios and possibilities. In good performance feedback context, we would want to see a balance between abstract (e.g. extensional) and concrete (e.g., intentional) language.

In the case of performance appraisal feedback we would expect effective individualized feedback and coaching to be stronger in the use of extensional language and weaker in the use of intentional language. More concrete and enumerative language would suggest more practical and individualized advice.

Question 3: Use of Concrete (Intentional) Language

Semantic profiles were constructed to gauge the use of concrete and comparative language (e.g., intentional language as it is characterized in linguistics) (Table 2). Four "intentional" characteristics were identified:

1. Use of language which compares – uses concrete examples for evaluation
2. Use of language which demonstrates conditions or dependencies – references examples which involve conditions or dependencies
3. Use of language which quantifies in a strict sense – references language which uses explicit quantities
4. Use of language which quantifies generally – references examples which have a quantitative aspect

Table 2. Indicators of Concrete (Intentional) Language.

Indicator #	Indicator Name	Description
11	Comparative Language	Comparative terms such as higher, lower, more, less…
12	Quantitative Language	Quantifying terms or precise numerical designations such as 60, sixty, second…
13	Conditional Language	Conditional terms such as if, but, except, perhaps, unless…
14	Consciousness of Projection Language	Consciousness of projection terms such as seems, appears, in my opinion…
15	Pseudo-Quantifying Language	Pseudo-quantifying terms or terms that loosely represent the idea of amount, size such as many, much, few, lots…

Question 4: Use of Abstract (Extensional) Language

Semantic profiles were constructed to gauge the use of abstract language (e.g., extensional as it is characterized in linguistics) (Table 3).   Four extensional criteria were identified:

1. Use of "comprehensive" or "all inclusive language – language which tends to generalize
2. Use of projection language – reflects possibilities, suppositions, and so on
3. Use of identification/predication language – which identifies and references examples or instances
4. Use of superlative or extreme language – references examples which may be extreme or at the edge

Table 3. Indicators of Abstract (Extensional) Language.

Indicator #	Indicator Name	Description
7	Allness Language	Allness terms as represented by such terms as all, every, entire, whole, none….
8	Superlative Language	Superlative terms such as best, worst, most, least, only, matchless,
9	Two-Valued Language	Two-valued terms such as either-or, if-when, if- not…
10	Identification-Predication Language	"Is" of statements which represented identification of objects. "Is" of as predication terms such as "roses are beautiful"

Pilot Data Set

Data was extracted from Traditional Performance Evaluation traditional evaluations with identifying information redacted and fully sanitized to assure confidentiality prior to analysis by the team.   All metadata that could have been used to identify the feedback provider or the feedback recipient was redacted.   Each feedback file was assigned a case number and cases were numbered and cross-referenced to facilitate comparative analysis and maintain consistency.

Comparative analysis was conducted on feedback in the year of the pilot and the preceding year before the pilot. Cases from the preceding year served as the control group.

Construction of the Semantic Models

The external team designed semantic models to represent each of the fifteen evaluative criteria. The Core Behavioral Competency criteria were translated into six semantic models,

one model for each competency. Semantic profiles for core behavioral competencies were constructed from the definitions and explanatory materials developed by human resources experts for use in the performance evaluation process. The materials were manually reviewed for key concepts and constructs. The concepts were them represented as they might be used in writing feedback, with grammatical expansions and synonym expansions. Table 4 below illustrates how many markers were identified for each criterion and built into the profile. The concepts and constructs were modeled using proximity operators because we expected semantic markers might be found in the same sentence or paragraphs. The semantic profiles were balanced across competencies – suggesting consistent definitions in the performance appraisal documentation.

Table 4. Core Behavioral Competency Semantic Markers.

Feedback Criteria and Semantic Profile	# Semantic Markers
Business Judgment	242
Client Results	240
General Behavioral Feedback	162
Leadership	304
Learning and Knowledge	369
Teamwork	420

The use of concrete and practical language was translated into four semantic models – one representing comparative language, one for conditional language, one tracking the use of pseudo-quantitative language, and one representing explicit quantitative markers. Existing and proven semantic models for intentional language were used for the analysis (Bedford 2011). And, finally, the criteria for use of abstract language were translated into four semantic models - one representing "allness," one to monitor the use of superlative language, one for the use of projective language, and one for example and identification tracking. Again, existing and proven semantic models were used for the analysis (Bedford 2011).

Semantic Processing and Data Analyses

The external team translated the semantic models into encoded rules for processing with the SAS Content Categorization Technologies. The semantic models were constructed by the external project team in consultation with the division project team. The level of semantics represented for each criterion was sufficiently refined to detect differences across cases and time periods. The semantics were defined at the concept and phrase level to address issues arising from possible variations in writing styles. The external team conducted a proof-of-concept test to validate the results of the semantic processing. The use of semantic technologies allowed the project team to complete the analysis with a minimal level of effort. The external review team also provided a neutral third-party evaluation as they were not involved in the feedback training, did not participate in the performance appraisal process, and were not known to the feedback providers. Figure 1 is a snapshot of the Business Judgment model. Figure 2 is a snapshot of the Comparative Language profile.
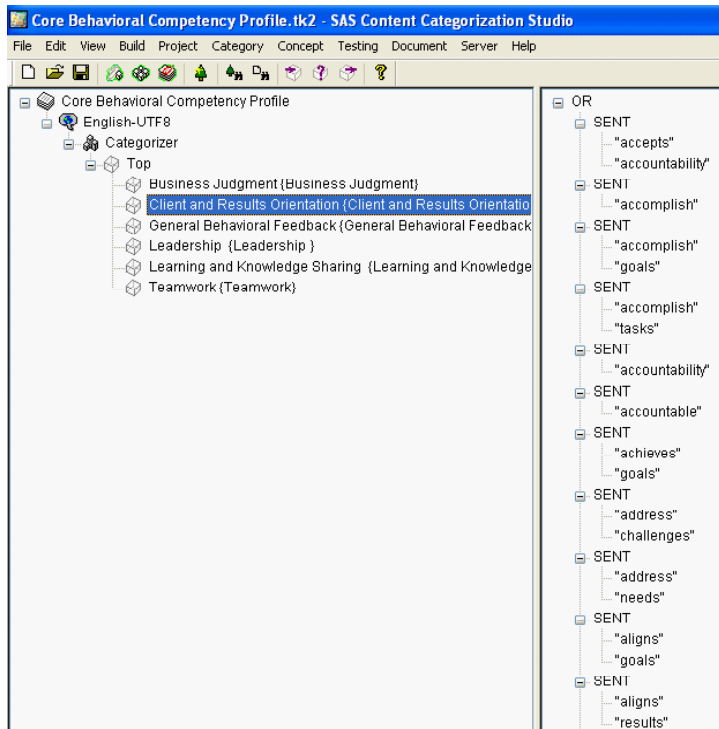
Margo THOMAS, Denise A. D. BEDFORD. A Case Study in the Effective Use and Quantitative Validation of Performance Advisers

PROBLEMS
OF MANAGEMENT
IN THE 21st CENTURY
Vol. 9, No. 1, 2014

73

Figure 1: Semantic Profile for Business Judgment Core Behavioral Competency.
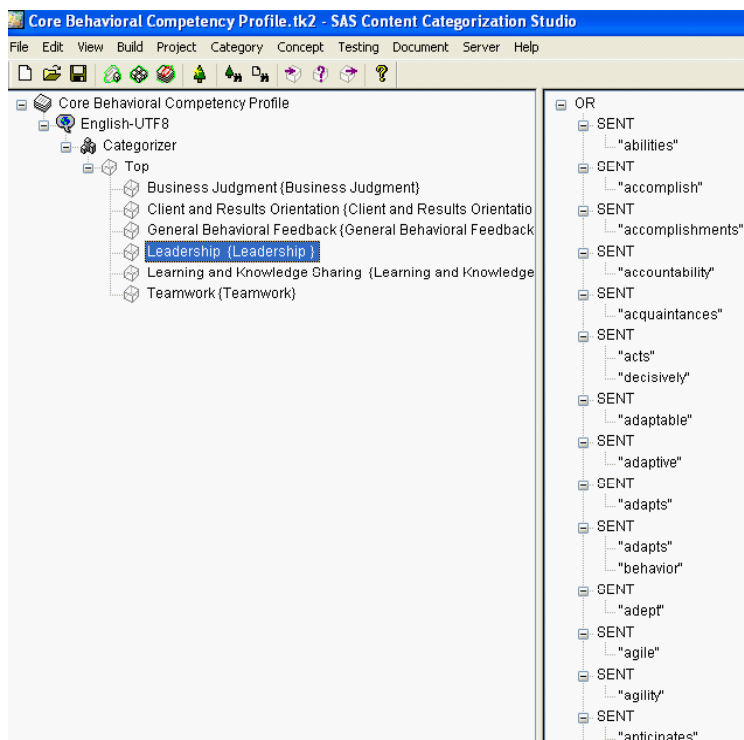

Figure 2: Semantic Profile for Use of Comparative Language.

74

Each feedback case was processed against all of the fifteen semantic profiles. In all, 42 matched cases were processed. For each case, we logged the number of semantic markers generated for each indicator as well as the individual semantic markers. The semantic processing generated a robust and in depth analysis of the feedback – case by case. For each profile, the semantic technologies generated two primary outputs: (1) an explicit list of semantic markers that were found in the feedback case; and (2) a numeric indicator of the goodness of fit of the document to each profile. In addition to providing an evaluation point for individual feedback cases, the outputs provided an aggregated view of feedback for the Year 1 Traditional Evaluation, the Year 2 Traditional Evaluation, Traditional Performance Evaluation and the pilot cases. These outputs allow us to evaluate trends across time – criteria by criteria.

The semantic processing and interpretation was consistent, objective, explicit and verifiable across all feedback cases. The use of openly developed, extensively reviewed and pre-tested profiles eliminated concerns about human subjectivity in interpretation. The encoding also made it possible for the external team to review trends and gain a deeper understanding of how the different criteria might be interpreted or used by feedback providers.

Use of Tag Clouds to Communicate Semantics

In addition to the analytical data, the project team generated tag clouds to support visual comparisons of the use of language from Year 1 to Year 2, and also covering the coaching pilot feedback. On average, the semantic technologies discovered several thousand semantic markers for each indicator. It was difficult for humans to understand the differences in richness and intensity of semantics by looking at long lists of concepts. Tag clouds were effective for visually demonstrating the differences in concepts, their richness and intensity across the three comparators. Sample tag clouds are presented for core behavioral competencies below.

Detailed Results for General Richness of the Feedback

Answers to each of the four research questions are discussed in detail below. In summary, we found that:

- Question 1 Result: The general richness of the feedback increased after the coaching pilot
- Question 2 Result: There was a major increase in the richness of discussion of core behavioral competencies
- Question 3 Result: The use of concrete language in the feedback improved after the coaching pilot
- Question 4 Result: The level of abstract and generalized feedback was held in check after the coaching pilot

Result 1: The richness of the Post-Coaching Feedback Increased

The simple amount of feedback increased from Year 1 through the Pilot Study (Table 5). The pilot study coaching, though, appears to have produced on average a four-fold increase in feedback content (400% increase).

Margo THOMAS, Denise A. D. BEDFORD. A Case Study in the Effective Use and Quantitative Validation of Performance Advisers

PROBLEMS
OF MANAGEMENT
IN THE 21st CENTURY
Vol. 9, No. 1, 2014

75

Table 5. Comparison of Semantic Richness of Three Test Sets.

Sample	Mean # Words	Mean # Paragraphs
2010 Traditional Evaluation Feedback	323.38	24.61
2012 Traditional Performance Evaluation Feedback	443.02	34.57
Year 2 Pilot Coaching Feedback	1315.15	50.4
Rate of Change Year 1 – Year 2	1.37	1.40

Result 2: The performance appraisal coaching increased the richness of discussion of core behavioral competencies

While there are variations in the results across core behavioral competencies, two things are clear from the results: (1) the coaching intervention was effective in increasing the richness and incidence of semantics; (2) the result of the coaching is evident in the Year 2 Traditional Performance Evaluation data. More specifically, we noted that the discussion of core behavioral competencies increased significantly from the Year 1 to the Year 2 Traditional Performance Evaluations. The richest feedback was found in the coaching discussions between the Performance Advisers and the staff members. By richness we mean the range and variety of core behavioral competency ideas and concepts that are used in the feedback.

Richer semantics suggests that there is a more substantive discussion of the competencies in the feedback, and that a broader range of the concepts that define the competency are being addressed. Business judgment is an example of a competency which was vaguely defined but present in the Year 1. Between Year 1 Traditional Performance Evaluation and Year 2, richness increased almost fourfold (Table 6). The semantic richness of coaching feedback was almost 15 times richer than the Year 1 Traditional Performance Evaluation. This means that an understanding of what we mean by business judgment is emerging and is being addressed across the board. The six core behavioral competencies were identified and profiled, including:

- Business Judgment (CBC1)
- Client Results Orientation (CBC2)
- General Behavioral Feedback (CBC3)
- Leadership (CBC4)
- Learning and Knowledge Sharing (CBC5)
- Teamwork (CBC6)

Table 6. Comparison of Core Behavioral Competency Treatment in Feedback.

	CBC 1	CBC 2	CBC3	CBC 4	CBC 5	CBC 6
Year 1 Traditional Evaluation Feedback Year 1Traditional Performance Evaluation	1.05	4.45	2.05	3.15	2.1	4.9
2012 Traditional Evaluation Feedback Traditional Performance Evaluation	3.64	8.02	3.85	5.57	4.78	6.80
2011 Coaching Data	15.33	19.96	14.74	17.74	15.48	22.44
Rate of Change Year 1 - Year 2	3.46	1.8	1.88	1.77	2.28	1.39

Business Judgment was a relatively new competency so we did not expect to see either a rich set of semantic markers, or a high level of incidence of use of those markers. Business Judgment emerges as a strong concept, with both richer semantics and higher incidence rates.

Client Results was a very well defined and understood competency. Richness and high

incidence were expected for this criterion. Year 1 Traditional Performance Evaluation data provides very thin treatment of this competency – with the obvious concepts having greater prominence. The Year 2 Traditional Performance Evaluation feedback increases richness and incidence – richness is improved with focus on actions towards results. The richest semantics, though, are found in the Pilot feedback – with a clear shift to the "results" aspect.

General Behavioral Feedback was as the name suggests a very general competency – the expectation was for increased incidence, though not necessarily richness because of its more general nature. This competency is thinly treated in the Year 1 Traditional Performance Evaluation data. The incidence of semantics clearly increases in Year 2, while the richness of the semantics is less evident. The Pilot feedback for this competency reflects the general nature of this competency.

Leadership was a very well defined and understood competency. Richness and high incidence were expected to increase for this criterion. The basic concept of leadership is evident and central in the Year 1 Traditional Performance Evaluation data, but there is not much evidence of rich semantics around leadership. This changes in the Year 2 Traditional Performance Evaluation data where a richer set of semantics comes through. The Pilot feedback, though, provides a very rich set of semantics around Leadership.

While there was a general focus on Learning and Knowledge Sharing across the organization, practice has evolved over time. We expected to see an increase in richness given the division's increased emphasis in the years leading up to the pilot. We observed from the semantics that the Year 1 Traditional Performance Evaluation data clearly focused on the training aspects of this competency – with thin semantics around that topic. In the Year 2 Traditional Performance Evaluation data we see a significant shift in the semantics from training to learning. In Year 2, the semantics of knowledge and knowledge sharing were clear and obvious. The richest semantics, though, are found in the Pilot data – a glance at the tag cloud for this competency demonstrates the progress made in understanding of what is meant by learning and knowledge sharing. The incidence of semantics is also greater in Year 2 and in the Pilot data.

Teamwork was a well-defined competency. We expected to see an increase in the semantic richness of discussions about this competency. Although teamwork is a core behavioral concept for the organization, we found that the focus on teamwork in the Year 1 Traditional Performance Evaluation data was relatively low. After the coaching pilot, we saw an increased discussion of the aspects of teamwork. However, even the Year 2 Traditional Performance Evaluation data is not as rich as the coaching pilot feedback.

Use of Tag Clouds to Communicate Semantics

As noted earlier, the project team generated tag clouds to support visual comparisons of the use of language in the coaching, and the pre- and post-pilot evaluations. Concepts displayed with larger fonts indicate greater incidence in the feedback. A more dense cloud suggests the use of richer language for that factor. The presentation of concepts in the clouds for comparison purposes here is challenging to read. Observations are presented for each factor. Figure 3 illustrates the semantic markers found in the Year 1 Traditional Performance Evaluation feedback. Notice the lack of key concepts such as teamwork and the low incidence rates of other team oriented concepts. Figure 4 presents the tag cloud for the Year 2 coaching feedback. Notice the dramatic increase in richness of discussion about teamwork. There are many more dimensions of the team concept visible in this cloud. Figure 5 represents the concepts that surfaced in the Year 2 feedback.

Figure 3: Year 1 Traditional Performance Evaluation Tag Cloud for Teamwork.

Figure 4: PA Feedback Tag Cloud for Teamwork.

Figure 5: Tag Cloud for Year 2 Traditional Performance Evaluation Feedback.

78

Figure 6: Year 1 Traditional Performance Evaluation Tag Cloud for Learning and Knowledge Sharing.

Figure 7: PA Feedback Tag Cloud for Learning and Knowledge Sharing.

Figure 8: Year 2 Traditional Performance Evaluation Tag Cloud for Learning and Knowledge Sharing.

Margo THOMAS, Denise A. D. BEDFORD. A Case Study in the Effective Use and Quantitative Validation of Performance Advisers

PROBLEMS
OF MANAGEMENT
IN THE 21st CENTURY
Vol. 9, No. 1, 2014

79

*Result 3: Concrete nature of the feedback language improved in
the post-coaching performance appraisals*

Four dimensions of the use of concrete language were defined, including:
• *Comparative Language Use(IT1)*
• *Conditional Language Use(IT2)*
• *Pseudo Quantifying Language Use (IT3)*
• *Explicit Quantifying Language Use (IT4)*

The use of concrete language increased in both the coaching and in the Year 2 Traditional Performance Evaluation feedback (Table 7). The use of concrete semantic markers increased both richness and incidence in three of the four indicators. This suggests that the coaching has had a positive effect. More coaching would likely contribute to more concrete semantics. Performance may improve as a result. Trust in the performance process may also improve.

Table 7. Comparison of Use of Concrete Language in Feedback.

Feedback Case #	IT 1	IT 2	IT3	IT 4
Year 1 Traditional Performance Evaluation Feedback	3.7	0.92	0.74	0.29
Year 2 Traditional Performance Evaluation Feedback	5.48	5.4	1.22	0.44
Year 2 Coaching Feedback	13.14	10.96	4.03	13.83
Rate of Change Year 1 – Year 2	1.48	5.87	1.65	1.52

In the Year 2 coaching pilot, all four intentional language indicators suggest that there is a trend towards more concrete references and actionable advice in the feedback. These trends may also suggest that the feedback is more specific to an individual's behavior and performance. There is a significant increase in the use of concrete language in both Year 2 and the Pilot data – this is important because it suggests more actionable feedback may be entering the discussions.

We expected to see greater use of comparator language in the Year 2 Traditional Performance Evaluation data and the Pilot data. An increase in the use of comparator language would suggest that coach was comparing behaviors, events and providing the recipient with information he/she could use understand past and to guide future actions. The Year 1 Traditional Performance Evaluation feedback data illustrated low level use of comparator language. The richness and incidence increased in the Year 2 Traditional Performance Evaluation feedback. We also observed that 100% of the Year 2 feedback cases included comparator language.

We expected to see greater use of conditional language in the Year 2 Traditional Performance Evaluation data and the Pilot data. The use of conditional language would indicate that the coach or manager was discussing conditions that might pertain to performance, or dependencies that may support or constrain behavior. The Year 1 Traditional Performance Evaluation data surfaced little evidence of conditional language. More conditions, more dependencies were evident in the Year 2 Traditional Performance Evaluation data – though the incidence rates are relatively low. The greatest incidence was found in the coaching data, as we expected.

We expected to see higher levels of pseudo-quantifying language – descriptors of quantities – in the Traditional Performance Evaluation and the Pilot feedback. Where coaches and managers reference quantities – whether in setting goals, creating benchmarks or referencing performance conditions, we would expect recipients to have more valuable feedback. We see the same essential semantics appearing across all three data sets – richness remains constant. Incidence, though, increases in Year 2 Traditional Performance Evaluation feedback and in the coaching feedback.

80

We expected to see higher levels in of quantitative language in the Year 2 Traditional Performance Evaluation and the coaching feedback. The rate of use of explicit quantifying language was low across all cases. The evaluation team suggested that this is not an indicator that may be as relevant to the division as was pseudo quantifying language.

Result 4. Use of abstract and generalized feedback was held in
check in the post-coaching feedback.

The use of extensional language is not necessarily an improvement in performance feedback as it represents the use of abstract and superlatives. This type of language does not provide staff with meaningful feedback on ways to grow or improve performance.

In the Year 2 coaching pilot, all four extensional language indicators were held in check, despite the overall increase in semantic richness and incidence (Table 8). This suggests that post-coaching feedback is trending towards actionable advice, concrete discussions and away from abstractions, superlatives and suppositions. The project team's general observations from the analysis suggest that the use of extensional language in the Year 1 Traditional Performance Evaluation feedback was low for all four factors. The incidence did not increase despite the overall increased richness in the coaching feedback and the Year 2 Traditional Performance Evaluation feedback.

Four dimensions of extensional language were defined including:
- Allness Language Use (ET1)
- Projection Language Use (ET2)
- Identification-Predication Language Use (ET3)
- Superlative Language Use (ET4)

Table 8. Comparison of Use of Abstract Language in Feedback.

	ET 1	ET 2	ET3	ET 4
Year 1 Traditional Performance Evaluation Feedback	1.35	0.55	1	2.65
Year 2 Traditional Performance Evaluation Feedback	2.7	0.66	1.25	4.18
Year 2 Coaching Feedback	5.96	2.51	4.0	9.92
Rate of Change Year 1 – Year 2	2	1.2	1.25	1.58

The fact that the use of extensional language are increasing at a slower rate than the use of intentional language is a good trend. It suggests that feedback providers are less prone to use all-encompassing and generalized statements than to give concrete advice.

The expectation was that allness language would be well managed and balanced with the use of concrete language. All inclusive or allness words tend to be more abstract. More abstract terms lead to generations and less actionable feedback for staff. We expected that allness semantics would increase at the same rate as concrete language. What we found, though, was relatively little change in richness or incidence in the Year 1 feedback and Year 2 feedback. This was a positive result. While new and different allness markers appeared in the coaching feedback, they were more tempered and less inclusive.

Similar to allness, superlatives tend to be more abstract and less concrete. Two factors lead us to expect a high rate of use of superlatives: (1) the organizations strong positive and respectful organizational culture, and (2) the high caliber nature of its staff. However, we expected that superlatives would be balanced with the use of concrete language, and would not grow at the same rate as concrete language. As expected, the growth between Year 1 and Year 2 Traditional Performance Evaluation data is small and the richness is not noteworthy. The use of this type of language appears to be moderating.

Margo THOMAS, Denise A. D. BEDFORD. A Case Study in the Effective Use and Quantitative Validation of Performance Advisers

PROBLEMS
OF MANAGEMENT
IN THE 21st CENTURY
Vol. 9, No. 1, 2014

81

We expected to see a variation in use of projection language, depending on the coaching or management style of the provider. Projection language speaks to future possibilities, to suppositions and assumptions, to potential. Projection language, like allness and superlatives, should be managed and balanced with use of concrete language. There was relatively little change in richness or incidence from Year 1 to Year 2 Traditional Performance Evaluation feedback. This is a good result – the trend is being managed in comparison to general semantics and the use of concrete language in particular.

In general, predication language includes verbs and verb phrases that reference examples and instances. The semantics for this criterion are by definition limited. We saw little change in richness or incidence from Year 1 to Year 2 Traditional Performance Evaluation feedback.

Overall Findings and Conclusions

We offer three general findings and/or observations. First, we found that the PA intervention was successful in improving performance evaluation feedback. This is an important finding because it tests a novel approach to developing and managing talent in a large, diverse organization with a complex mission and a traditional approach to performance management. Clearly there are important considerations and lessons that can be elicited for large organizations, recognizing that this performance feedback approach was piloted in a small, contained and voluntary context.

Second, the exploratory nature of the semantic analysis methods presents a quantitative, objective and verifiable method for discovering and evaluating changes in feedback. Semantic models and constructs are a viable approach to representing and characterizing performance evaluation feedback in an objective and consistent manner. The semantic profiles can be adapted to support other organization's competencies. Because the process is machine-based, it can be scaled to analyze hundreds or thousands of performance evaluation feedback cases. Semantic technologies which can leverage these models are efficient methods for evaluating the quality of performance evaluation feedback.

The tag clouds were important tools for communicating the semantic meaning of competencies added value in making it easier for the project team to visually understand and compare the semantics. The team welcomes feedback on other ways to meet this challenge.

Finally, the PA approach to evaluating the impact of pilot went beyond traditional quantitative analysis and qualitative participant feedback mechanisms (e.g., surveys, focus groups). The Kent State University team semantically modeling criteria and quantitatively analyzed the data. To our knowledge, this was an innovative application and context of semantic analysis. Therefore, there is more work to be done in terms of the scale effects and organizational setting for the application of semantic analysis to evaluate performance feedback quality.

From an organizational perspective, the performance pilot's outcomes and the findings about the quality of feedback offers one key conclusion. Organizational context and timing matters. The pilot was introduced in the context of a broader change management effort and was offered as a response to staff request for richer, better quality and more useful feedback. Therefore it can be argued that the voluntary participants self-selected into the pilot, possibly to improve their personal and professional outcomes from the broader change and were willing to invest in order to achieve better outcomes. In other words, these participants were more likely to find the pilot good and beneficial. This observation needs to be tested more rigorously, certainly before it applied generally in a large organization. .

82

References

Aguinis, H., Gottfredson, R. K., & Joo, H. (2012). Using performance management to win the talent war. *Business Horizons, 55 (6)*, 609-616.

Ashton, C., & Morton, L. (2005). Managing talent for competitive advantage: Taking a systemic approach to talent management. *Strategic HR Review, 4* (5), 28-31.

Baldwin, T. T., Bommer, W. H., & Rubin, R. S. (2012). *Managing organizational behavior: What great managers know and do*. McGraw Hill.

Beechler, S., & Woodward, I. C. (2009). The global "war for talent. *Journal of International Management*, *15* (3), 273-285.

Blackman, D., & Kennedy, M. (2008). *Talent Management: Developing or preventing knowledge and capability?* IRPSM, Brisbane, Australia.

Bryant, A. (2011). Google's quest to build a better boss. *New York Times*.

Cappelli, P. (2008). Talent management for the twenty-first century. *Harvard Business Review, 86* (3), 74.

Chang, T., Jianqiao, L., & Xinyan, W. (2008). The effect of team-based performance appraisal on knowledge sharing: Constructing and verifying an influencing model. *International Journal of Data Analysis Techniques and Strategies, 1* (2)*, 153-172.

Collings, D. G., & Mellahi, M. (2009). Strategic talent management: A review and research agenda. *Human Resource Management Review, 19* (4), 304-313.

Dries, N. (2013). The psychology of talent management: A review and research agenda. *Human Resource Management Review, 23* (4), 272-285.

Ellinger, A. D., Ellinger, A. E., & Keller, S. B. (2003). Supervisory coaching behavior, employee satisfaction, and warehouse employee performance: A dyadic perspective in the distribution industry. *Human Resource Development Quarterly, 14* (4), 435-458.

Farndale, E., Scullion, H., & Sparrow, P. (2010). The role of the corporate HR function in global talent management. *Journal of World Business, 45* (2), 161-168.

Fishman, C. (1998). The war for talent. *Fast Company 16,* 104.

Griffeth, R. W., & Hom, P. W. (2001). *Retaining valued employees*. Sage.

Gruman, J. A., & Saks, A. M. (2011). Performance management and employee engagement. *Human Resource Management Review, 21* (2), 123-136.

Hausknecht, J. P., Rodda, J., & Howard, M. J. (2009). Targeted employee retention: Performance based and job-related differences in reported reasons for staying. *Human Resource Management, 48* (2), 269-288.

Hiltrop, J. M. (1999). The quest for the best: Human resource practices to attract and retain talent. *European Management Journal, 17* (4), 422-430.

Kennedy, A. (2005). Models of continuing professional development: a framework for analysis. *Journal of In-service Education, 31* (2), 235-250.

Kwon, V., Bae, J., & Lawler, J. J. (2010). High commitment HR practices and top performers. *Management International Review, 50* (1), 57-80.

Lewis, R. E., & Heckman, R. J. (2006). Talent management: A critical review. *Human Resource Management Review, 16* (2), 139-154.

Lombardo, M. M., & Eichinger, R. W. (2000). High potentials as high learners. *Human Resource Management,* 39 (4), 321-329.

McCauley, C., & Wakefield, M. (2006). Talent management in the 21st century: Help your company find, develop, and keep its strongest workers. *Journal for Quality & Participation, 29* (4).

McDonnell, A. Lamare, R., Gunnigle, P., & Lavelle, J. (2010). Developing tomorrow's leaders−Evidence of global talent management in multinational enterprises. *Journal of World Business, 45* (2), 150-160.

Mellahi, K., & Collings, D. G. (2010). The barriers to effective global talent management: The example of corporate élites in MNEs. *Journal of World Business, 45* (2), 143-149.

Rhodes, C., & Beneicke, S. (2002). Coaching, mentoring and peer-networking: Challenges for the management of teacher professional development in schools. *Journal of In-Service Education, 28 (2)*, 297-310.

Sahai, S., & Srivastava, A. K. (2012). Goal/target setting and performance assessment as tool for talent management. *Procedia-Social and Behavioral Sciences, 37*, 241-246.

Schroevers, M., & Hendriks, P. (2012). Talent management in knowledge-intensive organizations. *New Research on Knowledge Management Models and Methods*, 355-370.

Schuler, R. S., Jackson, S. E., & Tarique, I. (2011). Global talent management and global talent challenges: Strategic opportunities for IHRM. *Journal of World Business, 46* (4), 506-516.

Tarique, I., & Schuler, R. S. (2010). Global talent management: Literature review, integrative framework, and suggestions for further research. *Journal of World Business,* 45 (2), 122-133.

Vural, Y., Vardarlier, P., & Aykir, A. (2012). The Effects of using talent management with performance evaluation system over employee commitment. *Procedia-Social and Behavioral Sciences, 58*, 340-349.

Weinberger, S. G. (20015). Developing a mentoring program. *Handbook of youth mentoring,* 220-234.

Whelan, E., & Carcary, M. (2011). Integrating talent and knowledge management: Where are the benefits? *Journal of Knowledge Management, 15* (4), 675-687.

Margo Thomas	Lead Operations Officer, Finance and Private Sector Development, International Finance Corporation- World Bank, I-Building, Washington DC 20433, USA. E-mail: mthomas@ifc.org
Denise A. D. Bedford	PhD., Professor, Kent State University, Kent, Ohio 44242, USA E-mail: Dbedfor3@kent.edu Website: http://www.linkedin.com/pub/denise-bedford/9/bb/abb