



THE LINE-OF-SIGHTSM ASSESSMENT

VALIDITY REPORT

Report on the Validity of the Line-of-SightSM
Assessment

Prepared by
Robert B. Bullock, Ph.D.

Updated January, 2016

TABLE OF CONTENTS

1. INTRODUCTION	2
I. PURPOSE AND ORGANIZATION OF REPORT	2
II. LINE-OF-SIGHT SM ASSESSMENT	3
A. THE FIVE KEYS TO STRATEGY EXECUTION AND MARKET DISCIPLINE	3
B. CBM RATING SCALE	4
III. LINE-OF-SIGHT SM FROM A MEASUREMENT PERSPECTIVE	5
A. FORMATIVE STRUCTURE	5
B. IMPLICATIONS OF STRUCTURE ON THE APPROACH TO ASSESSING VALIDITY	6
2. METHOD	7
I. SAMPLE POPULATIONS	7
A. MECHANICAL TURK SAMPLE	7
B. REAL-WORLD SAMPLE	8
II. ANALYSES	9
A. STATISTICAL ANALYSES USED IN THIS REPORT	9
B. STATISTICAL ANALYSES <u>NOT</u> USED IN THIS REPORT (AND WHY)	10
3. RESULTS	11
I. RELIABILITY	11
A. COMPOSITE RELIABILITY	11
B. INDICATOR MULTICOLLINEARITY	11
II. CONSTRUCT VALIDITY	13
A. ITEM-LEVEL VALIDITY	13
B. ITEM-LEVEL MULTICOLLINEARITY	14
C. CONSTRUCT VALIDITY OF SECOND-ORDER FACTOR STRUCTURE	16
I. RELATIONSHIPS BETWEEN FIRST-ORDER DIMENSIONS AND SECOND-ORDER CONSTRUCT	17
II. MULTICOLLINEARITY AMONG FIRST-ORDER DIMENSIONS	17
III. CONSISTENCY OF SECOND-ORDER FACTOR STRUCTURE ACROSS VERSIONS OF THE LOS AND SETTINGS	18
A. CONSISTENCY ACROSS VERSIONS	18
B. CONSISTENCY ACROSS SETTINGS	19
D. DISCRIMINANT VALIDITY	20
I. AVERAGE VARIANCE EXTRACTED	20
II. DISCRIMINANT VALIDITY USING FORNELL-LARCKER CRITERION ANALYSIS	21
E. CONVERGENT VALIDITY OF FORMATIVE MODEL	22
III. CRITERION-RELATED VALIDITY	25
IV. CONFIDENCE-BASED MARKING SCALE (CBM)	28
A. COMPARING CBM SCALE TO SEPARATE RATINGS OF ACCURACY AND CONFIDENCE	28
B. ASSESSING RESPONSE DISTRIBUTIONS OF CBM SCALE	30
4. APPENDIX	32
APPENDIX A: DESCRIPTIVE STATISTICS	32
1. MECHANICAL TURK SAMPLE	32
2. PRANA REAL-WORLD SAMPLE	33
APPENDIX B: SMARTPLS OUTPUT – HIGHER ORDER FACTOR STRUCTURE USING REPEATED INDICATOR APPROACH	34
APPENDIX C: SMARTPLS OUTPUT – REDUNDANCY ANALYSES	35
APPENDIX D: INSIGHTS FROM KSE IMPORTANCE ACROSS DIFFERENT VERSIONS OF THE LOS	37
5. REFERENCES	38

1. INTRODUCTION

I. PURPOSE AND ORGANIZATION OF REPORT

To evaluate the psychometric properties of the *Line-of-SightSM* (LOS) assessments, we conducted an assessment of its validity and psychometric properties using data from two samples*:

1. A sample of employees across the United States collected through Amazon's Mechanical Turk (**N = 163**)
2. Prana Business's internal database of raw LOS responses from their own clients and the clients of their Certified Practitioners (**N = 175**)

**Sample characteristics are provided in Section 2, Part I.*

The results are organized into the category of reliability or validity they represent. These include the following:

A. Reliability

- ✓ Composite reliability
- ✓ Indicator multicollinearity

B. Construct Validity

- ✓ Item-level validity
- ✓ Item-level multicollinearity
- ✓ Second-order factor structure
- ✓ Discriminant validity
- ✓ Convergent validity

C. Criterion-Related Validity

- ✓ Concurrent validity (Overall LOS)
- ✓ Concurrent validity (Keys to Strategy Execution)

D. Confidence-Based Marking Scale

- ✓ Comparing CBM to traditional approaches
- ✓ Response distributions using the CBM scale

All analyses described in this report were conducted using the software program **SmartPLS**, with the exception of multicollinearity, which was assessed using **SPSS**.

Before getting into the results, it will be important to establish a key feature of the LOS. The dimensions of the LOS – called the “Keys to Strategy Execution” (KSE) – are **formative** constructs. From a measurement perspective, this means that causality flows from item to factor. In other words, the full meaning of each KSE is derived from its indicators (more on this in the following section). The formative nature of the LOS dictated the approach taken to evaluate it.

II. LINE-OF-SIGHTSM ASSESSMENT

Line-of-SightSM (LOS) is a suite of three assessments that measure either (1) individual, (2) team, and (3) organizational performance on those critical few tasks that are essential for achieving strategic goals. These tasks are organized into the five *Keys to Strategy Execution* (KSE) plus another *Market Discipline* (MD) section. It is also important to note that the survey items and dimensions remain consistent across all three versions – the major difference is that their results are used to guide interventions at their respective levels (e.g., coaching and leader development; teambuilding and team development; organization development). The sections of the LOS and the method used to assess them are described in detail below.

A. THE FIVE KEYS TO STRATEGY EXECUTION AND MARKET DISCIPLINE

Using the Pareto Principle (i.e., the notion that 80% of a business's outcomes result from 20% of its processes), the developers of the LOS identified five dimensions that are both (a) within one's ability to control and (b) have a critical impact on the execution of an organization's strategy. These dimensions are the five KSEs, and are described below:

Strategic Understanding	This measures confidence and understanding of strategy across all levels. Strategy cannot be executed until it is known—only then can we link our efforts in a meaningful way with defined, long-term strategic objectives.
Leadership	This measures the factors of leadership that contribute to executing strategy, such as leadership credibility and constancy of purpose.
Activities & Structure	This measures whether employees operate within a structure that supports—rather than impedes—their efforts and links day-to-day activities, decisions, and actions to the big picture.
Human Capital	This measures the factors that contribute to the ability to attract, retain, and develop the talent required to execute strategy, cut through the clutter, and truly differentiate.
Balanced Metrics	This measures the extent to which metrics allow you to hone in on those vital few activities that drive the majority of results. Balanced metrics are quantitative and qualitative, lagging and leading.

At the foundation of all five KSEs is the bedrock concept of Market Discipline (MD), or the organization's value proposition that serves to organize and direct efforts towards meeting customer needs. There are three basic approaches to competitive advantage—although every company must demonstrate some basic competency at all three to produce sustainable performance, outstanding organizations typically have a clear and well-understood commitment to just one of them. The three approaches include (1) Operational effectiveness, (2) Customer intimacy, and (3) Product/service innovation. In the MD section respondents are asked to select the approach that most accurately describes their organization's strategy, processes, structure, and culture. Because the MD section is categorical (i.e., respondents are not asked to rate items but select which of three categories represents the best "fit"), it is not addressed in this report. Instead, this report focuses on the five KSEs, which are scored using an approach that is described below.

B. CBM RATING SCALE

Overall *Line-of-Sight*SM is not something that can be directly observed. Rather, it is the result of alignment in each of the KSEs and consistent organization around a single MD. To measure alignment, the LOS employs a unique approach called Confidence-Based Marking (CBM) to rate the items in each KSE. The CBM approach measures both accuracy and confidence through a single four-point scale*, which is described in the table below.

Response Option	Confidence		Accuracy	Result
1. I am confident this statement is accurate	High	+	High	= <i>Aligned</i>
2. I'm somewhat confident this is accurate; I wish I were more confident	Low	+	High	= <i>Somewhat Aligned</i>
3. I'm confident this statement is NOT accurate	High	+	Low	= <i>Misaligned</i>
4. I'm really not sure	Low	+	Low	= <i>Not Sure</i>

* Results from an assessment of the CBM scale can be found in Part IV of the Results Section.

The remainder of this page is intentionally left blank

III. LINE-OF-SIGHT FROM A MEASUREMENT PERSPECTIVE

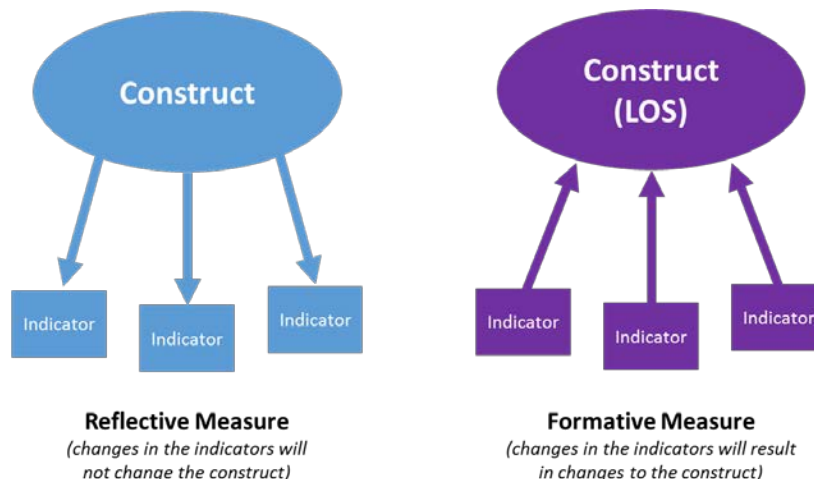
From a measurement perspective, the LOS is a *formative* assessment that is composed of five dimensions (the KSEs), each of which are measured using between three to seven items. The dimensions also load onto a single higher-order factor, called “overall line of sight.”

A. FORMATIVE STRUCTURE

To understand the concept of the LOS (and how to assess its validity and reliability), it will be important to understand the distinction between formative and reflective constructs.

In **reflective constructs**, some underlying factor *causes* behaviors that can be measured using assessments. Personality is a classic reflective construct. For example, your innate level of extraversion would *cause* you to exhibit things like talking a lot, enjoying parties, or being energized by parties. The flow of causality in a reflective construct is from the construct to the measure (i.e., being extraverted would *cause* you to score high on those characteristics).

In **formative constructs**, factors (like the KSEs) are seen as the result of a number of indicators (or a combination of indicators). Socio-Economic Status (SES) is a classic example of a formative construct. For example, SES is seen as the result of indicators like education, income, and occupational prestige. Causality here flows from the indicators to the construct (i.e., gaining an education and getting a job influences your SES.... on the other hand, SES cannot cause someone to finish college!). Another classic example of a formative construct is overall job satisfaction, which results from happiness with pay/benefits, boss, coworkers, etc. (Jarvis, MacKenzie, & Podsakoff, 2003). Both models are visualized in the figure below.



Overall LOS along with each of the KSEs are formative constructs. More specifically, the LOS can be conceptualized as a higher-order construct consisting of five formative indicators that each represent distinct facets that contribute to overall LOS. We believe LOS to be a formative construct for several reasons:

1. Overall, the construct of strategic *line of sight* is conceptualized as the outcome of strong performance on the KSEs. This means that causality flows in the direction from each KSE to overall LOS, which is a characteristic of a second-order formative construct. Likewise, each of the KSEs represent important and distinct dimensions that contributes to overall LOS, such that dropping

one of the KSEs could change overall LOS. Because of this, LOS is best identified as a higher-order formative construct explained by the first-order KSE constructs.

2. At the dimension level (i.e., looking at each of the KSEs), causality flows from the items to the construct, such that alignment (i.e., high accuracy and confidence) on the items are precursors to performance in the KSEs. The items used to measure KSEs are distinct and non-interchangeable. In reflective models, the items themselves are seen as interchangeable because they represent a spreading-out attempt to accurately measure an underlying trait using items that are similarly worded (e.g., measuring the personality trait “orderliness” using items such as, “Like order,” “Like to tidy up,” “Like routine,” and “Like to do things by the book”). In formative models, the items represent features that contribute, combine, or form some larger construct. In reflective models you can remove items without changing the underlying nature of the construct, but in formative models when you remove an item you run the risk of changing the nature of the construct (Bollen & Lennox, 1991). In the LOS, the items in each KSE clearly represent distinct features that contribute to the same outcome. For example, the Leadership KSE has three items – one measures leadership credibility, one measures leading through change, and one measures leadership communication. Removing one of these would be akin to changing this conceptualization of Leadership.

B. IMPLICATIONS OF STRUCTURE ON THE APPROACH TO ASSESSING VALIDITY

Because of this formative conceptualization, the approach to assessing the validity of the LOS is different than the approach taken by more standard reflective assessments. For example, rather than using covariance-based approaches to analyze data (e.g., AMOS, LISREL), formative constructs are assessed via variance-based methods (e.g., SmartPLS) (Haenlein & Kaplan, 2004). Likewise, the steps taken to validate formative constructs are different. For example, measures of internal consistency such as Cronbach’s Alpha are not relevant because the items in each KSE are not supposed to be highly correlated. In fact, high multicollinearity among indicators would suggest *poor* construct validity in formative models. This is because it would suggest that some of the items in a particular dimension are redundant and therefore unnecessary (Andreev, Heart, Maoz, & Pliskin, 2009). Therefore, we followed guidelines for assessing formative constructs as described by MacKenzie, Podsakoff, and Jarvis (2005), Andreev, Heart, Maoz, and Pliskin (2009), and Cenfetelli and Bassellier (2009).

****The remainder of this page is intentionally left blank****

2. METHOD

I. SAMPLE POPULATIONS

This study used data from two different samples, (A) a crowdsourced sample of US workers collected through Mechanical Turk, and (B) a real-world sample consisting of data from Prana's internal database. Each sample is described below. Descriptive statistics for both samples are provided in **Appendix A**.

A. MECHANICAL TURK SAMPLE

For most of our analyses, we relied on data we collected from 163 full-time US employees through Amazon's Mechanical Turk (MTurk). This is because many of the analyses required to assess the strengths and weaknesses of formative constructs require additional data beyond just the LOS questionnaire itself (e.g., universal indicators of factors, concurrent measures, etc.). Participants in the MTurk sample completed the Individual LOS as well as measures* of:

- Organizational innovation
- Flexibility of organization
- Quality of products and services
- Leadership effectiveness
- Satisfaction with organization
- Job satisfaction
- Turnover intention

** Note – all measures are based on the respondent's self-perceptions of those constructs.*

The MTurk sample was diverse and represented a good spread of job types and levels, gender, and age. The demographics of this sample are described in the table below.

Demographic Information from Mechanical Turk Sample

Variable	Category	Results
Gender	Female	45.1%
	Male	54.9%
Age	Mean	29.82
	Standard Deviation	8.81
	Range	19 to 60
Job Level	Technical or Specialist	29.4%
	Entry-level or Front-line	44.2%
	Supervisor or Project manager	14.7%
	Mid-level manager	9.2%
	Senior-level Manager	2.5%
Job Type (sorted high to low)	Technical or IT	17.8%
	Customer Service or Support	17.2%
	Administration	13.5%
	Sales or Marketing	12.9%
	Service Provider	12.3%
	Production or Manufacturing	8.6%
	Finance or Accounting	6.1%
	Temporary employee or Intern	5.5%
	Operations or HR	4.9%
	Consultant	1.2%

B. REAL-WORLD SAMPLE

We also conducted several analyses using real-world data from Prana's internal survey database. Whenever Certified Practitioners conduct a LOS survey with clients, respondents are given a link to Prana's survey engine where they take the assessment. This data consists of 175 responses in total. The table below shows the number of responses per version.

Sample Size for Each LOS Version from Prana Real-World Sample

Version	Sample Size
Individual LOS	66
Team LOS	79
Enterprise LOS	30

All three versions of the LOS contain similarly worded items but are targeted towards different levels. No demographic data (or any other data besides LOS responses) are available from this sample. Rather, these results were used to assess formative factor structure and version differences.

****The remainder of this page is intentionally left blank****

II. ANALYSES

The data was examined according to recommendations and guidelines for validating **formative measurements** as ascribed by Diamantopoulos and Winklhofer (2001), Jarvis et al. (2003), MacKenzie, Podsakoff, and Jarvis (2005), Cenfetelli and Bassilier (2009), Wong (2013), and others. The validity of the LOS was assessed using a number of statistical analyses which are described below.

A. STATISTICAL ANALYSES USED IN THIS REPORT

The various statistical analyses performed in this study, the type of reliability/validity evidence they provide, and their purpose are described below.

Analysis	Evidence For...	Purpose
Composite reliability	Reliability	Internal consistency reliability is not relevant for evaluating formative constructs, where items combine to form the full meaning of the construct and are not expected to be highly interrelated (Diamantopoulos & Siguaw, 2006; MacKenzie et al., 2005). However, because reliability estimates are such a strong norm, composite reliability (CR) was calculated using a reflective version of the measurement model.
Indicator multicollinearity	Reliability of a formative measure	To assess reliability (of a formative measure); indicator multicollinearity was assessed. This determines whether there is too much overlap between the items that measure a particular KSE. Too much overlap (i.e., high multicollinearity) suggests that items may not be contributing uniquely to the overall KSE.
Relationships between KSEs and indicators of organizational effectiveness	Concurrent validity	To assess concurrent validity (an element of criterion-related validity); relationships between KSEs and business outcomes and other indicators of organizational effectiveness were used to determine whether the LOS relates to other factors that are important for organizations. Correlations between KSEs and outcomes were also used to assess the unique contribution of the Confidence-Based Marking (CBM) approach to rating (compared to an alternative approach where agreement and confidence were rated using two separate rating scales).
Item-construct weights and loadings	Item-level validity	To assess item-level validity (an element of construct validity); significance of the weights and loadings between items and constructs was assessed to determine whether the items themselves significantly predict their KSEs.
First and second-order factor structures	Factorial validity	To assess factorial validity (an element of construct validity); first and second-order factor structures were assessed to determine whether items load onto their proper KSEs, and whether KSEs combine to form a single higher-order construct - overall LOS.
Fornell-Larcker criterion analysis	Discriminant validity	To assess discriminant validity (an element of construct validity); a Fornell-Larcker Criterion Analysis was conducted to assess whether the survey items combine to form dimensions (KSEs) that are statistically distinct from each other.

Redundancy analysis	Convergent validity	To assess convergent validity (an element of construct validity); a redundancy analysis was conducted to determine whether the items and KSEs are statistically related to external measures of constructs they should be related to.
Construct portability analysis	Generalizability	To assess generalizability; a construct portability analysis was conducted where factor weights were compared across studies and samples. This helps determine whether the LOS can be applied in a broad range of settings and environments with comparable results.

B. STATISTICAL ANALYSES NOT USED IN THIS REPORT (AND WHY)

Beyond describing the assessments we conducted for this report, it is also important to discuss what is not covered in this report and why. For example, this report does not include traditional assessments of reliability (i.e., Cronbach's Alpha; item-total correlations; squared multiple correlations) which are the most commonly reported measures of the "quality" of a scale. The rationale for statistical analyses not included in this report is explained in the table below.

Statistical Analyses Not Relevant for Evaluating Formative Measures

What was NOT Examined	Why
Cronbach's Alpha	Traditional reliability estimates such as Cronbach's Alpha are not relevant for this assessment because the LOS is a formative assessment (i.e., the items that measure each KSE are <u>not</u> interchangeable and combine to form the meaning of each KSE). As a formative assessment, the items in each KSE are not expected to be redundant or highly correlated so measures of inter-item correlation like Cronbach's Alpha are not applicable (Jarvis, MacKenzie, & Podsakoff, 2003). <i>Note. Composite Reliability (CV) estimates are provided in this report.</i>
Test-retest Reliability	Test-retest reliability was not assessed because we did not currently have a large enough sample of respondents who took the same assessment at multiple time-points.
Content Validity and Face Validity	Content Validity and Face Validity were not directly addressed because they were already used to guide the creation of the LOS and because they fall outside of the scope of the current statistical/psychometric purposes of this report.

The remainder of this page is intentionally left blank

3. RESULTS

Overall, results from this study provide strong support for the validity of the Line-of-Sight (LOS) assessments. Below, results are provided for each area of reliability and validity described in the Method section. This section is organized such that each subsection begins with a “bullet-point” statement of the result, then goes into detail on the approach and outcomes for each analysis.

I. RELIABILITY

A. COMPOSITE RELIABILITY

❖ **RESULT:** All dimensions showed adequate reliability.

In order to provide an estimate of the reliability of each dimension, a reflective model was created in SmartPLS. This model included all items and dimensions. Composite reliability estimates for each dimension are provided in the table below. While each KSE showed strong reliability (CR from 0.797 to 0.849), this type of analysis is not an appropriate method for evaluating formative constructs (Diamantopoulos, & Siguaw, 2006; MacKenzie, Podsakoff, & Jarvis, 2005). As such, this is the only analysis to specify the measurement model as reflective. In subsequent analyses, the LOS is always specified as a formative construct (i.e., causality flows from item to construct).

Dimension (KSE)	Items	Composite Reliability
Strategic Understanding	7	0.811
Leadership	3	0.849
Balanced Metrics	4	0.820
Activities and Structure	4	0.811
Human Capital	3	0.797

Note. Data from the real-world sample was used for this analysis. N=175.

The remainder of this page is intentionally left blank

B. INDICATOR MULTICOLLINEARITY

❖ **RESULT:** Multicollinearity was not an issue.

Results indicated the absence of collinearity (which is a good thing). Multicollinearity can be problematic in formative models because it indicates that items may overlap to the point of redundancy, thus leading to unstable indicator weights (Cenfetelli & Bassellier, 2009). To assess multicollinearity, we calculated variance inflation factor (VIF) scores for each item. Multicollinearity may be a problem when VIF scores are either below .20 or above 3.33 (Diamantopoulos & Siguaw, 2006). Results from our analysis were positive, with VIF values ranging from 1.113 at the bottom end to 2.233 at the top end – well within the ideal range. Data from the real-world sample was used for this analysis (N=175).

Variance Inflation Factor Values

Item	VIF
ACT1	1.532
ACT2	1.518
ACT3	1.387
ACT4	1.184
HUM1	1.226
HUM2	1.306
HUM3	1.185
LEAD1	1.462
LEAD2	1.373
LEAD3	1.556
MET1	1.212
MET2	1.658
MET3	1.290
MET4	1.428
STR1	1.995
STR2	2.233
STR3	1.630
STR4	1.208
STR5	1.516
STR6	1.600
STR7	1.113

****The remainder of this page is intentionally left blank****

II. CONSTRUCT VALIDITY

A. ITEM-LEVEL VALIDITY

❖ **RESULT:** All items loaded significantly onto their respective factors.

Factorial validity of the LOS is strong at the item level, such that that **each item contributes to its respective KSE**. However, two items – STR5 and STR6 – showed non-significant weights. A **weight** is a measure of the relative importance of the item when taking into account the contributions of the other items. The non-significant weight means that STR5 and STR6 do not predict *Strategic Understanding above and beyond* the other items. However, both items show significant loadings. A **loading** is a measure of the absolute importance of each item in contributing to its dimension without considering the effect of other items. This means that STR5 and STR6 are absolutely, but not relatively, important to *Strategic Understanding*. According to validity guidelines (e.g., Wong, 2013), this result means that you would be justified in either removing them from the LOS or leaving them in – the decision rests on whether or not they have practical value. Item-factor relationships are shown in the table below. Data from the real-world sample was used for both analyses. Significance values were estimated via bootstrapping (5,000).

Item-Factor Outer Weights

Item	KSE	Weight	SE	<i>t</i>	<i>p</i> *
ACT1	→ Activities & Structure	0.385	0.062	6.192	0.000
ACT2	→ Activities & Structure	0.346	0.055	6.267	0.000
ACT3	→ Activities & Structure	0.390	0.063	6.152	0.000
ACT4	→ Activities & Structure	0.246	0.059	4.186	0.000
HUM1	→ Human Capital	0.317	0.081	3.906	0.000
HUM2	→ Human Capital	0.472	0.070	6.784	0.000
HUM3	→ Human Capital	0.528	0.059	8.997	0.000
LEAD1	→ Leadership	0.491	0.066	7.452	0.000
LEAD2	→ Leadership	0.389	0.078	5.003	0.000
LEAD3	→ Leadership	0.358	0.074	4.804	0.000
MET1	→ Balanced Metrics	0.470	0.076	6.175	0.000
MET2	→ Balanced Metrics	0.231	0.082	2.825	0.005
MET3	→ Balanced Metrics	0.457	0.063	7.222	0.000
MET4	→ Balanced Metrics	0.227	0.081	2.812	0.005
STR1	→ Strategic Understanding	0.231	0.064	3.593	0.000
STR2	→ Strategic Understanding	0.257	0.071	3.608	0.000
STR3	→ Strategic Understanding	0.348	0.057	6.148	0.000
STR4	→ Strategic Understanding	0.302	0.056	5.382	0.000
STR5	→ Strategic Understanding	0.110	0.060	1.838	0.067
STR6	→ Strategic Understanding	0.055	0.074	0.751	0.453
STR7	→ Strategic Understanding	0.259	0.049	5.322	0.000

* *p*-values calculated via bootstrapping (5,000). Non-significant weights are highlighted in red. N=175.

Item-Factor Outer Loadings

Item	KSE	Loading	SE	t	p*
ACT1	→ Activities & Structure	0.778	0.043	18.194	0.000
ACT2	→ Activities & Structure	0.765	0.044	17.237	0.000
ACT3	→ Activities & Structure	0.771	0.041	18.887	0.000
ACT4	→ Activities & Structure	0.548	0.077	7.087	0.000
HUM1	→ Human Capital	0.656	0.075	8.702	0.000
HUM2	→ Human Capital	0.794	0.049	16.147	0.000
HUM3	→ Human Capital	0.791	0.046	17.014	0.000
LEAD1	→ Leadership	0.844	0.037	22.532	0.000
LEAD2	→ Leadership	0.769	0.057	13.557	0.000
LEAD3	→ Leadership	0.803	0.043	18.615	0.000
MET1	→ Balanced Metrics	0.732	0.058	12.529	0.000
MET2	→ Balanced Metrics	0.739	0.062	11.895	0.000
MET3	→ Balanced Metrics	0.738	0.054	13.733	0.000
MET4	→ Balanced Metrics	0.656	0.068	9.607	0.000
STR1	→ Strategic Understanding	0.708	0.058	12.199	0.000
STR2	→ Strategic Understanding	0.745	0.041	18.055	0.000
STR3	→ Strategic Understanding	0.796	0.037	21.323	0.000
STR4	→ Strategic Understanding	0.601	0.064	9.419	0.000
STR5	→ Strategic Understanding	0.449	0.115	3.924	0.000
STR6	→ Strategic Understanding	0.455	0.125	3.639	0.000
STR7	→ Strategic Understanding	0.430	0.094	4.580	0.000

* p-values calculated via bootstrapping (5,000). N=175.

B. ITEM-LEVEL MULTICOLLINEARITY

- ❖ **RESULT:** Multicollinearity was completely absent when looking at data from the Individual LOS and Team LOS. Multicollinearity was present in three items when looking at data from the Enterprise LOS, however this could be an artifact of small sample size (N=30).

The previous analysis indicated that each item significantly contributes to their respective construct. However, for formative models it is also important to ensure that items are not redundant (Diamantopoulos & Winklhofer, 2001). The absence of multicollinearity in formative models is important, as it indicates that the items in a construct do not tap into one single underlying aspect (as they would in a reflective model) but instead reflect distinct and unique aspects of their overarching factor (Diamantopoulos & Siguaw, 2006). To assess whether each item adds unique variance we assessed calculated VIF scores for each item after breaking out data into Individual, Team, and Enterprise versions of the LOS (results using all data are reported in section 1B above). Data were from Prana's real-world database, and it should be noted that the three LOS versions are identical in wording with the difference being that data are presented as individual breakout reports (Individual LOS), team results (Team LOS), or aggregated and reported across all respondents in an organization (Enterprise LOS).

Multicollinearity may be a problem when VIF scores are either below .20 or above 3.33 (Diamantopoulos & Siguaw, 2006). Results from our analysis were positive, with 95% of the VIF values ranging from 1.066 to 3.203. Three items from the Enterprise LOS had VIF values above 3.33. They were STR2 (4.033), STR5 (6.171), and STR6 (5.521). Thus, at the enterprise level these three items did show multicollinearity and may be redundant in the formation of the Strategic Understanding KSE. However, given the very small sample size in the Enterprise sample (N=30), further study is being required before making firm conclusions about the contribution of these items. Results for all three analyses are presented in the table below.

Item Multicollinearity Across Versions of the LOS

Item	VIF		
	Individual	Team	Enterprise
ACT1	1.179	1.809	2.523
ACT2	1.205	1.709	2.417
ACT3	1.211	1.436	1.708
ACT4	1.091	1.217	1.701
HUM1	1.352	1.183	1.62
HUM2	1.436	1.139	2.047
HUM3	1.195	1.19	1.397
LEAD1	1.282	1.587	1.874
LEAD2	1.338	1.337	2.274
LEAD3	1.454	1.524	2.099
MET1	1.159	1.351	1.066
MET2	1.294	2.12	1.969
MET3	1.187	1.469	1.505
MET4	1.298	1.627	1.623
STR1	1.463	2.303	2.665
STR2	1.75	1.985	4.033
STR3	1.787	1.479	3.203
STR4	1.278	1.09	1.69
STR5	1.204	1.278	6.171
STR6	1.38	1.336	5.521
STR7	1.128	1.155	1.499

Sample sizes: Individual (N=66), Team (N=79), Enterprise (N=30). VIF values above 3.33 are highlighted in red.

C. CONSTRUCT VALIDITY OF SECOND-ORDER FACTOR STRUCTURE

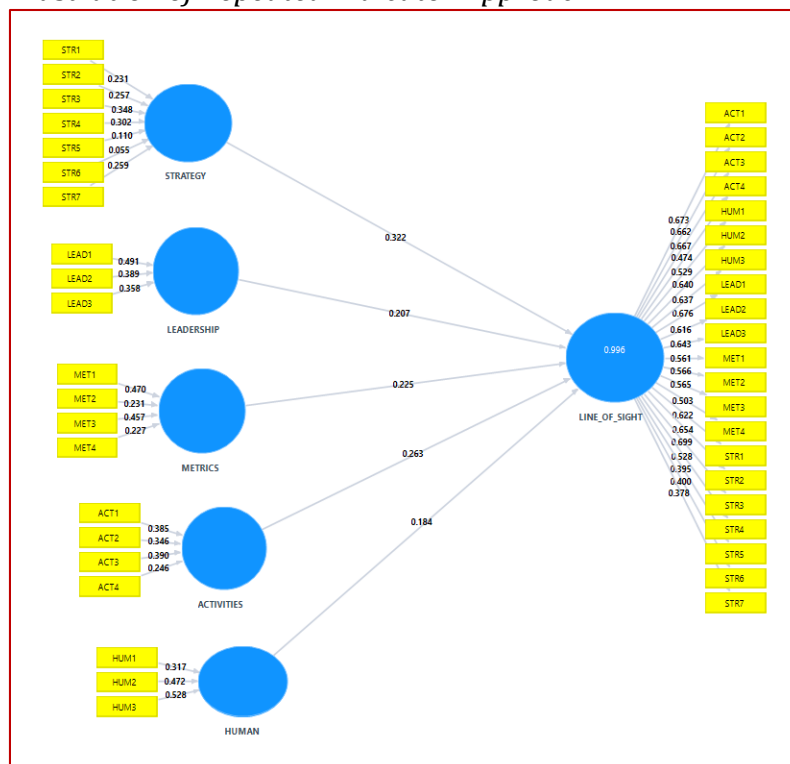
- ❖ **RESULT:** The LOS displays strong and consistent second-order factor structure, such that (i) all KSEs (i.e., first-order factors) showed significant relationships with “overall LOS,” (ii) there were no multicollinearity issues across the first-order KSEs, and (iii) the higher-order structure remained consistent across versions (i.e., Individual, Team, Enterprise) and settings (i.e., real-work sample vs. online sample).

The results from the first two analyses in this section provide validity evidence for the first-order factors (i.e., the items and KSEs) in that items loaded onto their respective KSEs, and multicollinearity is not an issue. Because the LOS posits that each KSE contributes to “overall LOS,” we assessed the factor structure of the entire LOS model, including both first-order factors (i.e., KSEs) and a single second-order factor (i.e., “overall LOS”). Specifically, this included three assessments:

- i. **Relationships between first-order dimensions (KSEs) and second-order construct (overall LOS)**
- ii. **Multicollinearity among first-order dimensions**
- iii. **Consistency of second-order factor structure across versions of the LOS and settings**

Because of the formative nature of the instrument (i.e., the items combine to form each construct with causal direction flowing from item to latent construct) path estimates were calculated using partial least squares analysis via SmartPLS. The sample consists of real-world data collected through Prana Business’s online survey engine. In developing and testing the second-order model, we used the repeated indicator approach as recommended by scholars (Becker, Klein, & Wetzels, 2012; Hair, Hult, Ringle, & Sarstedt, 2013). A screenshot showing this model is provided below and a larger image is provided in **Appendix B**.

Illustration of Repeated Indicator Approach



i. Relationships between first-order dimensions and second-order construct

In measurement models consisting of formative first-order and formative second-order factors, the first-order dimensions are not supposed to overlap but rather capture the breadth of the domain of the second order factor. Therefore, relationships between first and second-order formative constructs are assessed in terms of significance and strength (Mackenzie et al., 2005). Bootstrapping (5,000) analysis revealed significant paths between each KSE and a single overall dimension we labeled as overall LOS (i.e., overall line of sight), indicating that each KSE contributes significantly to overall LOS after considering the effects of all other KSEs. The table below shows the relationships between each KSE and overall LOS. It appears that Strategic Understanding contributed most to overall LOS (0.322), followed by Activities and Structure (0.263). Each KSE significantly contributes to a single higher-order LOS dimension.

Relationships Between First-Order KSEs and Second-Order LOS

KSE			Estimate*	SE	t	p
Activities & Structure	→	Overall LOS	0.263	0.012	22.446	0.000
Human Capital	→	Overall LOS	0.184	0.011	16.393	0.000
Leadership	→	Overall LOS	0.207	0.011	18.752	0.000
Balanced Metrics	→	Overall LOS	0.225	0.013	16.688	0.000
Strategic Understanding	→	Overall LOS	0.322	0.016	20.559	0.000

* Path estimates are standardized regression coefficients between each KSE and overall LOS. All paths are significant (p -value < .05), meaning that each one significantly contributes to an overall higher-order LOS factor.

ii. Multicollinearity among first-order dimensions

Multicollinearity assessment of the KSEs that shows that multicollinearity is not an issue at the dimension level. This means that the factors that make up overall LOS (the KSEs themselves) do not overlap too much. In formative constructs, each dimension should represent a distinct piece of the puzzle that combines to form an overall picture. If multicollinearity is too high ($VIF > 3.33$), this would suggest too much overlap between dimensions (imagine puzzle pieces bleeding into each other). Results are shown in the table below.

Multicollinearity Among First-Order Factors

KSE	VIF
Activities & Structure	2.553
Human Capital	2.246
Leadership	2.080
Balanced Metrics	1.731
Strategic Understanding	2.428

iii. Consistency of second-order factor structure across versions of the LOS and settings

To assess the consistency of the second-order structure, we conducted the same repeated indicator approach with different sets of data (i.e., breaking out Individual, Team, and Enterprise data, and comparing Prana's real-work data to the MTurk online sample) and compared results to assess the contributions of each KSE to overall LOS in different situations. Results are elaborated on below for differences in (a) version and (b) setting.

a. Consistency across versions

The factorial validity of the LOS is fairly consistent across versions (Individual, Team, Enterprise). Specifically, while the absolute values of estimates change slightly (which is expected because individuals have different needs from teams, and teams have different needs from organizations), their relative importance remains consistent. For example, at the enterprise level of the LOS, *Strategic Understanding* has a stronger contribution to overall LOS (0.384) than at the individual (0.305) or team (0.301) levels. Also, *Leadership* has a stronger path to LOS at the individual (0.224) level and less at the enterprise (0.173) level. This suggests that leaders contribute more to overall LOS for individual employees, while understanding of company strategy contributes more to overall LOS at the enterprise level (which makes logical sense).

Although the paths change slightly, the relative rank of importance stays fairly consistent across versions, with *Human Capital* contributing least and *Strategic Understanding* contributing most in both individual, team, and enterprise versions of the LOS in the real-world sample. This again provides evidence for the congruency of the three versions, but also offers a very interesting insight into the nature of strategic line of sight. A discussion of the practical significance of the difference in relative importance of KSEs to overall LOS across versions is provided in **Appendix E**. Finally, in every version of the LOS, two KSEs were consistently ranked as the most important for overall LOS: *Strategic Understanding* and *Activities and Structure*. Results are provided in the table below.

Validity Results for Formative Second-Order Construct Across Versions of LOS

KSE		LOS VERSION			COMBINED (N = 175)
		Individual (n = 66)	Team (n = 79)	Enterprise (n = 30)	
Activities & Structure	→ Overall LOS	0.262	0.264	0.229	0.263
Human Capital	→ Overall LOS	0.219	0.186	0.144	0.184
Leadership	→ Overall LOS	0.224	0.213	0.173	0.207
Balanced Metrics	→ Overall LOS	0.223	0.240	0.201	0.225
Strategic Understanding	→ Overall LOS	0.305	0.301	0.384	0.322

Values reflect path coefficients (i.e., standardized regression estimates) between each KSE and overall LOS. All paths are statistically significant ($p < .05$).

b. Consistency across settings

Comparing higher-order loadings on the Individual version of the LOS across the samples used in this report (real-world data from Prana's database vs. online data collected through Amazon's Mechanical Turk) provides us with an idea of the consistency of the Individual LOS in various settings. Overall, the path estimates remain fairly consistent across samples. The estimates are more similar in the Individual version from one sample to another than they are from one version to another, with the exception of *Balanced Metrics* (which varies less when comparing versions). Finally, the relative strength of path estimates are almost completely consistent, with the exception that *Human Capital* has the weakest loading in the real-world sample (2nd weakest in MTurk sample) and *Balanced Metrics* has the weakest loading in the MTurk sample (2nd weakest in the real-world sample). These results could be the result of inherent differences in the sample populations. For example, the high variation across path estimates for the MTurk sample could be the result of the sheer diversity in that population.

Regardless of these differences, the most important thing is that each path remains statistically significant across different populations. Our results provide support for this – all paths are significant in both populations. This provides evidence supporting the generalizability of the current conceptualization of the LOS across settings.

Comparison of Higher-Order Factor Loadings: Real-World Data vs. Online Sample

KSE		Prana Sample (N = 66)	MTurk Sample (N = 62)	Rank (A,B)
Activities & Structure	→ Overall LOS	0.262	0.284	2,2
Human Capital	→ Overall LOS	0.219	0.191	5,4
Leadership	→ Overall LOS	0.224	0.232	3,3
Balanced Metrics	→ Overall LOS	0.223	0.162	4,5
Strategic Understanding	→ Overall LOS	0.305	0.344	1,1

Data are from the Individual version of the LOS only. All paths are statistically significant ($p < .05$). A=Rank of importance for Prana's real-work sample. B=Rank of importance for MTurk sample.

The remainder of this page is intentionally left blank

D. DISCRIMINANT VALIDITY

- ❖ **RESULT:** The average variance extracted (AVE) for each KSE was >50% (with the exception of *Strategic Understanding* which displayed adequate AVE only after removing STR5 & STR6). The LOS displays adequate discriminant validity, such that each KSE adequately discriminates from other KSEs, in that their items account for more variance than the shared variance with other KSEs.

To assess discriminant validity we calculated the average variance extracted (AVE) by each KSE from their items. Convergent validity is supported when $AVE > .50$ or 50% (Fornell & Larcker, 1981). We then compared AVEs for each KSE to the percent of shared variance between each KSE and other KSEs. Discriminant validity for a KSE is met when AVE is greater than the percent of shared variance with any other KSE (Fornell & Larcker, 1981). Results are broken out into two sections:

- i. **Average variance extracted**
- ii. **Discriminant validity using Fornell-Larcker Criterion Analysis**

All analyses below were conducted using Prana's real-world sample (N=175).

i. Average variance extracted

Average variance extracted (AVE) is the average percent of variance captured by a dimension's indicators (survey items) in relation to measurement error. For example, in the table below you can see that the dimension of "Leadership" accounts for an average of 65% of the variance across the three items used to measure it. AVE values greater than .50 are said to have good convergent validity (Fornell & Larcker, 1981). Results showed that AVE was adequate ($> 50\%$) for all KSEs except for *Strategic Understanding*, which accounted for only 39.2% of the variance of the seven items used to measure it. Upon closer inspection we found that STR5 and STR6 contributed very little variance to *Strategic Understanding*. After removing these two items from the equation, AVE for *Strategic Understanding* jumped up to 58.0%. This suggests that STR5 and STR6 may not contribute as much to overall *Strategic Understanding* as the other items.

AVE for Each KSE

KSE	Items	AVE
Activities & Structure	4	52.1%
Human Capital	3	56.7%
Leadership	3	65.2%
Balanced Metrics	4	53.3%
Strategic Understanding (all items)	7	39.2%
Strategic Understanding (STR1,2,3,4,7)	5	58.0%

AVE=variance captured by a construct's indicators relative to measurement error, with values above .50 supporting convergent validity (Fornell & Larcker, 1981).

ii. Discriminant validity using Fornell-Larcker Criterion Analysis

Shared variance is the squared correlation between any two KSEs. These values can be described as the percent of overlap between those dimensions, such that higher values = stronger overlap between KSEs. To assess discriminant validity, we applied the Fornell-Larcker Criterion Analysis (Fornell & Larcker, 1981) to compare each AVE against the squared correlations between each KSE (shared variance). For adequate discriminant validity, the AVE of each KSE must be greater than the shared variance between that KSE and each other KSE. Results supported discriminant validity for all KSEs (note – the revised *Strategic Understanding* without STR5 and STR6 was used). Results are provided in the table below and show that the AVE values are greater than the shared variance values. This means that each of the KSEs are unique constructs that are statistically distinct from other KSEs.

AVE and Shared Variance Between KSEs

	ACT	HUM	LEAD	MET	STR
Activities & Structure	52.1%				
Human Capital	45.3%	56.7%			
Leadership	37.2%	39.7%	65.1%		
Balanced Metrics	33.2%	27.8%	21.5%	53.3%	
Strategic Understanding*	45.0%	30.6%	34.6%	26.8%	58.1%

*Bold values on the diagonal represent the AVE for that KSE. All other values represent the percent of shared variance between KSEs (e.g., Leadership and Balanced Metrics share 21.5% of their variance). ACT=Activities & Structure; HUM=Human Capital; LEAD=Leadership; MET=Balanced Metrics; STR=Strategic Understanding. Discriminant validity established when green values are greater than the other values in the same column or row (those values represent the correlations between KSEs). *Strategic Understanding calculated using STR1, STR2, STR3, STR4, & STR7.*

****The remainder of this page is intentionally left blank****

E. CONVERGENT VALIDITY OF FORMATIVE MODEL

- ❖ **RESULT:** Results from a redundancy analysis did not confirm the convergent validity of formatively measured KSEs.

From a statistical standpoint, the one dimension in which the LOS came up short was in regards to the convergent validity of the five KSEs. When we examined the relationship between each KSE and global measures of each dimension using redundancy analysis (Cenfetelli & Bassellier, 2009), all of the five KSEs show less than ideal convergent validity with global measures of each construct. Practically speaking, this means that the combined meaning of each KSE (combined in the sense that several indicators are aggregated to form each KSE) is somewhat different than the global item used to measure it. The analysis and results are described below.

To assess convergent validity (i.e., does this scale relate to things that it is supposed to relate to?) for each KSE, we conducted several redundancy analyses. In these analyses, we assessed the relationship between the KSEs and single-item global measures that summarize the essence of each KSE (path models shown in **Appendix C**). The strength of these relationships should ideally be above .80 to provide evidence of convergent validity (Chin, 1988). Data from the MTurk study were used because global items were included. All data used the Individual LOS. Global items used for each KSE in this analysis are provided in the table below.

Items Used in Redundancy Analysis

KSE	Global Item Used in Redundancy Analysis
Activities & Structure	All things considered, my company's structure supports the successful completion of tasks and projects.
Human Capital	All things considered, human capital is leveraged effectively at my company (e.g., talent is hired and developed, collaboration is promoted).
Leadership	All things considered, leadership (in general) is effective in my organization.
Balanced Metrics	All things considered, metrics are used effectively at my company (i.e., in a way that contributes to overall performance).
Strategic Understanding	All things considered, employees where I work have a strong understanding of the company's strategy and game plan.

The remainder of this page is intentionally left blank

Results from this analysis indicate less than ideal convergent validity for all five KSEs. All path coefficients were below the desired level of .80, although *Strategic Understanding*, *Human Capital*, and *Leadership* were close. Findings are shown in the table below and path models used in are also provided in **Appendix C**.

Path Coefficients Between Each Formative LV and Global Indicators

KSE			Path Coefficient	R ² for Global Construct*
Activities & Structure	→	Global Indicator	0.622	38.7%
Human Capital	→	Global Indicator	0.724	52.4%
Leadership	→	Global Indicator	0.700	49.0%
Balanced Metrics	→	Global Indicator	0.640	40.9%
Strategic Understanding	→	Global Indicator	0.773	59.7%

*Percent of variance in each global construct that is captured by the current items (i.e., the seven items in *Strategic Understanding* account for 59.7% of the variance in the global indicator). Global indicators are latent variables with single item indicators, loaded reflectively. Path coefficients are standardized regression estimates. Each analysis was run separately (see Appendix C for models). All paths are statistically significant ($p < .01$).

What could be causing this?

This could be due to inadequate conceptualization of the underlying nature of each construct when creating each of the single-item measures, or it could mean that the items in each KSE are slightly misaligned to their overarching constructs, or it could mean that the current item pool does not cover the entire breadth of each KSE. If the latter were true, this would suggest that adding additional items to capture the breadth of each KSE would improve these results. It is also important to note that this might be due to the coarseness of the four-point CBM scale used to score each item as well as the global indicators – the four-point scales mean less variability in responses, which in turn reduces the strength of relationships with other variables. If a five or seven-point scale were used, these path coefficients may indeed reach the .80 level.

To follow up on the findings from the redundancy analysis, we dug deeper to identify problem-items that could be changing the nature of each KSE *away from* their global conceptualizations. To assess whether individual items are conceptually related to their corresponding underlying constructs, we conducted five separate regression analyses (one for each KSE) in SPSS using the MTurk data. In these analyses, the items in a KSE were identified as predictor variables with the single-item global indicator used as a dependent variable. Results are provided in the table below. The coefficients reported in the table provide estimates of the relative contribution of each indicator, with higher estimates indicating stronger relationships with single-item global constructs. Items that did not load significantly onto their respective global indicators ($p > .10$) are highlighted in red. In other words, red items (and to a smaller extent, those items with smaller path coefficients below .200) represent the “black sheep” of each KSE in that they are not pointing in the same direction as the other items.

However, because of the formative nature of this instrument, items should not be removed on the basis of statistical analyses alone (Hair et al., 2013). Instead, these results should be used to guide future

revisions by directing questions about whether or not items contribute to the same big-picture (i.e., the meaning of their KSE) as the other items in their dimension. Of course, each single-item global indicator is not likely to fully capture underlying meaning of each KSE. Sometimes, even slight re-wording of items is all it takes to improve these results. These caveats need to be taken into consideration and suggest that these results are best served to direct future conversations about the meaning of KSEs and the adequacy of the items in each KSE. Items with weak or non-significant relationships to global indicators include: HUM1, LEAD2, STR5, STR6, and STR7 (listed at bottom of page). Results are provided in the table below.

Redundancy Analysis: Drilling Down to the Item Level

Global Indicator	Item	B Weight	SE	t	p	VIF
All things considered, my company's structure supports the successful completion of tasks and projects.	ACT1	0.031	0.078	0.404	0.687	1.674
	ACT2	0.351	0.068	5.200	0.000	1.494
	ACT3	0.169	0.066	2.561	0.011	1.574
	ACT4	0.166	0.080	2.058	0.041	1.295
All things considered, human capital is leveraged effectively at my company (e.g., talent is hired and developed, collaboration is promoted).	HUM1	0.062	0.061	1.021	0.309	1.223
	HUM2	0.315	0.069	4.601	0.000	1.385
	HUM3	0.526	0.060	8.702	0.000	1.210
All things considered, leadership (in general) is effective in my organization.	LEAD1	0.222	0.078	2.825	0.005	1.813
	LEAD2	0.284	0.081	3.501	0.001	1.889
	LEAD3	0.325	0.068	4.760	0.000	1.534
All things considered, metrics are used effectively at my company (i.e., in a way that contributes to overall performance).	MET1	0.124	0.062	1.982	0.049	1.111
	MET2	0.139	0.072	1.949	0.053	1.139
	MET3	0.187	0.061	3.068	0.003	1.159
	MET4	0.467	0.069	6.732	0.000	1.167
All things considered, employees where I work have a strong understanding of the company's strategy and game plan.	STR1	0.156	0.083	1.887	0.061	1.799
	STR2	0.339	0.084	4.032	0.000	2.156
	STR3	0.314	0.079	3.991	0.000	1.791
	STR4	0.144	0.063	2.266	0.025	1.755
	STR5	-0.096	0.068	-1.408	0.161	1.322
	STR6	-0.012	0.061	-0.202	0.841	1.296
	STR7	0.039	0.046	0.855	0.394	1.104

Data from MTurk sample used for this analysis. Regression coefficients are unstandardized B weights between items and global indicators. N=163.

Items with non-significant ($p > .10$) relationships to global single-item indicators:

- ACT1:** Managers use our company's strategy as a tool for helping employees stay focused on the critical few tasks that bring the most value to the company. ($B = 0.031, p = .687$)
- HUM1:** We receive training which enables us to perform activities that contribute to the effective execution of our company strategy. ($B = 0.062, p = .309$)
- STR5:** Understanding our company's strategy helps employees perform their jobs better. ($B = -0.096, p = .161$)
- STR6:** Members of my team feel a greater sense of purpose when they understand our business or functional unit strategy and how they contribute to its successful execution. ($B = -0.012, p = .841$)
- STR7:** Members of my team know who our main competitors are and how we differentiate from our competition. ($B = 0.039, p = .394$)

III. CRITERION-RELATED VALIDITY

A. CONCURRENT VALIDITY (OVERALL LOS)

- ❖ **RESULT:** Overall LOS displayed significant relationships with a number of positive indicators of organizational effectiveness.

When we looked at the relationships between overall LOS (conceptualized as the 2nd order factor consisting of the five KSEs with their respective formative indicators) and other organizational variables, we found strong evidence for its concurrent validity. In other words, strong alignment (and weak) across all five KSEs was associated with positive (and negative) evaluations of leadership effectiveness, quality of products and services, innovation, job satisfaction, and other outcomes. A path analysis approach was taken to assess the relationship between overall LOS and organizational variables. Overall LOS was constructed as a latent variable with five formative indicators. Each observed indicator was calculated using latent variable scores for a KSE. Estimates were calculated by drawing paths from overall LOS to each outcome. Outcomes were represented as endogenous latent variables with reflective indicators. The table below reports the coefficients of determination for each endogenous latent variable, or the proportion of variance that overall LOS explains in each outcome. Data from the MTurk sample were used because that sample includes measures of organizational variables.

There are two limitations to this analysis. First, all of the data are cross-sectional, meaning that participants completed the Individual LOS and all other measures in a single sitting. This can lead to inflated estimates, so to be conservative we present adjusted R^2 values below. Second, the outcomes were measured via self-report and represent employee perceptions of the various business outcomes. However, the scales used to measure these outcomes are straight from the research, meaning they are valid and reliable measures of what they purport to assess. The results from this analysis along with descriptions of the scales used to measure business outcomes are presented in the table and graphic below.

Relationship Between Overall LOS and Organizational Factors

Path	R^2	SE	<i>t</i>	<i>p</i>
LOS → Supportive Structure	0.499	0.058	8.556	0.000
LOS → Leadership Effectiveness	0.457	0.063	7.228	0.000
LOS → Effective Use of Human Capital	0.456	0.065	6.995	0.000
LOS → Overall Strategic Understanding	0.426	0.063	6.787	0.000
LOS → Quality	0.417	0.060	6.930	0.000
LOS → Innovation	0.374	0.061	6.148	0.000
LOS → Effective Use of Metrics	0.306	0.064	4.755	0.000
LOS → Satisfaction with Organization	0.290	0.064	4.542	0.000
LOS → Job Satisfaction	0.222	0.064	3.487	0.000
LOS → Turnover Intention*	0.112	0.051	2.213	0.027

*Note. * Path coefficient was negative (-0.335). R^2 =squared path coefficient, adjusted. SE=Standard Error. Bootstrapping (5,000) was used to assess significance. All paths were significant ($p<.05$). $N=162$.*

Descriptions of Organizational Variables and their Relationship with Overall

LOS Construct	Item(s)	R² *
Innovation	1. My company is innovative: new and better ways of doing things are readily accepted. 2. My company is flexible: we often adapt our processes as conditions change. 3. At my company, management is always looking for ways to improve products or processes. (7-point scale)	37.4%
Quality	1. My company has a reputation for providing high-quality products or services. 2. Where I work, we are always looking for ways to improve the quality of our products or services. 3. My company takes issues surrounding quality very seriously. (7-point scale)	41.7%
Job Satisfaction	Considering everything, I am satisfied with my current job. (7-point scale)	22.2%
Satisfaction with Organization	All things considered, my company is a good place to work. (7-point scale)	29.0%
Turnover Intention	I am seriously considering leaving my job (quitting) within the next 12 months. (7-point scale)	11.2%
Leadership Effectiveness	All things considered, leadership (in general) is effective in my organization. (CBM scale)	45.7%
Overall Strategic Understanding	All things considered, employees where I work have a strong understanding of the company's strategy and game plan. (CBM scale)	42.6%
Effective Use of Metrics	All things considered, metrics are used effectively at my company (i.e., in a way that contributes to overall performance). (CBM scale)	30.6%
Supportive Structure	All things considered, my company's structure supports the successful completion of tasks and projects. (CBM scale)	49.9%
Effective Use of Human Capital	All things considered, human capital is leveraged effectively at my company (e.g., talent is hired and developed, collaboration is promoted). (CBM scale)	45.6%

** Adjusted R², or the amount of variance accounted for by overall LOS, a second-order formative factor consisting of the five KSEs and their formative indicators.*

B. CONCURRENT VALIDITY (KEYS TO STRATEGY EXECUTION)

- ❖ **RESULT:** Individual KSEs showed significant and unique relationships with a number of positive indicators of organizational effectiveness, providing a nuanced view of criterion-related validity.

To further explore the relationship between the LOS and organizational factors, we examined relationships between individual KSEs and various outcomes. Using data from the MTurk sample, we calculated all estimates using a single bootstrapped (5,000) path analysis to estimate the paths between each KSE and the following organizational variables: Innovation, Quality, Job Satisfaction, Satisfaction with Organization, and Turnover Intention (items are shown in the table provided in Section 3A). KSEs were constructed as latent variables with formative indicators. Formative indicators represented the observed scores for individual survey items. All items were included in this analysis (i.e., no items were trimmed from their respective KSEs). To calculate estimates, paths were drawn from each latent KSE to each exogenous latent variables (i.e., innovation, quality, etc.).

Results indicate that the KSEs predict unique outcomes. For example, *Leadership* best predicted perceived innovation while *Activities & Structure* best predicted perceived quality of products and services. Likewise, *Strategic Understanding* was most related to employee satisfaction while *Leadership* and *Human Capital* were most related to turnover intention. Of the five KSEs, only *Balanced Metrics* showed no significant relationships with the factors listed above.

Relationships Between KSEs and Organizational Variables

Predictor	Outcome	Coefficient	SE	t	p
Activities & Structure	→ Innovation	-0.038	0.089	0.429	0.668
Activities & Structure	→ Quality	0.207	0.096	2.167	0.030*
Activities & Structure	→ Job Satisfaction	0.169	0.122	1.386	0.166
Activities & Structure	→ Org Satisfaction	0.062	0.104	0.591	0.554
Activities & Structure	→ Turnover Intention	-0.017	0.113	0.150	0.881
Human Capital	→ Innovation	0.275	0.090	3.071	0.002*
Human Capital	→ Quality	0.180	0.077	2.342	0.019*
Human Capital	→ Job Satisfaction	0.140	0.099	1.417	0.156
Human Capital	→ Org Satisfaction	0.258	0.085	3.054	0.002*
Human Capital	→ Turnover Intention	-0.150	0.115	1.301	0.193
Leadership	→ Innovation	0.376	0.084	4.502	0.000*
Leadership	→ Quality	0.193	0.097	1.985	0.047*
Leadership	→ Job Satisfaction	0.153	0.126	1.214	0.225
Leadership	→ Org Satisfaction	0.076	0.100	0.762	0.446
Leadership	→ Turnover Intention	-0.154	0.116	1.332	0.183
Balanced Metrics	→ Innovation	0.077	0.076	1.017	0.309
Balanced Metrics	→ Quality	0.051	0.072	0.709	0.478
Balanced Metrics	→ Job Satisfaction	-0.083	0.092	0.902	0.367
Balanced Metrics	→ Org Satisfaction	-0.005	0.087	0.060	0.952
Balanced Metrics	→ Turnover Intention	0.053	0.092	0.571	0.568
Strategic Understanding	→ Innovation	0.110	0.089	1.243	0.214
Strategic Understanding	→ Quality	0.159	0.079	2.003	0.045*
Strategic Understanding	→ Job Satisfaction	0.215	0.102	2.107	0.035*
Strategic Understanding	→ Org Satisfaction	0.285	0.096	2.961	0.003*
Strategic Understanding	→ Turnover Intention	-0.091	0.108	0.849	0.396

Note. * $p < .05$. Path coefficients are standardized regression coefficients. **Bold** paths are significant ($p < .05$). p -values calculated via bootstrapping (5,000). $N = 162$. Org Satisfaction = Overall satisfaction with organization.

IV. CONFIDENCE-BASED MARKING SCALE (CBM)

The LOS assessments are different from traditional surveys in that they employ a unique and non-traditional rating method. This method is called Confidence-Based Marking (CBM) and is different from standard agreement (Likert) scales in that it combines elements of both accuracy and confidence in a single four-point scale. To evaluate the CBM scale we conducted two assessments:

- A. Comparing CBM scale to separate ratings of accuracy and confidence**
- B. Assessing response distributions of CBM scale**

Results for each assessment are provided below.

A. COMPARING CBM SCALE TO SEPARATE RATINGS OF ACCURACY AND CONFIDENCE

- ❖ **RESULT:** The CBM scale is distinct from traditional scales in that it behaves like a combination of Accuracy and Confidence.

To assess the **CBM scale** (shown below) we compared the pattern of relationships between KSEs and organizational outcomes across different types of response scales. Specifically, we collected data from a separate sample of MTurkers (N=98) using the exact same LOS items and outcome measures from the previous MTurk sample with one major difference: accuracy and confidence were measured separately. In other words, rather than using the single four-point CBM scale, LOS items were assessed using two unipolar scales ranging from 1 (*Not at all Accurate/Not at all Confident*) to 7 (*Completely Accurate/Completely Confident*). This approach yielded two points of data for each LOS item – one for Accuracy and another for Confidence.

The purpose of this analysis was to determine whether the CBM scale truly results in different outcomes or whether participants were simply treating it as a simple measure of either accuracy. If the CBM scale is truly unique, it should result in a pattern of relationships with external variables that is distinct from the Accuracy and Confidence scales.

To test this, we calculated average scores for each KSE using CBM data, Accuracy data, and Confidence data. Next we calculated Pearson Correlation Coefficients (*r-values*) to assess the relationship between the three different scoring versions for each KSE and four organizational outcomes that were measured simultaneously (i.e., participants each completed the LOS and additional scales in the same sitting). Results from this analysis are provided in the table on the following page.

****The remainder of this page is intentionally left blank****

Comparing the CBM scale with Accuracy and Confidence scales using correlations between organizational outcomes and average KSE scores as calculated by each of the three scale versions

KSE	Outcome	Correlations by Response Scale		
		CBM	Accuracy	Confidence
Strategic Understanding	Satisfaction with Organization	.451**	.449**	.233*
	Job Satisfaction	.385**	.442**	.254*
	Innovation	.430**	.633**	.297**
	Quality	.498**	.663**	.448**
Leadership	Satisfaction with Organization	.451**	.633**	.185
	Job Satisfaction	.414**	.585**	.215*
	Innovation	.603**	.712**	.231*
	Quality	.568**	.678**	.419**
Activities & Structure	Satisfaction with Organization	.402**	.436**	0.159
	Job Satisfaction	.397**	.428**	.209*
	Innovation	.419**	.532**	.273**
	Quality	.544**	.589**	.376**
Human Capital	Satisfaction with Organization	.449**	.676**	0.104
	Job Satisfaction	.373**	.606**	0.069
	Innovation	.497**	.697**	0.073
	Quality	.544**	.705**	.220*
Balanced Metrics	Satisfaction with Organization	.284**	.548**	0.167
	Job Satisfaction	.211**	.530**	0.18
	Innovation	.370**	.573**	.202*
	Quality	.377**	.610**	.288**

** Correlation is significant at the .05 level (2-tailed). ** Correlation is significant at the .01 level (2-tailed). **CBM** = Original 4-point Confidence-Based Marking scale (N=161). **Accuracy** = Revised 7-point scale measuring the accuracy of each item (N=98). **Confidence** = Revised 7-point scale measuring the confidence of the previous accuracy response (N=98).*

Looking at the table it is clear that the correlation coefficients calculated using the CBM scale consistently fall in-between the coefficients calculated using the Accuracy and Confidence scales. Specifically, in all but a single KSE-Outcome pair, the CBM scale correlations were higher than the Confidence scale correlations and lower than the Accuracy scale correlations. This result is quite impressive in that it suggests that survey respondents truly seem to treat the CBM scale as a combination of both Accuracy and Confidence when they take the LOS. In short, this assessment provides support for the unique value of the CBM scale – not only does it result in significant relationships with important organizational outcomes, but it also seems to combine the results that would be found by using separate scales for Accuracy and Confidence (with the one caveat that this study used a cross-sectional design and is open to some inflation of estimates due to mono-method bias, but not to the extent that it would nullify these findings).

B. ASSESSING RESPONSE DISTRIBUTIONS OF CBM SCALE

When looking at the response distributions for each item, many showed significant positive skew and some were exponential in that they were lowest at one end and highest at the other end (exponential distributions are more common when measuring things like the frequency of personality disorders, the occurrence of earthquakes). In theory, after collecting a number of responses for a particular item or variable, they will form what is called a normal distribution. In reality, distributions are rarely if ever perfectly normal due to response skew (Furr & Bacharach, 2008). However, because many of the analyses conducted on this data rely on the assumption that response values are normally distributed, it is important to evaluate the normality of LOS items.

The results of this assessment are mixed. Using data from the MTurk sample, only 6 of 21 LOS items (not including the Market Discipline dimension) had normal distributions (i.e., z-scores for skew were within ± 3.29). The other 15 items had significantly ($p < .001$) non-normal distributions (i.e., z-scores for skew were above 3.29). The table here shows the percent of responses for each item and KSE.

It should also be noted that z-scores for skew values are sensitive to sample size. This is due to the fact that larger samples are associated with smaller standard errors, and because standard errors are the denominator in the z-score calculations (so smaller standard errors result in larger z-scores). Due to this problem, Field (2009) recommends that with large samples (around 200 or more), no significance criterion should be applied. Because the current sample is approaching 200, visual inspections of item distributions are recommended (Field, 2009).

*Frequency of response choices (largest in **bold**)*

DIMENSION	ITEM	1	2	3	4
Strategic Understanding	STR1	57.7%	35.6%	4.9%	1.8%
	STR2	53.8%	33.6%	11.2%	1.4%
	STR3	36.8%	42.3%	20.2%	0.6%
	STR4	45.7%	29.6%	20.4%	4.3%
	STR5	65.4%	24.1%	6.8%	3.7%
	STR6	54.9%	36.4%	3.7%	4.9%
	STR7	46.0%	31.1%	14.3%	8.7%
	TOTAL	51.5%	33.2%	11.6%	3.6%
Leadership	LEAD1	42.0%	40.7%	15.4%	1.9%
	LEAD2	42.3%	38.7%	18.4%	0.6%
	LEAD3	36.2%	38.0%	23.9%	1.8%
	TOTAL	40.2%	39.1%	19.3%	1.4%
Balanced Metrics	MET1	50.3%	32.5%	11.0%	6.1%
	MET2	46.3%	37.7%	13.6%	2.5%
	MET3	38.0%	37.4%	16.0%	8.6%
	MET4	41.4%	37.0%	19.1%	2.5%
	TOTAL	44.0%	36.2%	14.9%	4.9%
Activities and Structure	ACT1	37.4%	42.9%	16.6%	3.1%
	ACT2	54.0%	27.0%	16.0%	3.1%
	ACT3	44.4%	35.8%	13.6%	6.2%
	ACT4	51.2%	40.1%	8.0%	0.6%
	TOTAL	46.8%	36.5%	13.5%	3.2%
Human Capital	HUM1	44.8%	34.4%	19.6%	1.2%
	HUM2	52.1%	33.1%	13.5%	1.2%
	HUM3	36.2%	35.0%	28.8%	0.0%
	TOTAL	44.4%	34.2%	20.7%	0.8%

1 = I am confident this statement is accurate

2 = I'm somewhat confident this is accurate; I wish I were more confident

3 = I'm confident this statement is not accurate

4 = I'm really not sure

Upon visual inspection, the following characteristics were noted:

1. Positive skew is observable across items*

The most positive response, *I am confident this statement is accurate*, was typically the most frequently selected. When it wasn't, the second most positive response, *I'm somewhat confident this is accurate; I wish I were more confident*, had the highest frequency. The last option was the least frequent.

2. Distribution tails were smooth

Although positive skew was observed, the tails descended smoothly. This is good because it implies a pattern of responding that is consistent.

No bimodal distributions were observed (i.e., two separate peaks), which is good because they could suggest that external factors are influencing how people respond (e.g., population X chooses *I am confident* while population Y chooses *Not sure*).

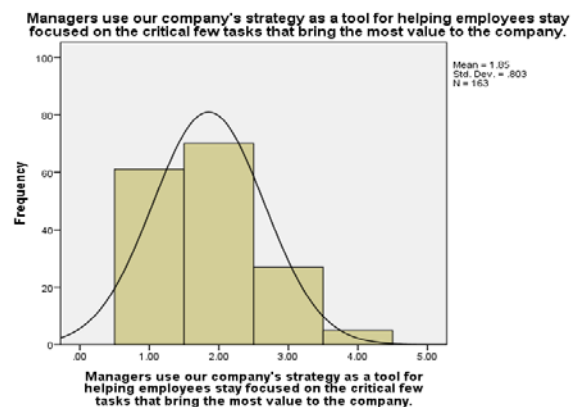
3. There were two typical response patterns

Most of the distributions for each item fell into one of two types of patterns – exponential (most frequent) or positive skew (less frequent). The two patterns can be seen in the histograms below.

Exponential Distribution



Positive Skew



Conclusion

Although positive skew was present, we decided not to transform the data for subsequent analysis due to the smooth distribution tails and adequate variance in responses. In future versions of the LOS, we recommend adding one or two more response choices to increase the granularity of the measure.

**With how the items were originally coded for this analysis (they were re-coded for other analyses so that positive regression estimates reflected positive relationships), a positive skew implies that people selected positive responses more-so than negative responses.*

4. APPENDIX

APPENDIX A: DESCRIPTIVE STATISTICS

1. MECHANICAL TURK SAMPLE

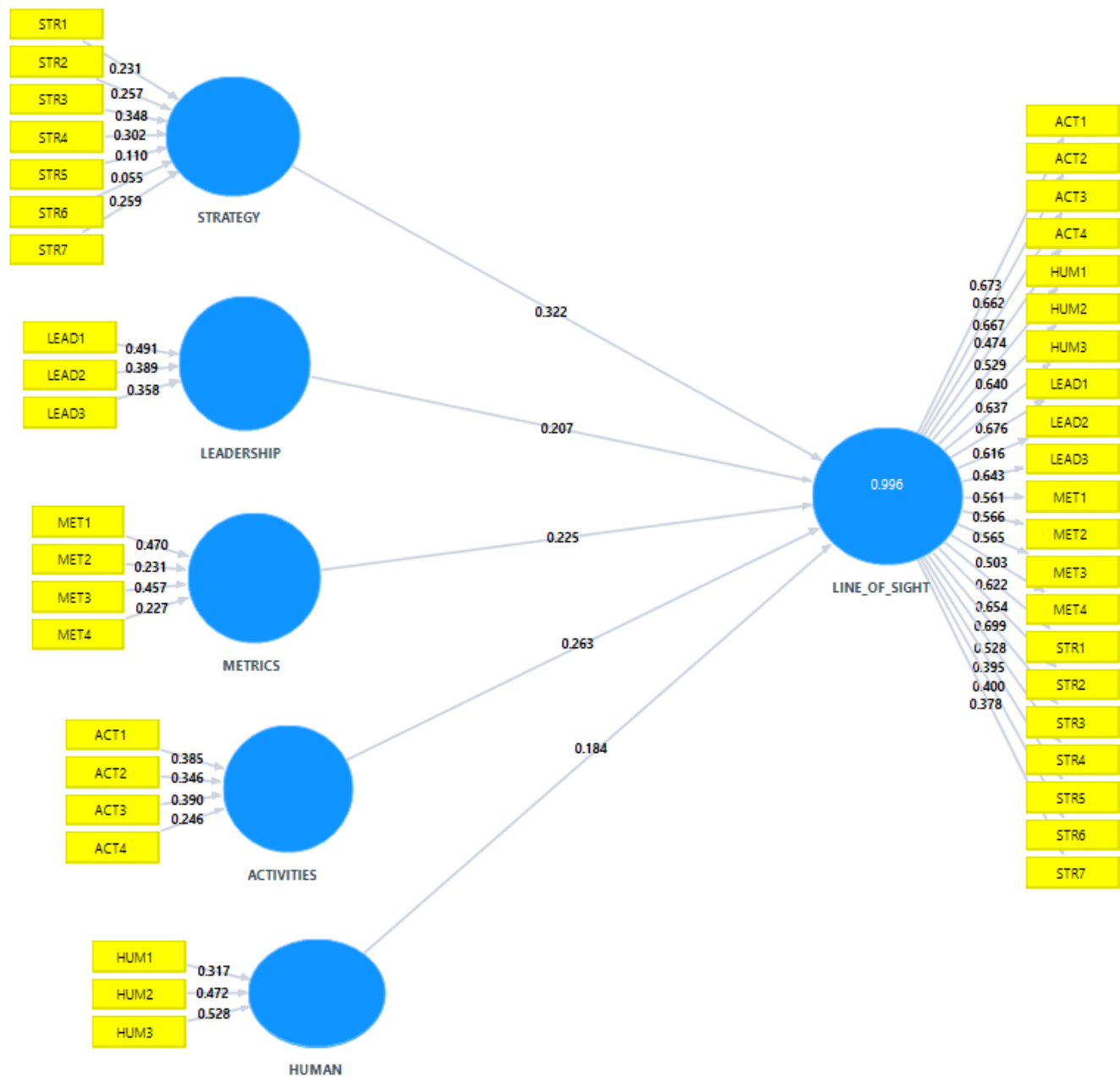
Category	Item	Mean	SE	SD	Median	Range	N
Strategic Understanding	STR1	3.491	0.053	0.679	4	3	163
	STR2	3.399	0.062	0.742	4	3	143
	STR3	3.153	0.059	0.758	3	3	163
	STR4	3.167	0.071	0.900	3	3	162
	STR5	3.512	0.061	0.782	4	3	162
	STR6	3.414	0.062	0.785	4	3	162
	STR7	3.143	0.076	0.967	3	3	161
Leadership	LEAD1	3.228	0.061	0.774	3	3	162
	LEAD2	3.227	0.060	0.764	3	3	163
	LEAD3	3.086	0.064	0.820	3	3	163
Balanced Metrics	MET1	3.270	0.070	0.889	4	3	163
	MET2	3.278	0.062	0.790	3	3	162
	MET3	3.049	0.074	0.942	3	3	163
	MET4	3.173	0.065	0.823	3	3	162
Activities & Structure	ACT1	3.147	0.063	0.803	3	3	163
	ACT2	3.319	0.067	0.851	4	3	163
	ACT3	3.185	0.070	0.893	3	3	162
	ACT4	3.420	0.052	0.666	4	3	162
Human Capital	HUM1	3.227	0.063	0.804	3	3	163
	HUM2	3.362	0.060	0.760	4	3	163
	HUM3	3.074	0.063	0.806	3	2	163
Organizational Variables	JOBSAT	4.708	0.128	1.623	5	6	161
	ORGSAT	5.056	0.114	1.442	5	6	161
	TURNOVER	3.765	0.158	2.017	4	6	162
	INNOV1	4.265	0.130	1.656	4	6	162
	INNOV2	4.519	0.128	1.627	5	6	162
	INNOV3	4.764	0.122	1.547	5	6	161
	QUAL1	5.094	0.123	1.553	5	6	160
	QUAL2	5.086	0.114	1.455	5	6	162
	QUAL3	5.342	0.119	1.509	6	6	161
Global Single-Item Indicators	STR_Ov	3.373	0.060	0.765	4	3	161
	LEAD_Ov	3.196	0.062	0.792	3	3	163
	MET_Ov	3.124	0.068	0.862	3	3	162
	ACT_Ov	3.309	0.059	0.750	3	3	162
	HUM_Ov	3.110	0.063	0.809	3	3	163

2. PRANA REAL-WORLD SAMPLE

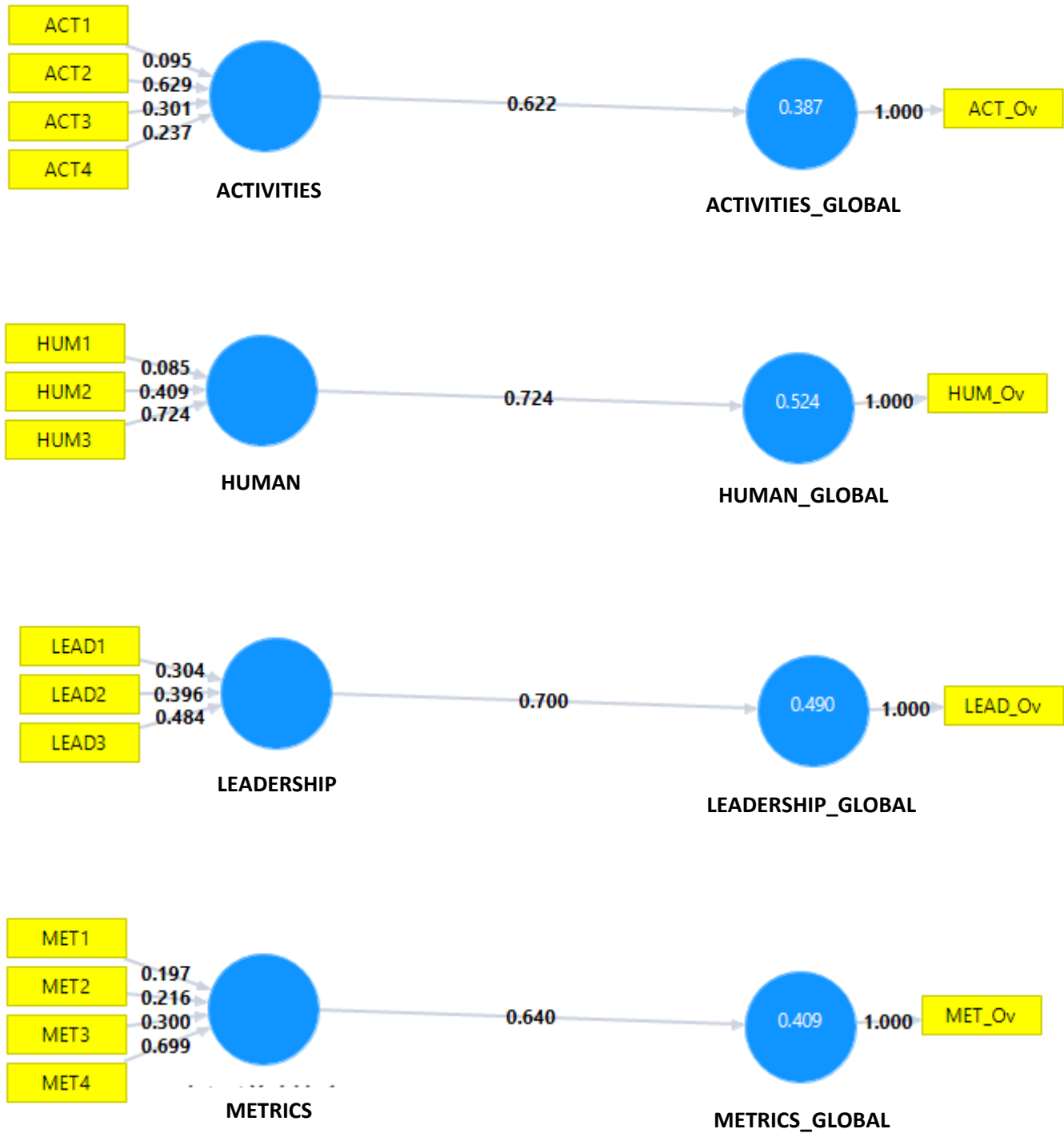
Category	Item	Mean	SE	SD	Median	Range	N
Strategic Understanding	STR1	3.380	0.061	0.813	4	3	175
	STR2	3.050	0.057	0.749	3	3	175
	STR3	2.550	0.057	0.748	3	3	175
	STR4	3.170	0.069	0.912	3	3	175
	STR5	3.670	0.053	0.698	4	3	175
	STR6	3.770	0.042	0.551	4	3	175
	STR7	3.100	0.061	0.810	3	3	175
Leadership	LEAD1	3.090	0.061	0.804	3	3	175
	LEAD2	2.910	0.063	0.829	3	3	175
	LEAD3	2.700	0.055	0.729	3	3	175
Balanced Metrics	MET1	3.170	0.074	0.977	3	3	175
	MET2	2.900	0.068	0.901	3	3	175
	MET3	2.910	0.068	0.905	3	3	175
	MET4	2.850	0.068	0.893	3	3	175
Activities & Structure	ACT1	2.710	0.063	0.836	3	3	175
	ACT2	2.830	0.060	0.791	3	3	175
	ACT3	2.990	0.070	0.928	3	3	175
	ACT4	3.250	0.061	0.806	3	3	175
Human Capital	HUM1	2.860	0.063	0.833	3	3	175
	HUM2	3.020	0.064	0.844	3	3	175
	HUM3	3.230	0.059	0.778	3	3	175

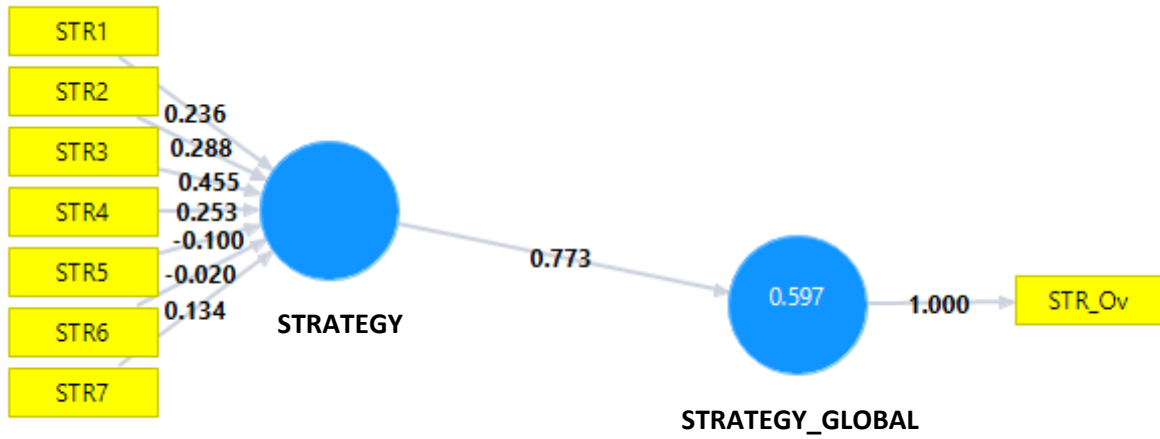
The remainder of this page is intentionally left blank

APPENDIX B: SMARTPLS OUTPUT – HIGHER ORDER FACTOR STRUCTURE USING REPEATED INDICATOR APPROACH



APPENDIX C: SMARTPLS OUTPUT – REDUNDANCY ANALYSES





The remainder of this page is intentionally left blank

APPENDIX D: INSIGHTS FROM KSE IMPORTANCE ACROSS DIFFERENT VERSIONS OF THE LOS

Slight differences in responding from version to version shows that, when used at higher levels (e.g., Team and Enterprise versions), certain KSEs start to carry more weight in accounting for overall LOS. This insight suggests that, when reviewing LOS results with clients or using them to plan out interventions, CPs may want to place more or less emphasis on different KSEs *depending on the LOS version they used* (i.e., the level of analysis). For example, *Strategic Understanding* is 25% more important in accounting for overall line of sight at the Enterprise level than it is at the Individual level. Therefore, if all else was equal, *Strategic Understanding* would be a higher priority when facilitating Enterprise LOS interventions compared to Individual LOS interventions.

Specifically, results show that...

- **Activities and Structure are most relevant for Team LOS**
- **Human Capital is most relevant for Individual LOS**
- **Leadership is most relevant for Individual LOS**
- **Balanced Metrics are most relevant for Team LOS**
- **Strategic Understanding is most relevant for Enterprise LOS**

The table below shows the contribution of each KSE towards overall LOS according to each version of the survey. Color has been applied to show the level that each KSE has the greatest impact on LOS.

KSE		LOS Version		
		Individual (n = 66)	Team (n = 79)	Enterprise (n = 30)
Activities & Structure	→ Overall LOS	0.262	0.264	0.229
Human Capital	→ Overall LOS	0.219	0.186	0.144
Leadership	→ Overall LOS	0.224	0.213	0.173
Balanced Metrics	→ Overall LOS	0.223	0.240	0.201
Strategic Understanding	→ Overall LOS	0.305	0.301	0.384

Overall, the relative contribution for each of the five KSEs is more balanced at the Individual level (i.e., each KSE has similar loadings onto overall line of sight) then differentiate more moving up to Team and Enterprise levels. This could be leveraged by focusing more developmental effort towards the more important KSEs at each level (e.g., to guide follow-up activities, provide a table that shows the relative importance of each KSE – for the Enterprise version of the LOS, this table would have *Strategic Understanding* at the top, so areas of “misalignment” in this KSE should be handled before anything else).

The remainder of this page is intentionally left blank

5. REFERENCES

- Andreev, P., Heart, T., Maoz, H., & Pliskin, N. (2009). Validating formative partial least squares (PLS) models: Methodological review and empirical illustration. *Thirtieth International Conference on Information Systems. Phoenix, Arizona, ICIS 2009 Proceedings*.
- Becker, J. M., Klein, K., & Wetzels, M. (2012). Hierarchical latent variable models in PLS-SEM: Guidelines for using reflective-formative type models. *Long Range Planning*, 45, 359-394.
<http://dx.doi:10.1016/j.lrp.2012.10.001>
- Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110, 305-314.
- Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. New York: Routledge.
- Cenfetelli, R. T., & Bassellier, G. (2009). Interpretation of formative measurement in information systems research. *MIS Quarterly*, 33, 689-707.
- Chin, W. W. (1998). The partial least squares approach for structural equation modelling. In G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 295-336). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Diamantopoulos, A., & Siguaw, J. A. (2006). Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. *British Journal of Management*, 17, 263-282.
- Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, 38, 269-277.
- Field, A. (2009). *Discovering Statistics Using SPSS* (3rd ed.). London: Sage.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18, 39-50.
- Furr, R. M., & Bacharach, V. R., (2008). *Psychometrics: An introduction*. Los Angeles: Sage.
- Gefen, D., & Straub, D. W. (2005). A practical guide to factorial validity using PLS-Graph: Tutorial and annotated example. *Communications of AIS*, 16, 1-28.
- Hair, J. F., Hult, G. T., Ringle, C. M., & Sarstedt, M. (2013). *A primer on Partial Least Squares Structural Equation Modelling (PLS-SEM)*. Thousand Oaks, California: SAGE Publications, Inc.
- Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, 30, 199-218.
- Haenlein, M., & Kaplan, A. M. (2004). A beginner's guide to partial least squares analysis. *Understanding Statistics*, 3, 283-297.
- MacKenzie, S. B., Podsakoff, P. M., & Jarvis, C. B. (2005). The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. *Journal of Applied Psychology*, 90, 710-730.
- Wong, K. (2013). Partial least squares structural equation modeling (PLS-SEM) techniques using SmartPLS. *Marketing Bulletin*, 24, 1-32.