Performance Based Call Routing Pilot Test Report

March 25, 20XX

By

Tom Pyzdek

Table of Contents

Executive Summary

A test was conducted during the month of February, 20XX to determine if routing calls to the available agent with the highest average contact resolution rates would improve the overall rate of contact resolution for the entire group. A small number of agents who handled the same type of calls was selected for the test. The test indicated a statistically significant increase of 2% in overall contact resolution. Routing calls in order of contact resolution had no statistically significant impact on either agent satisfaction or average handle time.

Recommendation

A 2% improvement in average contact resolution in a group that already has very high and very consistent contact resolution is worth pursuing with other groups. I recommend that interested business units proceed to the next phase of this project, which is deployment of the new routing protocol to a larger group. In addition to verifying that the predicted improvement can be replicated, during Phase II we will need to consider what is needed to make the change welcome to agents and what technology changes are needed, if any.

Test Overview

Nineteen XYZ technical support agents were divided into test and control groups. Groups were matched to have statistically similar baseline contact resolution rates. Calls were routed at random to agents in the control group. Calls were routed to agents in the test group based on their average contact resolution rates. I.e., the available agent with the best average contact resolution rate received the call. The test was "blind," i.e., agents, coaches, and supervisors were not aware that a test was being conducted.

To avoid the possibility of bias due to receiving calls before or after the test agents, the control group was split into "before" and "after" subgroups. The before subgroup of controls received calls (at random) before the test group. When all agents in the before subgroup were busy, calls were routed to the test group in order of their individual average contact resolution rate. When all agents in the before control subgroup and the test group were busy, calls were routed at random to agents in the after subgroup. Each agent spent one week in the before subgroup, one week in the after subgroup, and two weeks in the test group.

Agent satisfaction could not be measured directly because of a moratorium on employee surveys for six weeks prior to the "Great Place to Work" survey. Agent satisfaction was estimated by asking coaches and supervisors to rate their perception of each employee's satisfaction using a five point satisfaction/dissatisfaction scale.

Appendix: Details of analysis

Experimental Hypotheses and Hypothesis Test Results

H_0: Baseline test group and control group overall contact resolution will be the same.

Agents were divided into control and test groups prior to the experiment. The contact resolution rates for the baseline period (a recent quarter preceding the experiment) are shown below. The mean difference in contact resolution was 0.04%.

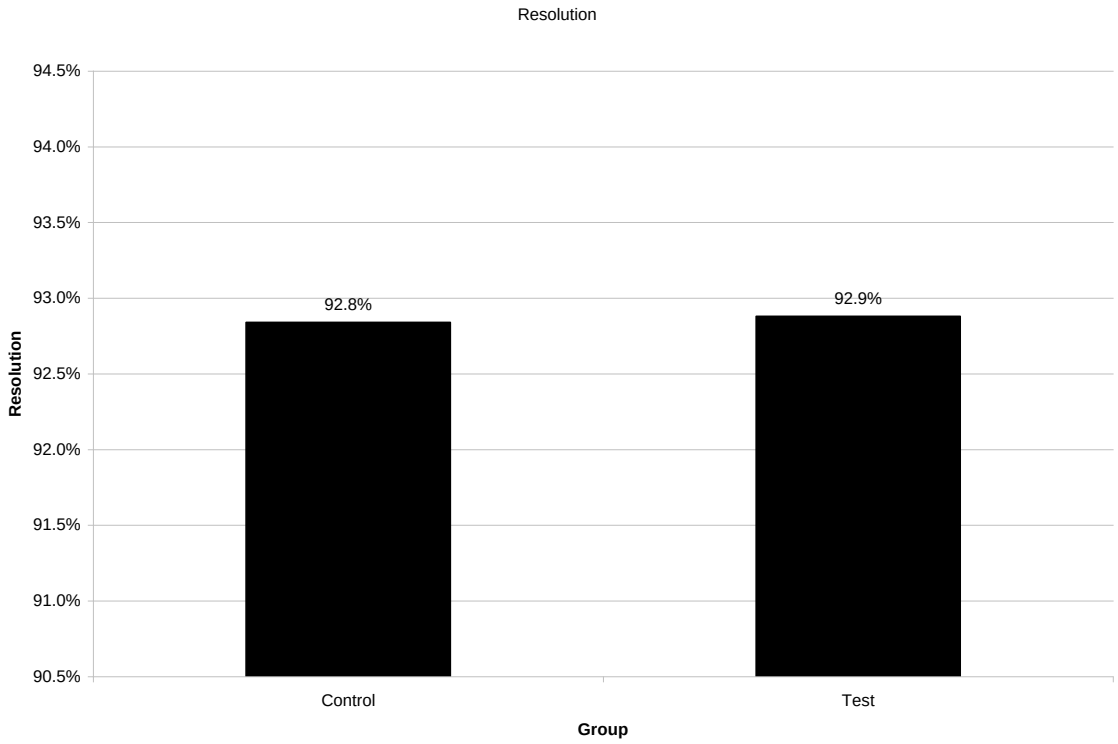
group comparisons

Group	No	Yes	sum	p-bar
Test	291	4089	4380	7.12%
Control	236	3297	3533	7.16%
	527	7386	7913	

A chi-square test of the contact resolution rates of two groups was conducted.

Chi-square Expected Counts		
Group	No	Yes
Test	292	4088.3
Control	235	3297.7
P =	**0.95**	

Conclusion: We fail to reject H_0 (P = 0.95) and conclude that prior to the experiment the test and control groups had the same contact resolution rates.



H_1: Test group overall contact resolution will be the same as the combined control groups.

Analysis

Group * Resolution Crosstabulation

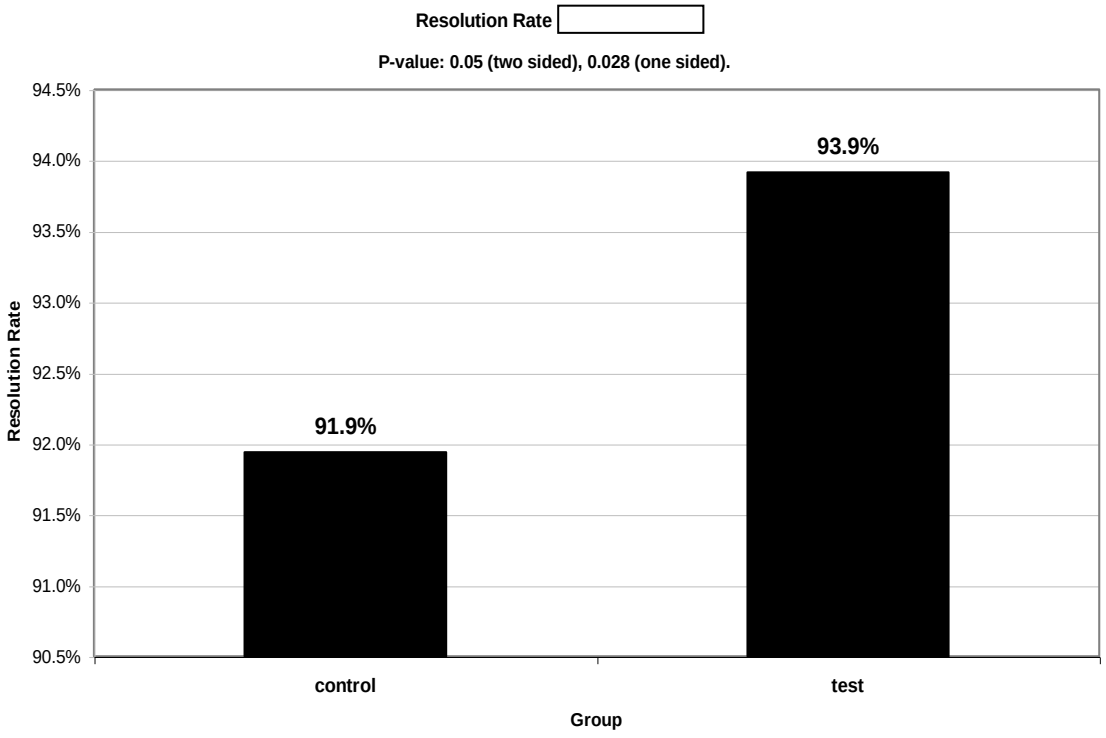
Count

		Resolution		
		no	yes	Total
Group	control	113	1289	1402
	test	77	1188	1265
Total		190	2477	2667

Chi-Square Tests

	Value	df	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Fisher's Exact Test			**.050**	**.028**
N of Valid Cases	2667			

Conclusion: We reject H_1 (P = 0.050 (two sided), 0.028 (one sided)) and conclude that the resolution rate of the test group is significantly higher than that of the control group.[1]



[1] Test group results are 2% higher.

H_2: Test group contact resolution improvement will be unchanged as test group utilization increases.

Nonparametric Correlations[2]

		(ACD+Hold+ACW)/ Staffed Time	Average Resolution
Spearman's rho	(ACD+Hold+ACW)/ Staffed Time	Correlation Coefficient	1.000	-.046
		Sig. (1-tailed)	**.**	**.207**
		N	324	324
	Average Resolution	Correlation Coefficient	-.046	1.000
		Sig. (1-tailed)	.207	.
		N	324	324

Conclusion: We fail to reject H_2 (P = 0.207) and conclude that there is no correlation between occupancy and average contact resolution.

[2] A non-parametric correlation coefficient was used because neither variable is normally distributed.

H_3: Agent satisfaction will be the same for the test and the control groups.

Analysis[3]

A repeated measures Univariate ANOVA was conducted.[4]

Tests of Between-Subjects Effects
Dependent Variable: Satisfaction rating

Source		Type III Sum of Squares	df	Mean Square	F	P	Noncent. Parameter	Observed Power(a)
EXPGRP (test group vs. controls)	**Hypothesis**	.184	1	.184	.064	**.801**	.064	.057
	Error	103.724	36.223	2.863(b)				
AGENT(EXPGRP)	**Hypothesis**	105.810	36	2.939	9.562	.000	344.223	1.000
	Error	51.334	167	.307(c)				
RATER	**Hypothesis**	1.380	2	.690	2.245	**.109**	4.490	.452
	Error	51.334	167	.307(c)				
RATER * EXPGRP	**Hypothesis**	1.107	2	.553	1.800	**.169**	3.600	.372
	Error	51.334	167	.307(c)				

a Computed using alpha = .05

b .971 MS(AGENT(EXPGRP)) + 2.877E-02 MS(Error)

c MS(Error)

Conclusions

1. The EXPGRP comparison indicates that there is no significant difference between average agent satisfaction ratings for the test and control groups (P = 0.801).
2. The differences in agent satisfaction ratings between raters across the control and test groups is not significant (P = 0.109).[5]
3. The difference in agent satisfaction ratings between raters within the control and test groups is not significant (P = 0.169).

[3] This analysis excluded baseline ratings.

[4] With a repeated measures ANOVA the group-to-group comparison is made after factoring out agent-to-agent differences. I.e., we are interested in how individual agents' performance changes as they are moved between test and control groups, not in the differences between agents.

[5] The results of one of the raters were excluded due to inconsistency with the other three raters. See H_9 for details.

H_4: "Before" and "After" control group contact resolution will be the same.

Analysis

A one-way ANOVA was conducted and a "contrast" requested to compare the two control groups. The contrast is shown in the table below as contrast #2.

Contrast Coefficients

	Experimental group (coded)		
Contrast	**Control After**	**Control Before**	**Test**
1	1	1	-2
2	1	-1	0

Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
CR_BAR	Assume equal variances	1	-.0395	.02134	-1.851	2340	.064
		2	-.0210	.01476	-1.421	2340	.155
	Does not assume equal variances	1	-.0395	.02127	-1.856	2286.977	.064
		2	**-.0210**	.01568	-1.337	1176.681	**0.181**

CONCLUSION: Using contrast 2 we fail to reject H_4 (P = 0.181) and conclude that the contact resolution rates for the before and after control groups are equal.

H_5: "Before" and "After" control group agent satisfaction will be the same.

Analysis

A repeated measure Univariate ANOVA was conducted.[6]

Tests of Between-Subjects Effects
Dependent Variable: Satisfaction rating

Source		Type III Sum of Squares	df	Mean Square	F	P	Noncent. Parameter	Observed Power(a)
Intercept	Hypothesis	1407.310	1	1407.310	894.452	.000	894.452	1.000
	Error	57.855	36.772	1.573(b)				
GROUP	Hypothesis	0.02	1	0.02	0.013	**.910**	0.013	0.051
	Error	57.855	36.772	1.573(b)				
AGENT(GROUP)	Hypothesis	59.384	36	1.650	5.576	.000	200.743	1.000
	Error	18.637	63	.296(c)				
RATER	Hypothesis	.867	2	.433	1.465	.239	2.930	.302
	Error	18.637	63	.296(c)				
RATER * GROUP	Hypothesis	.498	2	.249	.842	.436	1.685	.188
	Error	18.637	63	.296(c)				

a Computed using alpha = .05

b .944 MS(AGENT(GROUP)) + 5.627E-02 MS(Error)

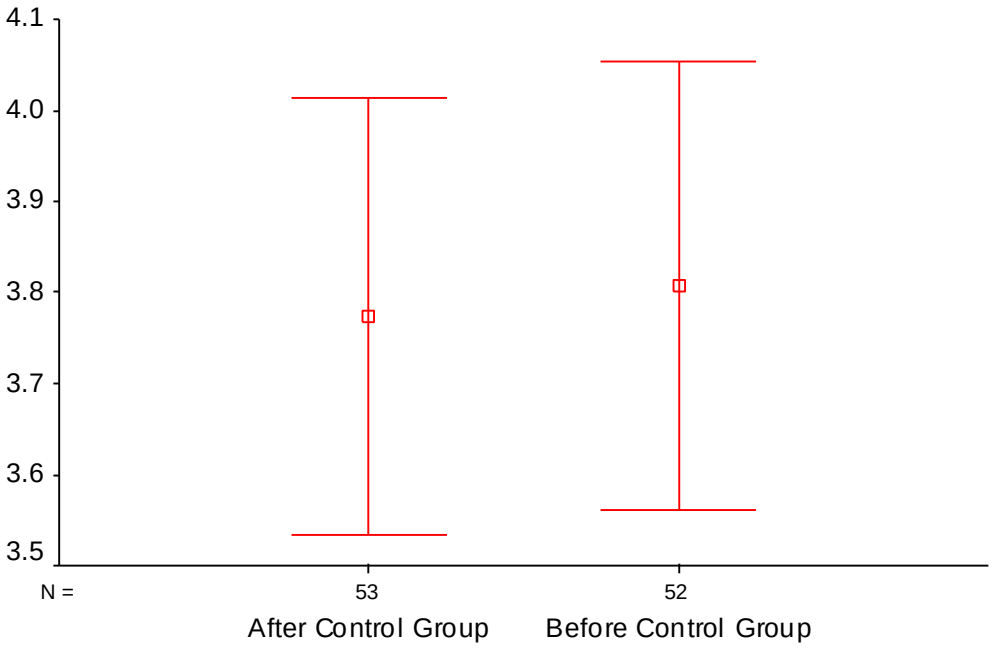
c MS(Error)

Conclusion: We fail to reject H_5 (P = 0.910) and conclude that there is no significant different between average agent satisfaction ratings for the before and after control groups.[7]

[6] The group-to-group comparison is made after factoring out agent-to-agent differences.

[7] Analysis excluded test and baseline groups and the ratings of one of the raters. The differences between the remaining raters is not significant, see H_9 for details.

Before versus After Control Group



Group

Excludes Rater X's ratings

H_6: "Before" and "After" control group calls-per-half-hour will be the same.

Test: ANOVA with contrasts. Agents: controls.

Descriptives
number of calls

	N	Mean	Std. Error
Control After	1458	1.52	.022
Control Before	1620	1.55	.021
Test	2851	1.55	.016
Total	5929	1.55	.011

ANOVA

number of calls

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.132	2	.566	.819	.441
Within Groups	4096.868	5926	.691		
Total	4098.000	5928			

Contrast Coefficients

	Group Membership		
Contrast	Control After	Control Before	Test
1	-1	1	0

Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	**Sig. (2-tailed)**
number of calls	Assume equal variances	1	.03	.030	1.081	5926	**.280**

Conclusion: We fail to reject H_6 (P = 0.280) and conclude that the difference in mean number of calls handled by the two control groups is the same.

H_7: NP score will be the same for "Before" and "After" control groups and Test group.

Test: ANOVA of mean net promoter score. Agents: all.

Descriptives
NP_BAR

	N	Mean	Std. Deviation
Control After	588	7.8342	2.42453
Control Before	635	7.7740	2.49204
Test	1120	7.8107	2.30386
Total	2343	7.8067	2.38559

ANOVA
NP_BAR

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.141	2	.570	.100	.905
Within Groups	13327.276	2340	5.695		
Total	13328.416	2342			

Contrast Coefficients

	Experimental group (coded)		
Contrast	Control After	Control Before	Test
1	**1**	**1**	**-2**
2	**1**	**-1**	**0**

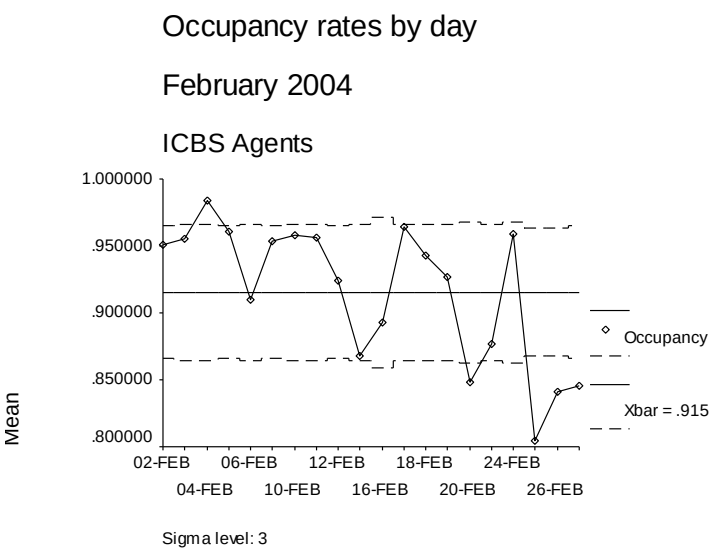
Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	**Sig. (2-tailed)**
NP_BAR	Assume equal variances	1	-.0132	.19747	-.067	2340	.947
		2	.0602	.13658	.441	2340	.660
	Does not assume equal variances	**1**	**-.0132**	.19681	-.067	2335.965	**.946**
		2	**.0602**	.14063	.428	1218.014	**.669**

Conclusion: We fail to reject H_7 (P > 0.66 for all contrasts) and conclude that the difference in net promoter score between before and after control group is not statistically significant,.

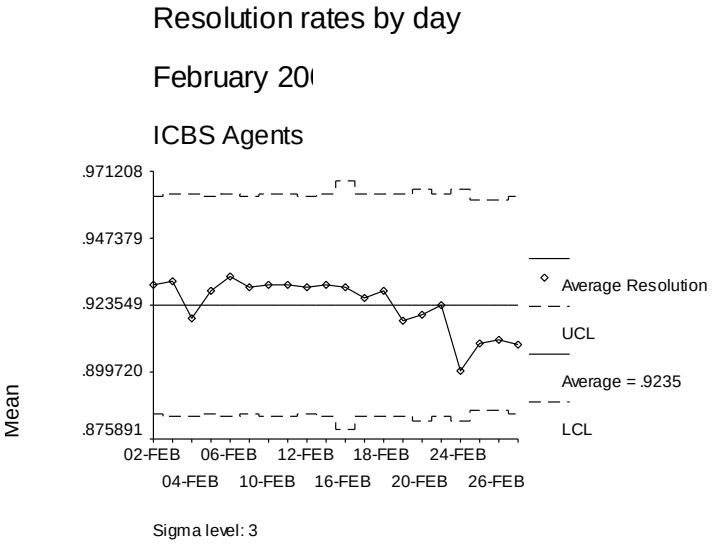
H_8: No time-related patterns in key metrics.

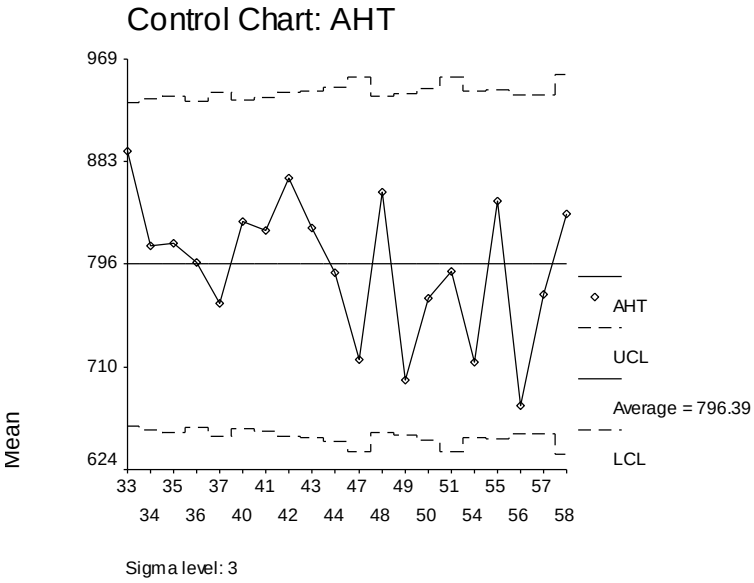
Test: X-bar chart by day for each metric.



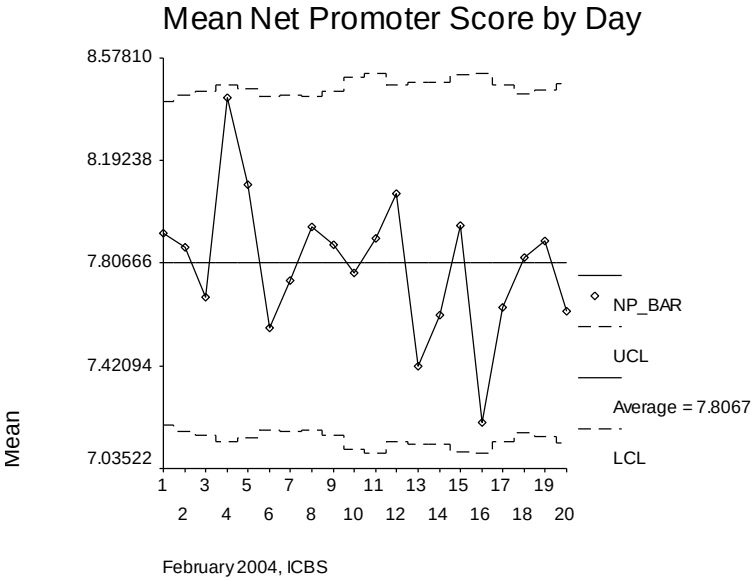
Conclusion: Reject H_8 for occupancy and conclude that it varied due to special causes throughout February.

Conclusion: We fail to reject H_8 for contact resolution and conclude that it was not influenced by special causes in February.



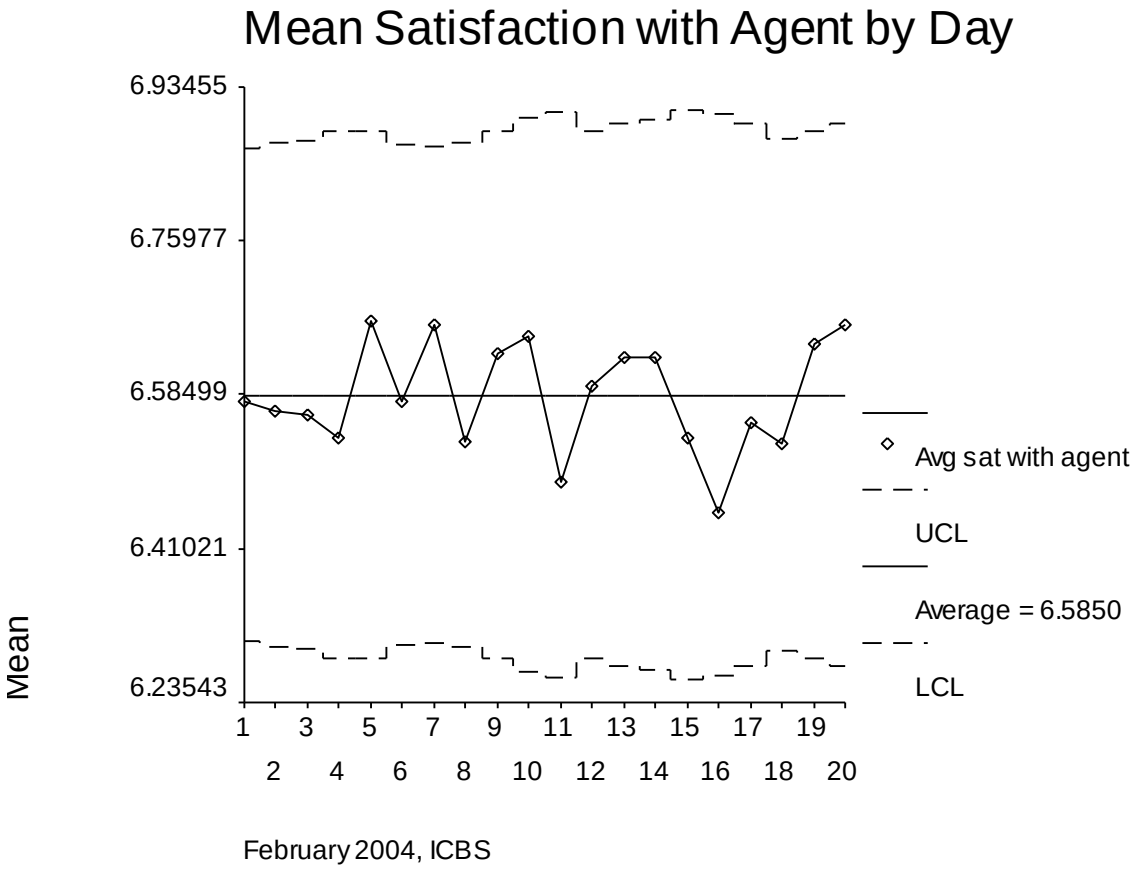


Conclusion: We fail to reject H_8 for AHT and conclude that AHT was not influenced by special causes in February.[8]



Conclusion: We fail to reject H_8 for net promoter scores and conclude that they were not influenced by special causes in February.

[8] AHT averaged 13.3 minutes in February for XYZ.



Conclusion: We fail to reject H_8 for satisfaction with agent and conclude it was not influenced by special causes in February.

Conclusions regarding H_8:

Although statistical stability was not exhibited by occupancy[9] during this pilot test, the conclusions are not affected because all comparisons of interest are between test and control groups. I will assume that the special causes affecting the overall process occupancy metric impacted both test and control groups approximately the same. No other important variables exhibited special cause variation.

[9] *Recommendation*: Management should use control charts on an ongoing basis for key process metrics to detect and identify the special causes when they occur.

H_9: Agent satisfaction ratings will be consistent for different raters.

Test: ANOVA with contrasts.

Descriptives
Satisfaction rating

	N	Mean	Std. Deviation
Rater X	95	3.39	.879
Kirby	95	3.69	.923
Leigh	95	3.61	.971
Myrna	76	3.76	.831
Total	361	3.61	.913

ANOVA
Satisfaction rating

	Sum of Squares	df	Mean Square	F	Sig.
Between Raters	7.081	3	2.360	2.875	**.036**
Within Raters	293.063	357	.821		
Total	300.144	360			

Conclusion: We reject H_9 (P = 0.036) and conclude that there is a difference between raters.

Further analysis revealed that Rater X's ratings caused the inconsistency. The analysis was repeated without Rater X's ratings.

Multiple Comparisons of raters (excluding Rater X)
Dependent Variable: Satisfaction rating
Tamhane[10]

(I) Rater	(J) Rater	Mean Difference (I-J)	Std. Error	**Sig**.
Kirby	Leigh	.13	.149	**.760**
	Myrna	-.10	.142	**.858**
Leigh	Kirby	-.13	.149	**.760**
	Myrna	-.23	.149	**.322**
Myrna	Kirby	.10	.142	**.858**
	Leigh	.23	.149	**.322**

Conclusion: With these three raters we reject H_9 (P > 0.32 for all comparisons) and conclude that the raters are consistent.[11]

[10] Tamhane test does not require equal variances.

[11] As a result of these findings Rater X's rating were dropped when analyzing agent satisfaction.

H_{10}: Agents handle the same number of calls for each split.

Test: Chi-square. Rows = Agents, columns = Splits.

LOGIN_ID * SPLIT_NBR Crosstabulation
Count

		SPLIT_NBR		Total
		144	380	
LOGIN_ID	11373	101	228	329
	11496	57	120	177
	11627	83	180	263
	11851	102	204	306
	12031	88	212	300
	32205	63	142	205
	35420	134	273	407
	35460	119	259	378
	35476	95	227	322
	35509	88	198	286
	35524	60	209	269
	35637	58	129	187
	35656	88	185	273
	35665	115	219	334
	35678	105	228	333
	35740	138	254	392
	35852	69	191	260
	35867	62	144	206
	35871	89	190	279
Total		1714	3792	5506

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	19.776	18	.346
Likelihood Ratio	**20.371**	**18**	**.312**
Linear-by-Linear Association	.036	1	.849
N of Valid Cases	5506		

a 0 cells (.0%) have expected count less than 5. The minimum expected count is 55.10.

Conclusion: We fail to reject H_{10} and conclude that agents handle the same numbers of calls from each split. Don't analyze splits separately.[12]

[12] These two splits accounted for 93% of the calls in February. Two other splits were excluded from the analysis.

H_{11}: Agent satisfaction will not decline as occupancy increases.

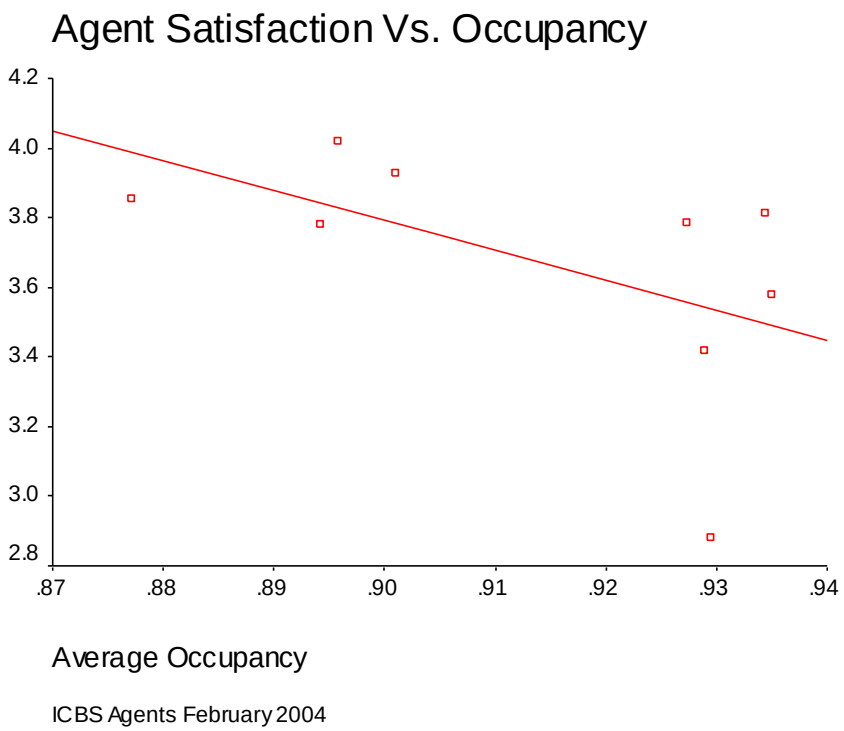
Test: Correlation between agent satisfaction and occupancy.

Nonparametric correlations

			(ACD+Hold+ACW)/ Staffed Time	average satisfaction
Spearman's rho	(ACD+Hold+ACW)/ Staffed Time	Correlation Coefficient	1.000	-.109
		Sig. (1-tailed)	**.**	**.025**
		N	324	324
	average satisfaction	Correlation Coefficient	-.109	1.000
		Sig. (1-tailed)	.025	.
		N	324	324

* Correlation is significant at the .05 level (1-tailed).

Conclusion: We reject H_{11} (P = 0.025) and conclude that agent satisfaction declines as occupancy increases.



Discussion: Agent satisfaction declines as occupancy increases. Although the relationship is weak (r = -0.11) it is real and needs to be addressed. Also, the increase in average occupancy was **not** caused by performance based routing. It may simply be that when a pool of agents is very busy their satisfaction declines.