AN INITIAL EXAMINATION OF OUTCOMES FROM SERVANT LEADERSHIP IMPLEMENTATION AT A NATIONAL BEHAVIORAL HEALTHCARE ORGANIZATION

Anthony J. Roberson M.A. & Kristen Bielecki, Ph.D.[1]
Devereux Children's Behavioral Health Services

While the notion of leaders placing the interests of their followers before their own is a concept that pre-dates the bible (e.g., Plato's *Republic*; see Williamson, 2008), the term "servant leadership" (SL) was not introduced into the modern vernacular until Greenleaf's writings in the 1970's (Greenleaf, 1977; 1998). Since then, SL has emerged in the popular and scholarly literature as a promising positive leadership framework embraced by several high profile organizations such as Starbucks, Wegmans Food Market, Cleveland Clinic, and many others. The simple logic behind SL is that modern leaders have an obligation to actively support the wellbeing of their followers as a valued end in itself. There is no doubt that such a management approach stands in stark contrast to traditional capitalistic ideas of effective leadership— where the central focus is on promoting company growth over the broader humanistic needs of employees (Biglan, Lee, & Cody, 2016). Yet, the emerging evidential and ethical basis for adopting a person-centered leadership style suggests that bolstering the personal growth and engagement of employees not only supports the flourishing of the organization, but aids in the cultivation of a more nurturing society writ large (Biglan et al., 2016). Despite this grand promise, the available scholarship is still relatively nascent and in need of more rigorous investigation (Van Dierendonck, 2011). The intention of the proposed study is to offer such an empirical evaluation of implementation and outcomes from a company-wide SL initiative at a national behavioral healthcare organization. This study will function both as an internal evaluation of the initiative's effectiveness—helping the organization's leaders use the results to shape specific company policies and practices—and as a contribution to the organizational leadership literature more broadly.

To date, the majority of writings on SL have focused on clarifying various operationalizations of the concept; however, there remains a great deal of disagreement around what SL actually is. In Van Dierendonck's review (2011), the author notes that Greenleaf himself left the exact definition of SL imprecise and may not have even intended for it to be strictly defined at all. Nevertheless, this has not stopped later theorists from proposing several competing conceptual models, each with overlapping and distinct elements. Unfortunately, this work may have collectively done more to muddy the waters than provide consensus. For instance, while some frameworks describe SL solely in terms of characteristics of servant-leaders themselves (e.g., Spears, 1995; Laub, 1999), others conceptualize SL as a larger multi-component process (e.g., Patterson, 2003; Winston, 2003). Such fundamental discrepancies between models result in SL appearing as a poorly distinguished slurry of latent character traits (e.g., a need to serve), observable behaviors (e.g., providing direction), mediating factors (e.g., mutual respect between leaders and followers), and outcomes (e.g., employee engagement). Given that comprehensive component analyses of SL features have yet to be conducted (and such an undertaking is well beyond the scope of the proposed study), Van Dierendonck's synthesized definition of SL behavior will have to suffice for now. His six-pronged definition views servant-leaders as individuals who (1) endeavor to develop others, (2) show humility, (3) accept people for who they are, (4) provide direction, (5) are authentic, and (6) act as stewards for the collective good (2011).

The lack of consensus for a precise description of SL should not suggest that management behavior aligned with the spirit of the above definition is not very much worthwhile. While relatively little attention has been paid to investigating consequent effects of SL initiatives compared to antecedent determinants (Van Dierendonck, 2011), the available quantitative evidence paints an optimistic picture for what outcomes might be expected. For example, organizational embrace of SL as a core value has been associated with various desirable outcomes such as lower employee disengagement, reduced staff turnover intention, and increased employee helping behaviors (Hunter, Neubert, Perry, Witt, Penney, & Weinberger, 2013).

---

[1] Correspondence regarding this study should be directed to Tony Roberson at tonyjroberson@gmail.com or Dr. Kristen Bielecki at kbieleck@devereux.org.

Further, empirical evidence from the wider literature suggests that increased employee wellbeing may be linked to a diverse set of positive outcomes at multiple organizational levels. For instance, increased staff engagement in hospital settings has been associated with fewer patient safety incidents (Harter, Schmidt, Killham, & Agrawal, 2013) and higher patient satisfaction ratings (Gallup, 2003).

Although the principles of behavior analysis (BA) have not been explicitly proposed as an explanatory framework for why person-centered leadership styles like SL produce such outcomes, scaling up these principles to the level of systems behavior may provide useful insights (Biglan & Glenn, 2013). The BA approach appeals to the environmental contingencies that can be directly modified in order to shape behavior to the desired state. In this view, servant-leaders verbally specify wellbeing-promoting[2], prosocial[3] rules to govern the behavior of followers across the organization. The consistent reinforcement of these rules establishes what Glenn (2004) refers to as *metacontingencies* composed of a series of *interlocking behavioral contingencies* (IBCs) where the behavior of followers, as a group, is dependent on the behavior of leaders. Collections of IBCs produce *macrobehavior* of groups of people. This is what we more commonly think of as cultural practices. Then, the relation between measurable products (e.g., higher employee job satisfaction ratings) resulting from macrobehavior (e.g., employees given more choice in their work activities) is known as a *macrocontingency* (Glenn, 2004). Similarly, desirable consumer-level outcomes (e.g., Gallup, 2003) can be explained by consumer behavior interacting with the positive aggregate product that is the macrobehavior from servant-leadership.

Considering the context sketched above, the overarching goal of the present study was to explore the influence of organization-wide SL programming across a variety of outcome domains. Such an evaluation may be valuable in helping steer the organization's efforts with a data-driven assessment interpreted in a way that allows policy and practice changes to be made based on empirically derived principals from behavior analysis. Due to the need for additional study of practical outcomes associated with SL, and the novel application of a behavior analytic framework in interpreting the mechanisms of change, this study aims to contribute to the larger academic literature on person-centered positive organizational leadership.

Seven specific outcome measures were included in this study (see below). The broad domains of interest include data from client and customer satisfaction surveys, client program records, and human resources metrics, each evaluated before and after SL training. These results may help clarify which domains are meaningfully linked to organizational SL implementation as well as the strength of change over time.

The central hypothesis (**H1**) is that there will be improvement on all quantitative metrics assessed following SL training across sites compared to pre-training. Specifically, increases are predicted for (1) client/caregiver global satisfaction scores, (2–3) client satisfaction survey scores at discharge (two items), (4) percentage of positive discharges, and (5) frequency of internal promotions. Decreases are predicted for (1) number of client safety incidences, (2) rate of employee turnover. The second hypothesis (**H2**) is that centers that have a greater degree of SL plan milestones reached at post will show larger magnitudes of desirable change across all metrics during the study period. The final hypothesis (**H3**) is that the magnitude of change for employee metrics will be generally larger compared to client and caregiver metrics over the study period due to employees contacting the SL change contingencies sooner and more directly.

---

[2] A behavior analytic evaluation of an individual's level of wellbeing involves "(1) one's verbal behavior with respect to one's other behavior and environment, (2) the sets of direct acting contingencies operating in one's environment, (3) the level of coercive control in one's environment, and (4) the degree to which individuals have an optimal level of choice" (Houmanfar, Alavosius, Morford, Herbst, & Reimer, 2015, p. 9).

[3] Mapping onto the definition of wellbeing, a behavior analytic description of prosocial behavior "(1) operates in the context of positive reinforcement contingencies for others, (2) minimizes aversive or coercive conditions and the contingencies of others while not explicitly operating as part of those conditions or contingencies, and (3) aids others in identifying or achieving optimal levels of choice" (Houmanfar et al., 2015, p. 9).

Method

Participants and Setting

This study was conducted with The Devereux Foundation, a non-profit behavioral health organization based in the U.S. All data were drawn from Devereux archives that the company gathers on clients in residential programs, client caregivers, employees involved in direct client service (e.g., clinical or direct care staff) indirect service to the organization (e.g., corporate leadership, human resources). These data are gathered as part of standard operational practice across Devereux centers in 13 states. Exact sample sizes varied by outcome domain and are summarized in Table 1.

Table 1. Summary statistics for all outcome domains compared pre and post the national SL champion training.

Outcome Measure	*N*	Pre		Post	
		M	*SD*	*M*	*SD*
MREs	578	184.29	145.73	147.57	116.02
% Positive Discharges	296	59.13	20.76	50.62	29.74
Employee Turnover	723	109.50	73.17	89.14	56.45
Internal Promotions †	310	19.50	17.56	19.07	17.10
SL Satisfaction Survey Mean	228	-	-	4.24	0.90
Day of Discharge Survey: Satisfaction †	1718	3.32	0.80	3.21	0.85
Day of Discharge Survey: Needs Met †	1674	3.33	0.80	3.27	0.85

Note: For non-survey data, *M* and *SD* statistics reflect values for an average center and *N* reflects counts of programs within centers used as the level-1 observations in the models. Survey data is pooled across all centers and *N* reflects counts of individual survey responses included in models.
† No significant effects.

Materials

External Stakeholder & Customer Satisfaction Surveys. As part of routine satisfaction monitoring efforts, Devereux began disseminating optional SL-aligned satisfaction surveys to (1) all clients over 18 years of age, (2) client caregivers for clients under 18, and (3) external professionals that work in coordination with the organization in some capacity. The surveys contain 18 statements across six domains (i.e., Programs and Services, Customer Service, Employee Interactions, Communication, Teamwork, Credibility) to which respondents indicate the relative-frequency of each item on a five-point scale ranging from 0 = *Never* to 4 = *Always*, or *Unsure*. For the purposes of the analyses, global satisfaction was used as the primary outcome measured by the composite score of all 18 items. Internal consistency of the 18-item composite score was strong after controlling for the specific domains (ω hierarchical = .93). Results were only available at Post making Pre/Post comparisons impossible for this study. See Table 1.

Discharge survey. The client discharge survey is an optional questionnaire offered to all clients and concerned persons upon discontinuation of services. The survey contains several items regarding experiences with Devereux with associated four-point Likert-type response options relevant to the item content. Two specific items were used for this study: "To what extent has our program met your needs?"

and "Overall, how satisfied are you with the services you have received?" See Table 1 for descriptive summaries.

        Program data. Client-level data gathered directly from each program included (1) the number of safety incidents/major risk events (MREs) and (2) the percentage of positive discharges both gathered from archival records. A discharge was considered "positive" if the listed discharge outcome was "successful," the client was reunified with family, was adopted, or moved to a lower level of care if no other information was available. See Table 1 for descriptive summaries.

        Human resources data. Employee data was gathered and analyzed from human resources records. These metrics included records of (1) employee turnover and (2) internal promotions. Results of the SL-aligned employee survey were also intended to be included but the data were not available during the period of the study and this domain was thus discarded. See Table 1 for descriptive summaries of employee turnover and promotions.

Procedures

        The driving force behind each center's SL plan are the SL "Champions" specific to each center. Champions work to create a project charter that outlines the vision of SL at their center that their executive director reviews and approves. All Champions received SL training to increase collective knowledge among the employees at their center beginning with the national SL conference that took place October 31–November 1, 2018. In addition, members of the corporate leadership team consulted on-site with each program beginning January 1, 2019 and concluding by the end of April to provide support and guidance more specific to the particular culture and needs of the facilities. Although Devereux's SL initiative is ongoing, for practical purposes, data were analyzed through April 30, 2019.

        This study primarily utilized a pre/post design to monitor the changes in each of the outcomes (DVs) specified above in the 6-month period before and after November 1, as available, at each Devereux center and aggregated nationally. In addition, because each center is asked to develop a unique timeline for SL rollout, leading to varying implementation schedules, the degree to which centers have reached their charter milestones by the end of the study period was considered to test associations between degree of implementation and outcomes.

Analyses

        Multilevel modeling (MLM) procedures were used to formally test the association and magnitude of effects in each outcome domain while controlling for the inherent clustering of variance at each center. MLM is a flexible statistical framework that allows for modeling relations among variables of various types (e.g., continuous, categorical, count) at higher and lower levels of nesting simultaneously (Bryk & Raudenbush, 2002). For the purposes of this study, 10–14 clusters (i.e., 10–13 centers + Devereux Corporate only for employee outcomes) were included as a random level-2 variable in all models while programs within centers were used as the level-1 unit of analysis for all outcomes excluding surveys where client and caregiver responses were treated as level-1 variables nested within centers. Although centers and programs were modeled with random intercept terms where appropriate, the contribution of random slopes to overall model fit was not tested in the interest of parsimony.

        In order to address **H1**, a level-1 fixed effect term of Time (i.e., 0 = Pre, 1 = Post-training) was included as the lone predictor in each outcome's model. **H2** was then tested by adding a level-2 continuous variable of percentage of SL charter milestones complete by the end of post to the original models used to test **H1**. After results from these models are gathered. **H3** will also be evaluated by comparing model effects informally between employee and client/caregiver outcomes, where employee outcome models are predicted to have larger effects on average. Standard Hox (2010) model building procedures will be used to test the incremental addition of terms to the MLM for each outcome.

Results

        Comparisons of samples sizes as well as means and standard deviations of all outcomes at Pre and Post are presented in Table 1. For non-survey data, *M* and *SD* statistics reflect values for an average center

and *N* reflects counts of programs within centers used as the level-1 observations in the models. Survey data is pooled across all centers and *N* reflects counts of individual survey responses included in models. Centers had an average of 42.9% (*SD* = 23.1%) of their charter milestones reached by the end of the study. The Implementation term was grand-mean centered in all models to aid interpretation of model effects.

       Major risk events (MREs). MLMs for MREs in each program were modeled on a Poisson distribution to accommodate the count data. Results revealed the full model containing main effect terms for Time and Implementation as well as a Time x Implementation interaction was the best fitting, overall. Results from this preferred model are presented in Table 2 where small but significant effects were found for all three terms suggesting that for every 1% increase in charter implementation at Post, there was a 0.5% reduction in the odds of MREs.

Table 2. Preferred MLM for major risk events (MREs).

Fixed Effects	Est. (log odds)	SE	Z	p	Odds Ratio
(Intercept)	1.627	0.250	6.510	<.001	5.089
Time[Post = 1]	-0.167	0.032	-5.243	<.001	0.846
Implementation	0.023	0.010	2.174	< .05	1.023
Time x Implementation	-0.005	0.001	-4.464	< .001	0.995
Random Effects	**Var.**	**ICC**			
Level-2 Intercept: Center	0.948	.393			
Level-1 Intercept: Program	0.733	.509			

       Positive discharges. MLMs for the rate of positive discharges within programs were modeled on a standard Gaussian distribution as the data were continuous and residuals normally distributed. Results indicated that a model containing both main effects of Time and Implementation were significant unique predictors but the interaction term did not improve model fit and was thus excluded from the preferred model presented in Table 3. The Time main effect suggests that there was an approximate 6% decrease in positive discharges at Post, contrary to the hypothesis. However, the significant main effect of Implementation was small but positive suggesting a 0.6% increase in positive discharges for every additional percentage of milestones reached.

Table 3. Preferred MLM for percentage of positive discharges.

Fixed Effects	Est.	SE	t	df	p
(Intercept)	60.827	4.488	13.554	11.134	<.001
Time[Post = 1]	-6.402	2.869	-2.231	134.421	< .05
Implementation	0.635	0.194	3.278	10.401	< .01
Random Effects	**Var.**	**ICC**			
Level-2 Intercept: Center	562.4	.114			
Level-1 Intercept: Program	143.8	.444			
Residual	559.4				

Employee turnover. MLMs predicting employee turnover within programs were modeled on a Poisson distribution to accommodate the count data. Results suggested that there was a small but significant main effect of time where the odds of an employee exiting the company reduced by 8.5% at post overall. However, neither the main effect of Implementation nor the interaction term were significant and were thus excluded from the preferred model presented in Table 4.

Table 4. Preferred MLM for frequency of employee turnover.

Fixed Effects	Est. (log odds)	*SE*	*Z*	*p*	Odds Ratio
(Intercept)	1.033	0.097	10.655	<.001	2.809
Time[Post = 1]	-0.089	0.039	-2.262	< .05	0.915
Random Effects	**Var.**	**ICC**			
Level-2 Intercept: Center	0.094	.106			
Level-1 Intercept: Program/Department	0.480	.540			

Internal promotions. MLMs predicting the number of promotions earned by employees were modeled on a Poisson distribution to accommodate the count data. No significant effects were found for Time, Implementation, or their interaction leading to the conclusion that there is insufficient evidence that Time or Implementation are related to promotions in the context of the study.

SL satisfaction survey. MLMs concurrently predicting mean response to the 18-items on the client and caregiver satisfaction surveys were modeled on a Gaussian distribution given that the data were continuous and residuals normally distributed. Results suggest a small but significant reduction in global satisfaction at Post, contrary to the hypothesis. Full model results are presented in Table 5.

Table 5. Preferred MLM for SL satisfaction survey overall mean.

Fixed Effects	**Est.**	*SE*	*t*	*df*	*p*
(Intercept)	4.212	0.085	49.545	8.486	< .001
Implementation	-0.020	0.005	-4.128	13.108	< .01
Random Effects	**Var.**	**ICC**			
Level-2 Intercept: Center	0.032	.045			
Residual	0.688				

Note. Pre/Post comparisons were not possible due to survey response data only being available during Post.

Day of discharge survey. MLMs predicting discharge survey responses for the "Needs Met" and "Satisfaction" items were modeled on a cumulative logit distribution to accommodate the ordinal data. No significant effects were found in any model for both items. This suggests that there is insufficient evidence that Time or Implementation is significantly related to day-of-discharge survey responses.

Discussion

Summary of Findings

This study examined seven different outcomes of interest to a large behavioral healthcare organization undergoing the process of incorporating practices of servant leadership at each of its centers across the country. Three hypotheses were posited: (**H1**) improvements will be found for each of the outcomes during the 6 months following national SL training compared to the 6 months prior, (**H2**) greater amounts of center-specific charter implementation will be associated with larger improvements from Pre to Post, and (**H3**) generally, employee-level outcomes will show greater improvement compared to client-level outcomes.

H1 was partially supported as there was a statistically significant reduction in MREs and employee turnover from Pre to Post. However, contrary to the hypothesis, the rate of positive client discharges actually decreased over the course of the study and all other outcomes had no significant changes from Pre to Post. The magnitude of effects were also uniformly small.

Evidence for **H2** was similarly mixed. For MREs, while there was a small but significant interaction of Time x Implementation, suggesting greater reductions at Post for centers with more of their SL charter vision implemented, charter implementation in general was curiously positively associated with MREs regardless of Time. Moreover, implementation was negatively related to SL satisfaction responses to a small degree (see Table 5) again counter to what was expected. Implementation was, however, positively associated with rate of positive discharges independent of Time. There were no other meaningful associations with Implementation noted.

Finally, in regard to **H3**, there was one meaningful effect of Time on employee turnover but all other models with significant effects related to client-level outcomes. The client-level MRE model had the strongest associated effects, contrary to what was expected. It is interesting that this was the case given that client-outcomes are more distal in theory than employee outcomes. Future research should incorporate more direct measures of employee satisfaction and engagement along with the same client metrics to see if the trend replicates or if the direct employee surveys are more sensitive to SL implementation and outweigh the presently observed effects of the client outcomes.

Limitations

The results of the study should be interpreted in light of some important limitations. Given the correlational nature of this study design, it cannot be known if observed effects are causally related. It is essential to understanding these results that any observed effects be prudently and conservatively interpreted to not misattribute a causal direction. Relatedly, appropriate caution should also be used to not over-interpret the size of effects found here given the relatively short amount of time SL implementation has been active at the organization. It is important to bear in mind that the Post period reflects only the first 6 months of SL implementation nationally and most of the centers had only reached approximately half of their intended milestones. Future study is needed to compare these metrics again after full implementation at all centers is complete and over a longer period (e.g., 1 year Pre and Post) in order to better gauge the impact of slow systems change and increase our confidence that observed changes may be related to the macrocontingencies (Glenn, 2004) associated with SL taking effect.

Unfortunately, too few responses were returned for the external partner survey ($N = 30$ responses across 11 centers) to analyze with sufficient power to detect effects and was thus excluded from the current study. SL-aligned employee satisfaction survey data also could not be included as data collection began after the end of Post. These data should be included in future research as employee satisfaction is theorized to mediate client outcomes and could be incorporated in more sophisticated structural models that may strengthen the case for causation and lead to a more sophisticated understanding of how the variables of interest influence one another.

References

Biglan, A., & Glenn, S. S. (2013). Toward prosocial behavior and environments: Behavioral and cultural contingencies in a public health framework. In G. J. Madden (Ed.), *APA handbook of behavior analysis* (pp. 255–275). Washington, DC: American Psychological Association.

Biglan, A., Lee, J., & Cody, C. (2016). The evolution of capitalism. In R. D. Zettle, S. C. Hayes, D. Barnes-Holmes, & A. Biglan (Eds.), *The Wiley handbook of contextual behavioral science* (pp. 496–512). West Sussex, UK: Wiley Blackwell.

Gallup (2003). Strengthening care through engagement. Retrieved from http://www.gallup.com/strategicconsulting/157250/strengthening-care-engagement.aspx.

Glenn, S. S. (2004). Individual behavior, culture, and social change. *The Behavior Analyst*, *27*, 133–151.

Greenleaf, R. K. (1977). *Servant leadership: A journey into the nature of legitimate power and greatness*. New York: Paulist Press.

Greenleaf, R. K. (1998). *The power of servant-leadership*. San Francisco: Berrett-Koehler.

Harter, J. K., Schmidt, F. L., Agrawal, S., Plowman, S. K., & Blue, A. (2013). The relationship between engagement at work and organizational outcomes. *Gallup Poll Consulting University Press, Washington*.

Houmanfar, R. A., Alavosius, M. P., Morford, Z. H., Herbst, S. A., & Reimer, D. (2015). Functions of organizational leaders in cultural change: Financial and social well-being. *Journal of Organizational Behavior Management*, *35*, 4–27.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York, NY: Routledge.

Hunter, E. M., Neubert, M. J., Perry, S. J., Witt, L. A., Penney, L. M., & Weinberger, E. (2013). Servant leaders inspire servant followers: Antecedents and outcomes for employees and the organization. *The Leadership Quarterly*, *24*, 316–331.

Laub, J. A. (1999). Assessing the servant organization: Development of the Organizational Leadership Assessment (OLA) model. *Dissertation Abstracts International,* 60 (02): 308A (UMI No. 9921922).

Patterson, K. A. (2003). *Servant leadership: A theoretical model.* Doctoral dissertation, Regent University. ATT No. 3082719.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: SAGE Publications, Inc.

Spears, L. C. (1995). *Reflections on leadership: How Robert K. Greenleaf's theory of servant-leadership influenced today's top management thinkers.* New York: John Wiley.

Van Dierendonck, D. (2011). Servant leadership: A review and synthesis. *Journal of Management*, *37*, 1228–1261.

Williamson, T. (2008). The good society and the good soul: Plato's *Republic* on leadership. *Leadership Quarterly*, *19*, 397–408.

Winston, B. E. (2003). *Extending Patterson's servant leadership model: Explaining how leaders and followers interact in a circular model.* Paper presented at the Servant Leadership Roundtable at Regent University, Virginia Beach, VA.