Evaluation of Shift Positive Method

Q1: Does engaging in the Shift Positive method lead to improvement/professional growth & development?

To answer this question, ally assessments regarding client improvement/progress scores were combined (via simple averaging) across all of a given client's goals to give a single overall score for each ally's assessment of a client's overall.

An overall **score of zero indicates that a client exhibited no improvement**, on average, across all their goals according to a given ally's assessment. **Negative scores indicate that a client exhibited decreased effectiveness**, on average, across all their goals according to a given ally's assessment. **Positive scores indicate that a client exhibited enhanced effectiveness**, on average, across all their goals according to a given ally's assessment.

The average score across all ally assessments was 1.64 (sd=0.71; SE=0.03). Cohen's d suggested that this average score was 2.31 standard deviations higher than zero, indicating that the **shift positive method exhibited a large effect on client effectiveness across all goals**. A single sample t test confirmed that this average score was significantly different from zero ($t(532)=53.27$, $p<0.0001$), indicating a **meaningful improvement in client effectiveness across all goals.**

Figure 1. Average ally assessment of Client Progress

One limitation of the approach outlined above is that it assumes all ally assessments are drawn at random from a normal distribution with an average score of zero. This assumption lies at the core of determining whether our observed scores are "statistically significantly" different from zero, because the shape of the normal distribution allows us to estimate how rare it would be to

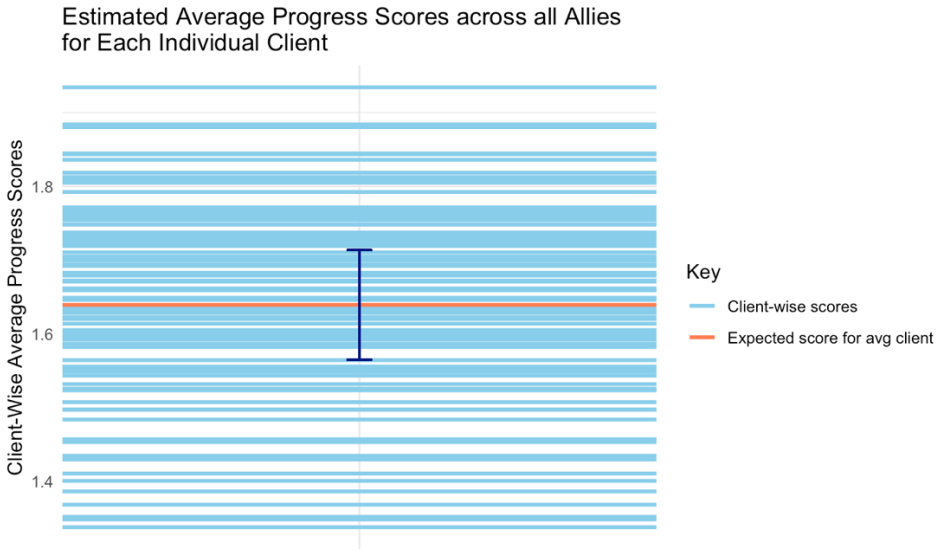
observe the actual average score that we've observed given how tightly clustered our data is around this average score.

Since ally assessments were nested within different corresponding clients, we can't assume that they were 'drawn at random', because allies who are assessing the same client are likely to give more similar assessments to each other than to allies who are assessing a different client. This makes our estimates look more tightly clustered around their average than they would be if they were truly drawn at random, which inflates the statistical significance of our results. Multilevel, or Mixed-Effects models, are a tool that can be used to account for the fact that allies who are assessing the same client are likely to give more similar assessments to each other than to allies who are assessing a different client.

Mixed effects models accomplish this by describing our data at two levels: The within-client (fixed effect) level, and the between client (random effect) level. The fixed effects estimate of these models tell us whether ally assessments are significantly different from zero for a (statistically) average client, while the random effects tell us how these ally assessments vary across clients.

Mixed effects modeling yielded similar results to that of t-testing, with fixed effects estimates telling us that the **expected ally assessment score was** 1.64 (SE=0.04; $p<0.0001$) for a statistically average client, which was itself **significantly greater than zero** (notice that the standard error was slightly larger, lending to a slightly less, albeit still significant result compared to standard t testing). Mixed effects models also have the added benefit of being able to estimate how this expected ally assessment score varies across clients. The random effect estimates tell us that the fixed effect estimate of 1.64 (95% CI [1.57, 1.71]) exhibit a standard deviation of 0.22, suggesting that **a given client would need to be approximately seven standard deviations below the fixed effect estimate in terms of their client assessment scores to exhibit a non-significant progress score**, which gives us a **high degree of confidence that the shift positive methodology will lead to significant improvement/professional growth & development across a wide array of potential clients**.

Figure 2. (Mixed-Effect Model Implied) Average ally assessment of Client Progress across individual Clients along with a (theoretical) statistically average Client



Q2: Correlation between client progress and how frequently the client shared goals with allies.

Like with average ally assessments of client improvement/professional growth & development, it is important to distinguish between within-client and between-client correlation with covariates like how frequently the client shared goals with an ally.

Between-client correlations refer to the correlation between average client level scores across all allies, whereas within-client correlations refer to the correlation between ally-specific deviations from these average client level scores. If these two distinct patterns of correlation aren't distinguished, then the correlations we observe based on raw scores that don't account for the fact that ally assessments tend to be more similar within clients will tend to reflect some mix of the two, which can be problematic if within-client correlations differ from between client correlations. As such, we should examine all three forms of correlation to ensure that they yield similar interpretations of the data.

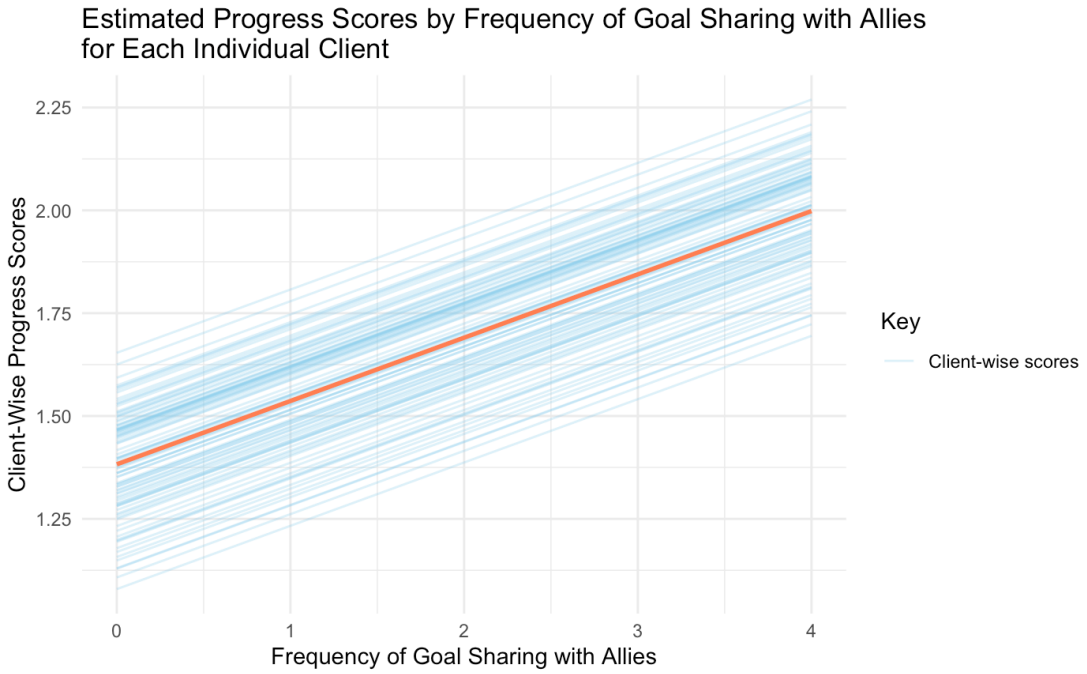
- Total correlation: r = 0.22 (95% CI [0.135, 0.295]), t(529)=4.46, p<0.0001
- Within-Client correlation: r = 0.21 (95% CI [0.13, 0.29]), t(529)=4.36, p<0.0001
- Between-Client correlation: r = 0.22 (95% CI [0.14, 0.30]), t (543)= 4.56, p<0.0001

In this case, **all three correlations give similar results**, indicating that there is a **significant positive relationship between client progress and how frequently the client shared goals with Allies**, which is robust to differences in level of analysis. This means that 1) Clients who share goals more frequently on average across all their allies tend to exhibit higher levels of average reported progress across all their allies, and 2) For any given client, allies who are shared with more often tend to give higher progress ratings.

Mixed effects models allow us to quantify the magnitude of within-client relationships via their fixed effects. With respect to the within-client relationship between goal sharing and client progress, mixed effect modeling found significant fixed effects (within-client effects for a statistically average client). These models implied that **allies whose client never shared goals**

with them were expected to report progress scores of 1.38 (app. 16% lower than the average expected progress score of 1.64 (see section Q1)), with these scores expected to increase by 0.16 (app. 11.6%, p < 0.0001) for every unit increase in reported frequency of goal sharing. Furthermore, random slope variance, which describes the expected degree of variability in the relationship between two variables specified in a model's fixed effects, was found to be non-significant. This means that the **strength and significance of the relationship between within-client goal sharing and progress scores can be expected to remain stable across a wide array of potential clients.**

Figure 3. (Mixed-Effect Model Implied) Relationship between Goal Sharing and Client Progress across individual Clients and a (theoretical) statistically average Client



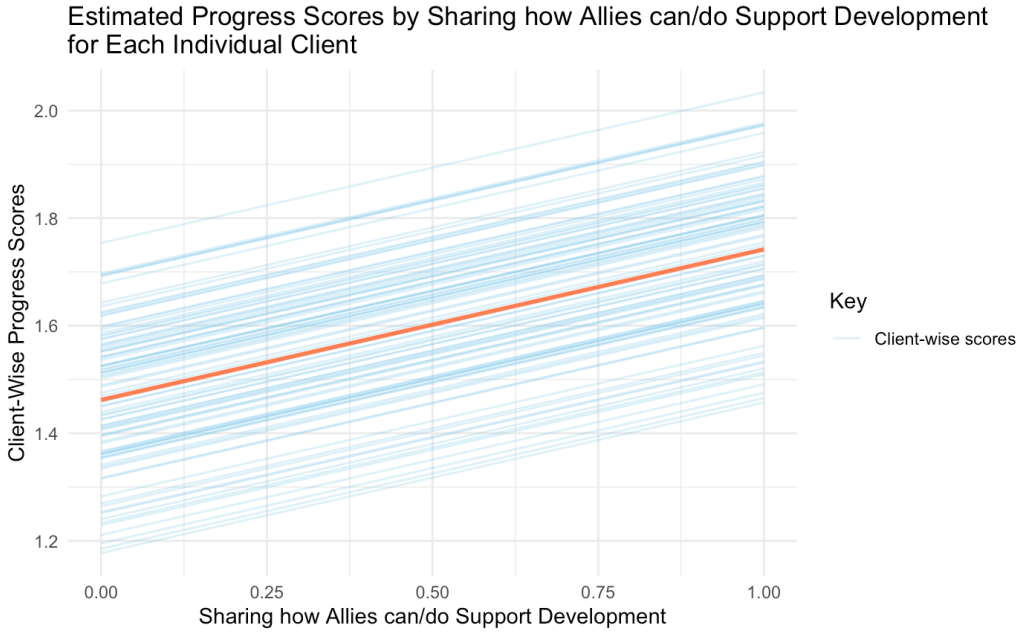
Q3: Correlation between client progress and whether the client spoke with allies about how they can support the client's development

- Total correlation: r = 0.190 (95% CI [0.107, 0.271]), t(530)=5.12, p<0.0001
- Within-Client correlation: r = 0.186 (95% CI [0.103, 0.267]), t(530)=5.00, p<0.0001
- Between-Client correlation: r = 0.192 (95% CI [0.110, 0.272]), t (543)= 5.20, p<0.0001

In this case, **all three correlations give similar results**, indicating that there is a **significant positive relationship between client progress and speaking with allies about how they can/do support the client's development,** which is robust to differences in level of analysis. This means that 1) Clients who speak to a higher proportion of their allies about how they can/do support the client's development tend to exhibit higher levels of average reported progress across all their allies, and 2) For any given client, allies who are spoken to (by the client) about how they can/do support the client's development tend to give higher progress ratings.

With respect to the within-client relationship between progress scores and speaking with allies about how they can/do support development, mixed effect modeling found significant fixed effects (within-client effects for a statistically average client). These models implied that **allies whose client never spoke to them about how they can/do support the client's development were expected to report progress scores of 1.46 (app. 11% lower than the average expected progress score of 1.64 (see section Q1)), with these scores expected to increase by 0.28 (app. 19.2%, p < 0.0001) for allies whose client had spoken to them about how they can/do support the client's development.** Furthermore, random slope variance was found to be non-significant, which means that **the strength and significance of the relationship between within-client progress scores and speaking to allies about how they can/do support the client's development can be expected to remain stable across a wide array of potential clients.**

Figure 4. (Mixed-Effect Model Implied) Relationship between Client Progress and Speaking to Allies about how they can/do support Development across individual Clients and a (theoretical) statistically average Client



<u>Q4: Correlation between client progress and how frequently the allies helped the client work towards their development.</u>

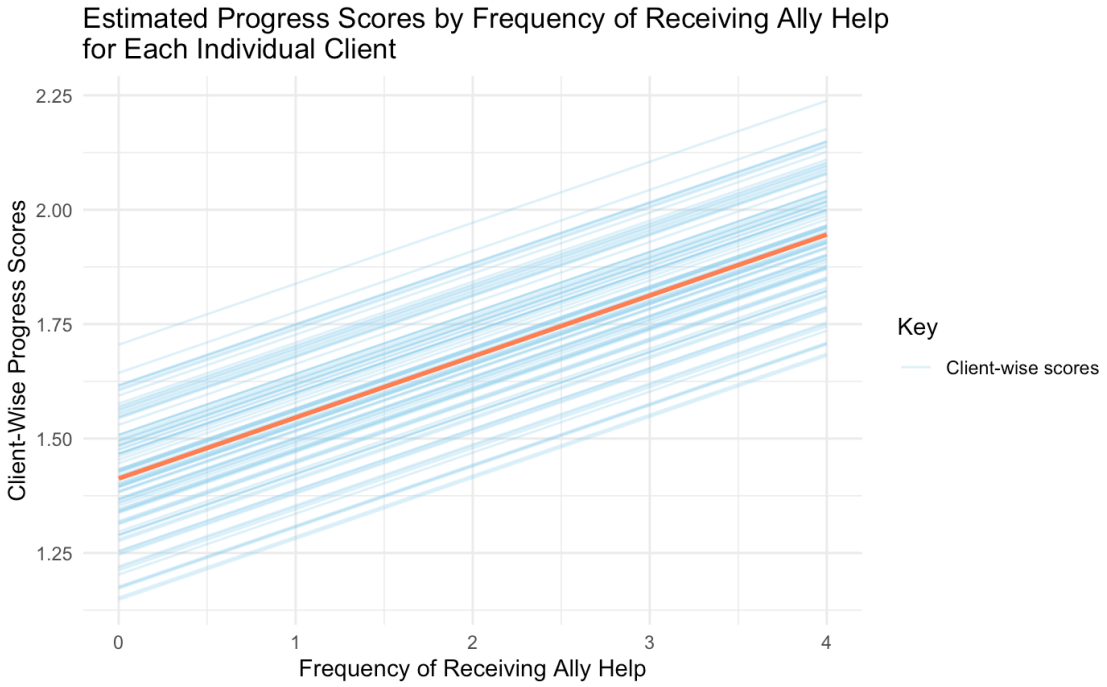
Total correlation: r = 0.22 (95% CI [0.135, 0.295]), t(530)=5.12, p<0.001
Within-Client correlation: r = 0.21 (95% CI [0.13, 0.29]), t(530)=5.00, p<0.001
Between-Client correlation: r = 0.22 (95% CI [0.14, 0.30]), t (543)= 5.20, p<0.01

In this case, **all three correlations give similar results**, indicating that there is a **significant positive relationship between client progress and how frequently allies helped the client work towards their development**, which is robust to differences in level of analysis. This

means that 1) Clients who receive more frequent help on average across all their allies tend to exhibit higher levels of average reported progress across all their allies, and 2) For any given client, allies who help the client more frequently tend to give higher progress ratings.

With respect to the within-client relationship between ally help and progress scores, mixed effect modeling found significant fixed effects (within-client effects for a statistically average client). These models implied that **allies who never helped the client work towards their development were expected to report progress scores of 1.41 (app. 14% lower than the average expected progress score of 1.64 (see section Q1)), with these scores expected to increase by 0.13 (app. 9.2%) for every unit increase in reported frequency of helping.** Random slope variance was found to be non-significant, which means that **the strength and significance of the relationship can be expected to remain stable across a wide array of potential clients.**

Figure 5. (Mixed-Effect Model Implied) Relationship between Ally Help and Client Progress across individual Clients and a (theoretical) statistically average Client



Supplementary Analysis

1. Predicting 'Not Applicable' Response to Question on Frequency of Ally helping from 1) Ally Role and 2) Having been spoken to by the client about how one can/does support their development

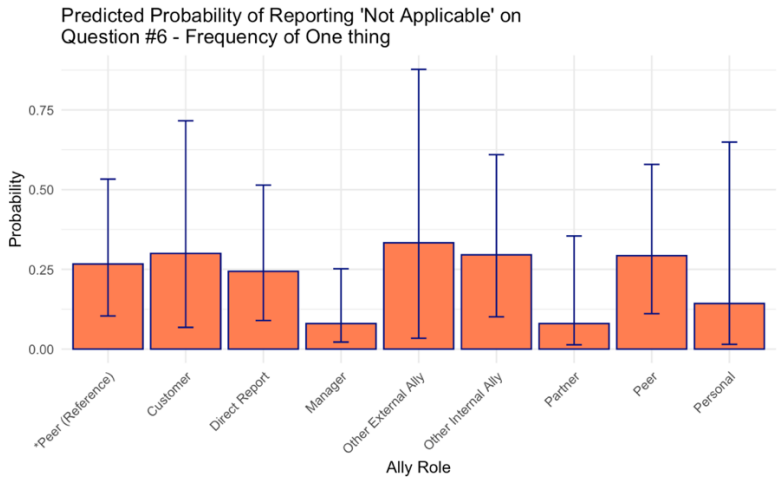
App. 22.5% of allies responded 'Not Applicable' to Question #6 on how frequently they help the client with their development goals. This is problematic because correlations between Client progress and how frequently allies help the client with their development goals are based on a substantially lower proportion of respondents than some of the other correlations discussed above, since we have to drop these 'Not Applicable' responses when computing these correlations. If the population of respondents used to compute these correlations are systematically different, then we become limited in the degree to which we can compare the correlations discussed above (e.g., if the correlation between Client progress and question#6 is based on a population that systematically contains fewer direct report allies because they're are replying 'Not Applicable' to question#6 at a substantially higher rate than we would expect by chance, then we can't directly assess whether the true correlation between client progress and Question #6 is higher or lower than that with say, Question#4 on Goal sharing).

To assess potential factors underlying a client's propensity to give a 'Not Applicable' response to question #6, a number of statistical models were specified to **assess whether 1) Ally Role and 2) Having been spoken to by the client about how one can/does support their development impact the probability of giving a 'Not Applicable' Response to Question #6.**

Statistical testing via Intraclass correlations ($r < 0.05$) found that responses didn't tend to cluster within specific clients to the same degree as they did with the primary analysis, so all analyses were conducted using simple linear (logistic) regression rather than multi-level modeling.

Using peer allies as a reference class, the first model examined the effect of different roles on an ally's probability of responding 'Not Applicable' to Question #6. According to a logistic regression model, all roles demonstrated similar propensities to peer allies, with the exception of **Managers**, who **were significantly less likely than peers to respond 'Not Applicable' to question #6** ($B=-1.43$; $SE=0.69$; $p<0.04$).

Figure S1.



A second model was then specified to assess how an ally's response to Question #5 on whether the client had spoken to them about how they could or do support client development impacted their propensity to respond 'Not Applicable' to question#6. This model revealed that an ally's response to question 5 was a stronger predictor of whether they would respond 'Not Applicable to question #6 than their role, with **allies who had spoken to their corresponding client about how they could or do support client development being significantly less likely to respond 'Not Applicable' to question #6** (B=-3.27; SE=0.632; p<0.0001). **After controlling for ally responses to question #5, ally role no longer served to predict propensities to respond 'Not Applicable' to question 6**, with all ally roles demonstrating similar rates to the peer reference class.

Figure S2.

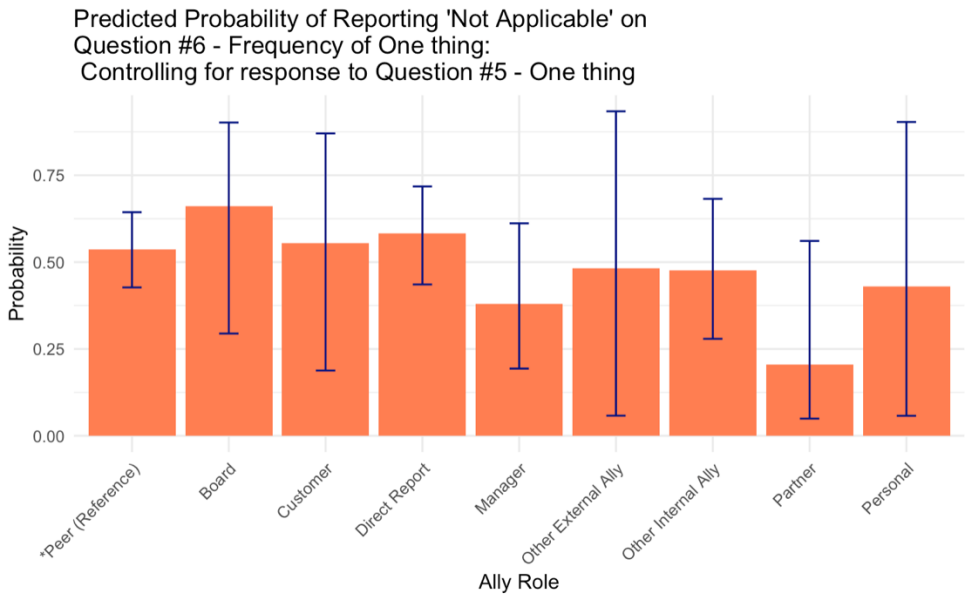
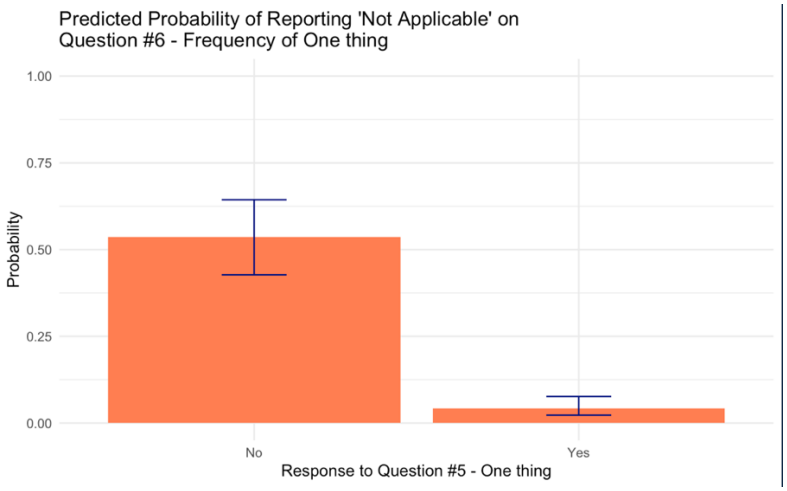


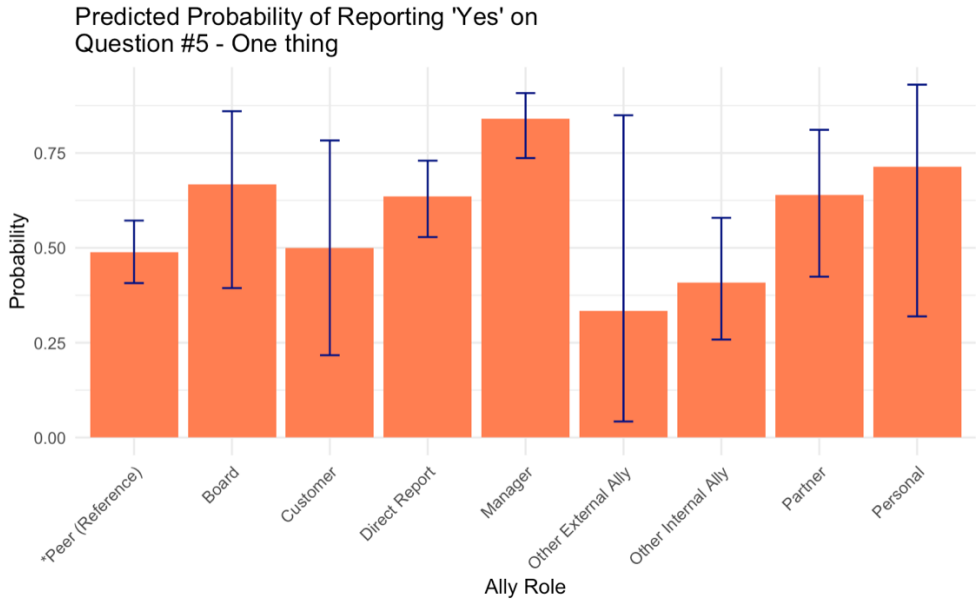
Figure S3.



This pattern of results, whereby Ally role looks like a significant predictor of responding 'Not Applicable' to question#6, until we control for whether an Ally was spoken to by the client about how they could or do support client development, paired with the possibility that an ally's role may predict whether a client speaks to them about how they can or do support development, signals a potential **mediation process, whereby** ally roles only exercise an indirect effect on ally propensities to respond 'Not Applicable' to question# 6 through their effect on responses to Question #5. To examine this possibility, a number of additional models were specified to **assess the degree to which ally roles affected client propensities to discuss how the ally could/does support development**, as well as the **degree to which this effect subsequently spilled over into impacting an ally's propensity to respond 'Not Applicable' to question #6, given that allies who had spoken to their corresponding client about how they can/do support development were substantially less likely to give 'Not Applicable' responses to Question #6.**

The first of these models found ally roles to be a significant predictor of an ally's propensity to have discussed how they can/do support development with their corresponding client. Specifically, **relative to the reference class of peer allies**, **direct reports** (B=0.60; SE=0.22; p<0.01) and **managers** (B=1.70; SE=0.32; p<0.0001) **were both significantly more likely to have discussed ways they could/do support client development**. All other ally roles demonstrated similar rates to peer allies.

Figure S4.

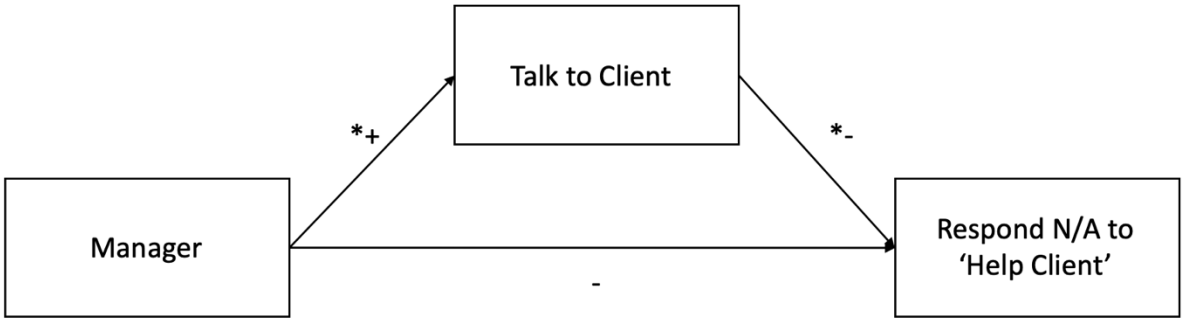


A bootstrapped Sobel test was then run to assess whether the effect of ally role on an ally's propensity to have discussed how they can/do support development with their corresponding client 'spilled over' into their propensity to respond 'Not applicable' to Question #6 on helping frequency. This test indicated that **the indirect effect of ally on role an ally's propensity to respond 'Not applicable' to Question #6 via its effect on an ally's propensity to have discussed how they can/do support development with their corresponding client was**

significant for direct reports (B=-2.09; 95% CI [-3.65, -0.52]) **and managers** (B=-6.07; 95% CI [-8.582, -3.564]), **both of whom were less likely than peers to have responded 'Not applicable' to Question #6 due to their increased likelihood of having actually discussed how they can/do support client development.**

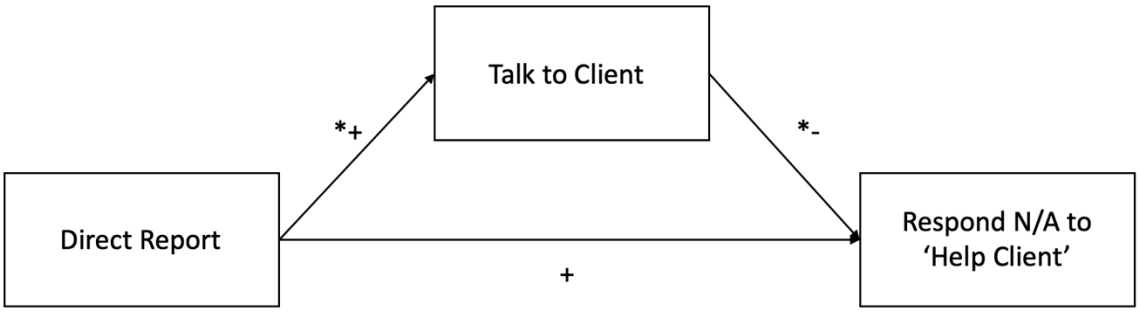
One thing to note is that without controlling to speaking to the client, the effect of being a direct report on an ally's propensity to respond 'Not Applicable' to question #6 looks non-significant. Given the pattern of mediation observed above, this is the result of a **suppressive effect, whereby the overall effect of being a direct report is 'hidden' by the contrasting direction of the direct and indirect components of this effect.** This is because the direct effect of being a direct report on an ally's propensity to respond 'Not Applicable' to question #6, after controlling for whether the ally spoke to the client about how they can/do help, is positive, which contrasts with its negative indirect effect, making it seem like there is no effect occurring when we fail to decompose these effects appropriately. **The same is not the case for managers, whose direct and indirect effects are in alignment with one another**. See figures S5 and S6 for a graphical depiction of the relational structures involving direct Manger and Direct Report status with propensities to speak

Figure S5. Graphical depiction of the relational structure between ally role (Manger), Question#5 ('One Thing'), and whether an ally will respond to Question#6 (Frequency of 'One Thing')



Note: *Indicates significant effects

Figure S6 Graphical depiction of the relational structure between ally role (Direct Report), Question#5 ('One Thing'), and whether an ally will respond to Question#6 (Frequency of 'One Thing')



Note: *Indicates significant effects

These results outline a limitation of the primary analysis, which is that **correlations between client progress and frequency of ally helping (main analysis, Question 4) is based on a subset of the data that contains a disproportionately higher number of direct report and manager allies**, who are more likely than other allies to give a response regarding how frequently they help clients work towards their developmental goals. As such**, we should be careful in generalizes those results to the same populations for which we may want to generalize the results for (main analysis) Questions 2 and 3**.

In either case, the key takeaway here is that, **regardless of an ally's role, speaking to them about how they can support your development is a key facet underlying the degree to which they will feel capable of helping you in working towards your developmental goals further down the road.** Being helped by allies is a key factor underlying developmental progress, as outline above (see primary analysis, Question 4), so **clients should aim to make concerted efforts to discuss with allies how they can or do support their developmental journey's regardless of the ally's role relative to the client.**

12/5/24

Re: Statistical analysis for Change Positive on the Shift Positive method

Hi Pete,

I'm doing well, hope all is good on your end. Questions 1 and 2 get at the more basic assessment of client improvement. The key thing to focus on here is that the unit of analysis is individual evaluations of improvement, of which we had 544 across 94 clients. (because clients were evaluated by multiple allies).

Question 1 uses a t test to give a baseline measure of average improvement across these individual evaluations, and Question 2 uses mixed effects models to get at the question of how these evaluations vary across clients.

The basic descriptives around those figures are as follows: **526 out of 544 ally evaluations exhibited positive improvement in developmental goals**. The average degree of improvement across these individual evaluations was +1.64 (on a scale of -3 to +3), and the p value for that was <0.000000000001, basically implying that if we were to randomly take another 544 evaluations, we would expect a **less than 1 in a trillion chance** of the average evaluation score being zero or below. That in and of itself isn't super helpful from a narrative crafting perspective, so it makes more sense to look at client-wise improvements averaged across all allies.

When evaluations were averaged out across allies (Question 2), all 94 Clients demonstrated positive improvement in developmental goals, with very little variability around the average degree of improvement, which again was +1.64. The standard deviation across client average scores was 0.22, which means that a client's improvement score, averaged across all their corresponding allies, would need to be 7 standard deviations away from the average to be a zero or less, which corresponds to approximately 1 in 700 billion client assessments being zero or below. The lowest improvement for a client was 0.533, which means that even if the results you got in this round of evaluations were some kind of crazy fluke, and that the lowest scoring client's improvement was really representative of the type of average improvement people should expect to see, a client's improvement score, averaged across all their corresponding allies, would still need to be 2.4 standard deviations away from that average to be a zero or less, which corresponds to approximately approximately 1 in 120 client assessments being zero or below. Those numbers might seem a bit out there, but they're reflective of the fact that client improvement scores were super consistent across all the clients assessed, **which means that if someone signs up for your program, they are highly likely to see an improvement score that's really close to 1.64**

I ran an **additional analysis** that can serve as a supplement that might make for better narrative crafting, which was **running t tests separately for each individual client (See excel sheet attached).** That analysis found **statistically significant improvement in 78 out of 94 clients (83%)**, which is a more digestible metric, but also less rigorous from a statistical perspective because the variation in scores within a given client is based on such a small number of ally assessments. The issue with this is that it can lead to a high degree of false negatives because the smaller the number of allies assessing a given client is, the more likely it is that those allies will give highly discrepant scores, making the average score between them look less precise. This means that, **of the clients who did not show significant improvement, it's highly likely that they would have if they'd been evaluated by more people**. I tested that explicitly by looking at differences in the number of allies between clients with significant average scores and those without (See excel sheet attached), and found that **clients with significant t-tests were evaluated by 3.45 more allies on average than those with non-significant t-tests**, and that difference itself was statistically significant ($p<0.0001$), which supports the idea that most of those non-significant t tests would have been significant if those clients were evaluated by more allies.

The mixed effects modeling approach I took for question 2 averages out the within-client variation in ally assessments to give a better estimate of how much you can expect ally assessments to vary around any given client's average score (**when they are evaluated by an average number of allies**), as well as how much you can expect the average client-level scores to differ from each other when generalizing to future clients. I think that method, while more complex, gives a better picture of just how effective your program was for the clients you evaluated, whereas just running t tests within each client individually severely underestimates that effectiveness. **Sorry for the long-winded answer on this, you have a really impressive set of results here and I just want to make sure I'm doing them justice.** Hope this helps, happy to chat through ideas for presenting this information any time if you'd find that useful as well.

Cheers,
Danny

From: Pete Berridge <pete@shiftpositive360.com>
Sent: Thursday, December 5, 2024 1:40 PM
To: Daniel Martin <daniel.j.martin@cgu.edu>
Subject: Re: Statistical analysis for Change Positive on the Shift Positive method

Danny,

How are you doing? I just wanted to reach out to ask a basic question that Jamie and I were asking today about the results. It's the first most basic question - how many people improved or significantly improved for participating in the Shift Positive 360? I see that it says the Shift Positive method exhibited a large effect on client effectiveness across all goals. How can we say that in a straightforward way, e.g., "9 out of 10 participants showed significant or highly significant improvement in their developmental goals"?

I'm just looking for that headline about improvement before getting into the detail of sharing goals, allyship, etc.

Thank you Danny,

Pete

Pete Berridge, MAPP, PCC
www.shiftpositive360.com
pete@shiftpositive360.com
612-270-0860

October 29, 2024

Hi Pete,

Hope all is well. I've managed to make a first run at analyzing the progress survey data. Attaching my report here, along with the code I put together to generate the results (just for transparency's sake, happy to walk you through any piece of that if it's at all of interest to you). In general, I found a lot of positive results. It looks like engaging in the Shift Positive method does lead to improvement/professional growth & development. There is also a strong positive correlation between client progress and 1) how frequently the client shared/talked about her/his/their goals with the allies/respondents, 2) whether the client spoke to allies about how they could or do support he client's development ("One Thing") , and 3) extent to which the allies/respondents helped ("Frequency of One Thing"), all of which can be expected to hold across a wide array of (potentially) diverse clients.

I also ran some supplementary analyses that showed that clients were more likely to speak to allies about how they could or do support he client's development ("One thing") if the allies in question were managers or direct reports, which in turn lead to an increased propensity for ally's to actually answer the "Frequency of One Thing" Question (i.e. not respond 'Not Applicable'). I thought that spoke to the importance of clients making a concerted effort to think about ways that non manager or direct report allies can support their development, and subsequently share those thoughts with those other classes of ally to ensure that they're receiving support from more sources. Happy to hop on a call to discuss further. Hope it helps!

Cheers,
Danny

Daniel Martin
Claremont University