Dual Monitor Study

A Lean Six Sigma Black Belt Project at Chewy.com

About Me

Tyler Vincent

University of Southern California

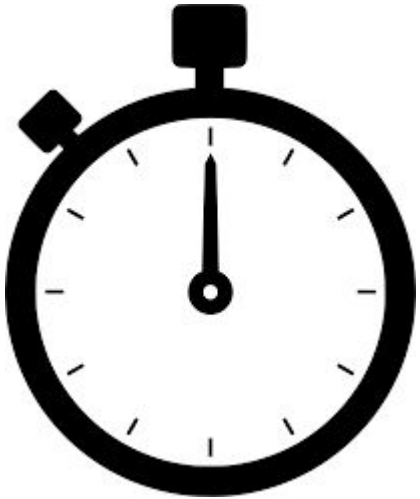
B.S. Industrial Systems Engineering
Honors with a Minor in Writing

Prev. Lean Six Sigma Engineer II at
Chewy.com

Industry Specific Metrics

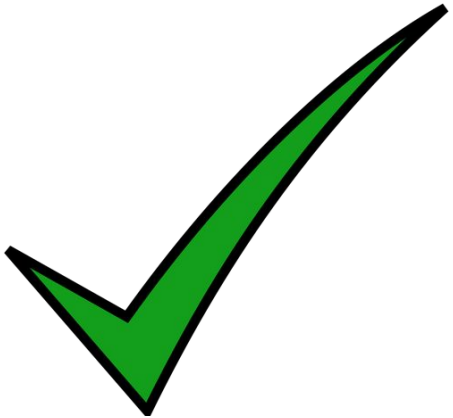
Average Handle Time

- Average time spent assisting each customer



First Contact Resolution

- Resolve customer problems completely in the first interaction.



Overview & Background

The Problem

High cognitive load & variance on customer service agents setups, is decreasing our ability to consistently satisfy our customer in a timely fashion

Why It Matters?

Improved agent performance leads to better customer service

Potential cost savings through:

- Reduced rework costs by increasing first contact resolution
- Reduced labor costs by decreasing the average handle time
- Reduced churn by streamlining the Customer and Employee experience with satisfaction

The Pitch

Proposed Project

- A month long at home study
- A larger sample size of 150 TX and FL agents

Stakeholders

- CS Operations
- IT
- Facilities
- Finance
- Quality and Assurance

Stakeholder Buy-In: Clear communication of potential benefits & Regular stakeholder updates

Goal

Primary Metrics

- Reduction in Average Handle Time (AHT)
- Increase in First Contact Resolution (FCR)

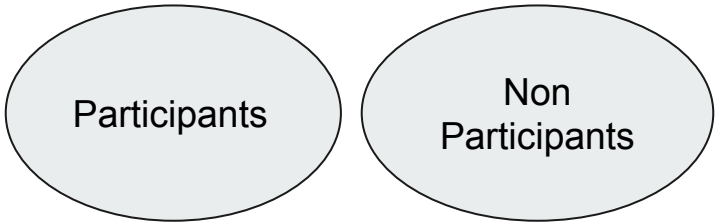
Secondary Metrics

- Customer Satisfaction CSAT
- Employee Satisfaction
- Window Switches (Discrete)
- Mouse Movement (Continuous)
- Work Not Inbound (WNI)
- Contacts Per Hour (CPH)

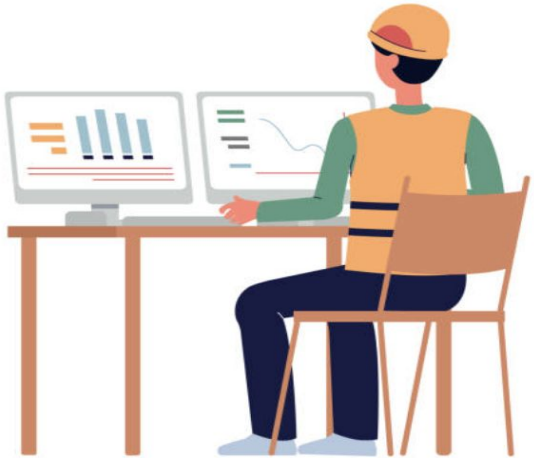
Best Goal Baseline

Study Execution

1. **Data Collection Methods**
   - Performance:
     - Self queried SQL data from data lake
     - Clickstream Data from Oracle via SPLUNK processing language
   - Qualitative:
     - Customer Satisfaction via Quality data connectors in Tableau
     - Administered surveys



2. **Analysis Tools**
   - Statistical Process Control
   - Hypothesis Testing
   - Correlation Coefficients



Findings

- ★ Decreased Task time: AHT for Voice was positively impacted by provision of a second monitor.
- ★ Unaffected Quality: FCR was not statistically different.
- ★ AHT for Email increased slightly; however, email agent data was significantly impacted by special cause variation
- ★ Increased employee satisfaction, reduced movement

Key Observations & Further Analysis

- Improves user satisfaction and reduces stress levels
- Improves task time in all cases but one
- Identified unexpected special cause variation event
- Variation in equipment setup and initial training significantly impacts performance despite tenure

- Action steps taken:
  - Identified and excluded dates most affected by this variation for a clearer data analysis.
  - Reported this variation to stakeholders for transparency and accurate interpretation of results.

Solution

1. Dual Monitor Opt-In option for all CS agents.

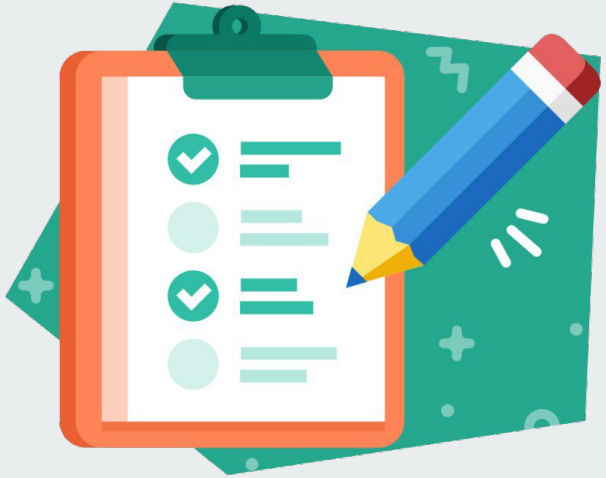
2. Dual Monitors are only offered after 30 days (i.e., once they've graduated from training) to minimize onboarding complexity and equipment costs*

3. Creation of Knowledge Base training for best Dual Monitor Setup as well as best practices/suggested setups for each channel type

4. Roll out our second monitor opt-in process for existing TM and new hires after 30 days of service in future onboardings*

* NOTE: Equipment breakage hovered around 63% yearly, (getting only 37% of equipment returned yearly with attrition). Costs related to breakage impacted the entitlement (ROI) minimum, eliminating some other options. In addition, Chewy attrition was ~50% at the time of this project annually..

Sustainment Plan



1. New training for all agents equipped with a second monitor that highlighted best practices
2. Developed troubleshooting aid for agents and funnel for IT to handle dual monitor requests
3. Monitor the KPIs following deployment including a 3 and 6 month check in with stakeholders

Impact

Projected Annualized Cost Impact

$716K

Averaged savings from reduced AHT

Preference for two monitors

96.8%

Agents overwhelming prefer two monitors

Savings from Voice Agents

-5.3 sec/contact

For a 5M monthly contacts

Lessons Learned

- Reduce selection bias through random sampling
- Align leadership expectations up front about statistical certainty against time/cost constraints
- Allow for more flexibility and build in more fat into statistics so when circumstances change, so that there is more breathing room if the data sample must be reduced
- Factor in possible special cause variations during study design to account for their impact
- Buy a second monitor


THANK YOU

Appendix

Timeline

Preliminary Study Conclusion

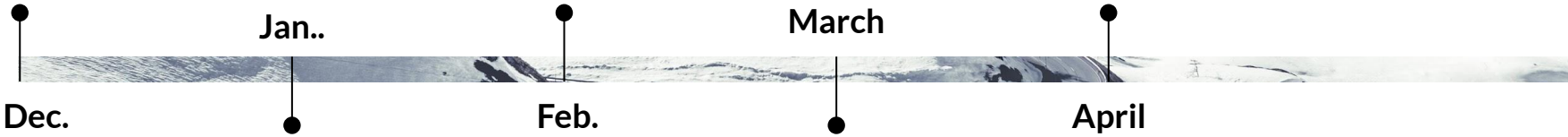
Findings from the first study lead me to recommend with a second study for further certainty

Start Study

After distributing monitors, begin monitoring data for and participants for issues.

Implementation

Roll out dual monitors to entire agent population, update training with SOP and best practice, setup sustainment plan and 6 month audit of progress



Establish project scope

Introduce new parameters and size that have a significant relationship with primary metrics. Include stakeholders and timeline

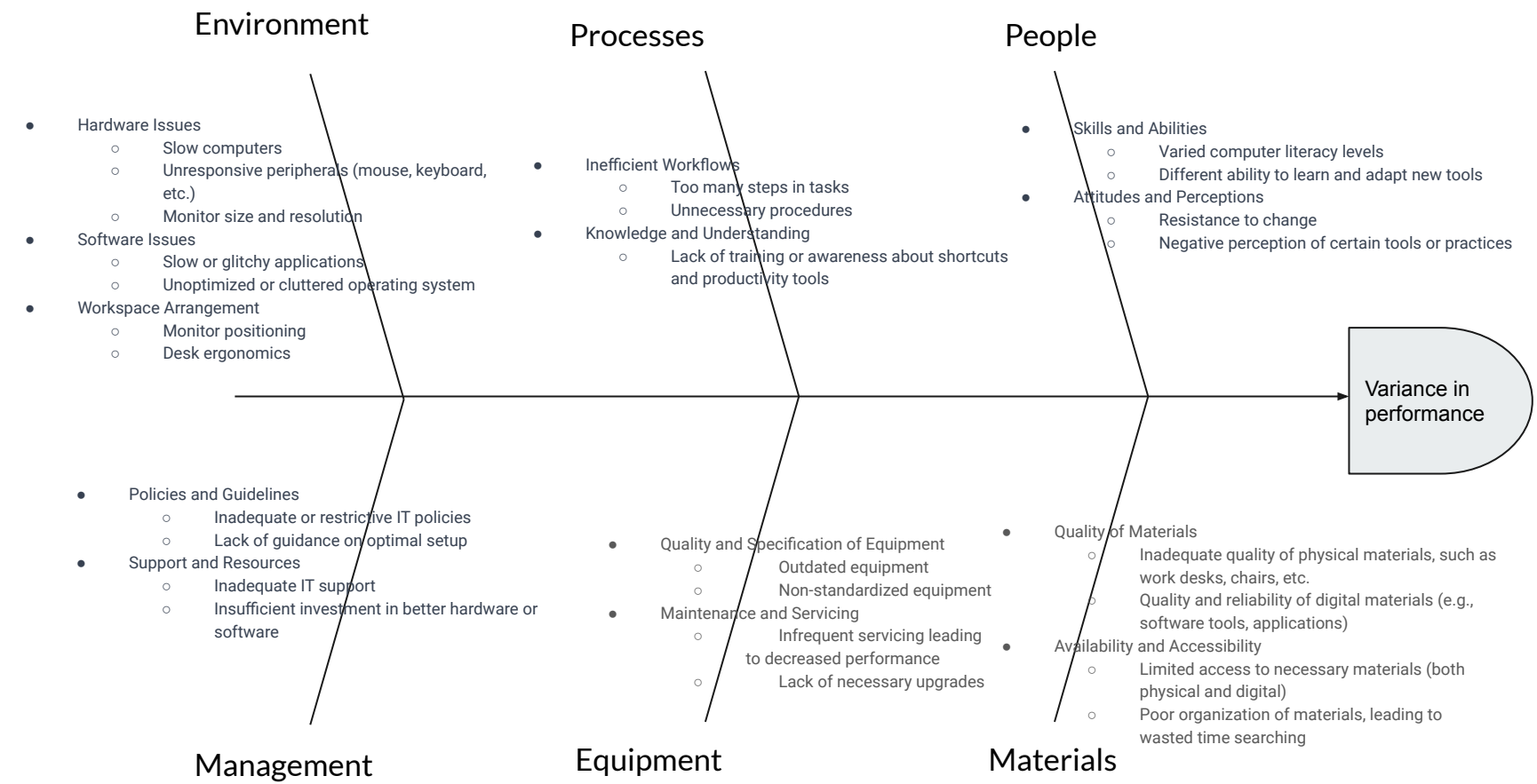
Conclusion & Analysis

Write white paper on findings, resolve any issues, decided recommended solution with leadership

2021

2022

Fishbone Ishikawa Diagram

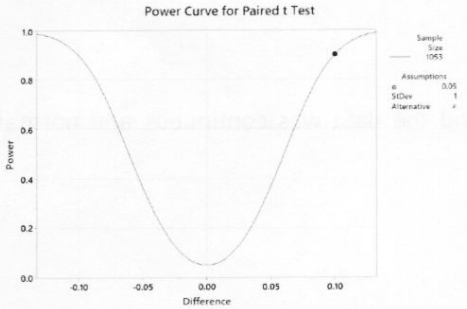


RACI

Task/Role	Sr Director, CS Operations	Associate Director, CS Workforce Mgmt.	Associate Director, Technical Support Ops	Sr Director, CS Program Mgmt.	Sr Director, CS Product	Director, Continuous Improvement	LSS Engineer 2 (ME)	Finance Manager	Associate Director, Facilities	Associate Director, IT
Project Initiation and Planning	A	C	R	C	C	I	R	C	I	C
Preparation and Setup	C	C	C	C	C	I	R	A	R	A
Pre-Study Training	C	R	C	C	C	I	A	I	C	R
Implementation and Data Collection	C	A	C	C	R	I	R	I	I	C
Data Analysis and Reporting	C	C	C	C	C	I	A	C	I	I
Implementation Plan Development	A	C	R	C	R	C	R	I	I	C
Full-Scale Implementation and QA	A	R	C	R	C	C	R	C	R	A
Post-Implementation Review and Closure	A	R	C	R	C	R	A	C	I	C

Sample Size Estimation (1 out of 2)

This graph illustrates how much accuracy the preliminary study had—95% certainty in detecting a 10% shift in Std. Dev. The second part calculate how many data points could we get with a larger study. The next slide goes into the difference that it will make.



Conclusion: We would need 1053 data points for each population (one monitor & two monitors) to detect a 10% shift in standard deviation with 95% certainty.

Paired t Test
Testing mean paired difference = 0 (versus ≠ 0)
Calculating power for mean paired difference = difference
α = 0.05 Assumed standard deviation of paired differences = 1

How many data points will a 4-week study yield for each population?*

Voice: $16{,}800 = 120\ Agents\ x\ \frac{10\ interactions}{day}\ x\ \frac{7\ days}{1\ week}\ x\ 2\ weeks$

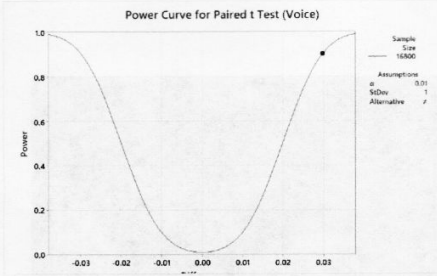
Email: $4{,}200 = 30\ Agents\ x\ \frac{10\ interactions}{day}\ x\ \frac{7\ days}{1\ week}\ x\ 2\ weeks$

*agents often take more than 10 interactions a day, but this scenario assumes a low data collection rate of 10 / day

Sample Size Estimation (2 out of 2)

Here, I calculate how much a larger sample will improve the accuracy. Actual data sample was much larger, and was even more accurate than being able to detect a 3% shift with 99% certainty

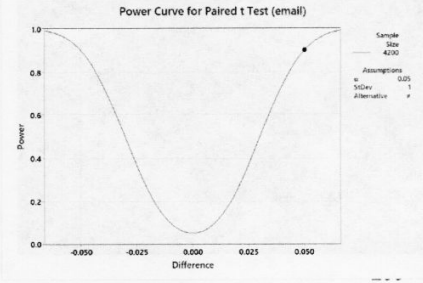
How accurate will that make our hypothesis test results (Paired T Test)?



Results

Sample Size	Power	Difference
16800	0.9	0.0297633

Conclusion: With 16,800 data points, we could detect a 3% shift in standard deviation with 99% certainty with voice.



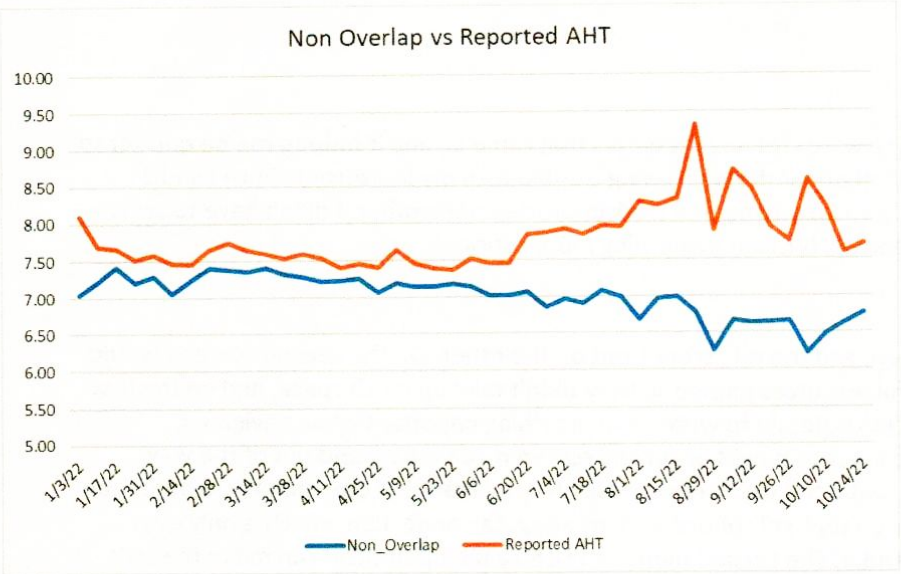
Results

Sample Size	Power	Difference
4200	0.9	0.0500291

Conclusion: With 4,200 data points, we could detect a 5% shift in standard deviation with 95% certainty with email.

With agents taking often 4-5x more than the amount of incidents calculated above (10), we expect to present our final findings with greater confidence than shown here. Even in a worst-case scenario, there will be more than enough data to make decisions with minimal alpha/beta risk.

Isolating Special Cause Variation



Blue line indicates the true email AHT without rollover calls that email agents were taking.
Orange line indicates the email AHT used in this study, including rollover calls

*This graph reflects the task time of email agents versus time. I removed any agents taking rollover calls, who were part of this special cause variation, and the leftover agents' avg task time is shown by the **blue** line. Preliminary results prior to identifying the special cause, would be the **orange** line.*

Charter

Problem Statement:
In our remote work environment, customer service agents primarily worked with a single monitor, leading to high variance and underperformance in agent tasks due to increasing complexity of tasks and multiple systems to manage simultaneously.

Project Objective:
Evaluate the effect of dual monitors on agent performance, specifically looking at task completion speed, task quality, and agent satisfaction.

Background:
This project arises out of the observed extreme variability in simple tasks for users of the same tenure based on their setup. Preliminary studies suggest a potential decrease in Average Handle Time (AHT) when a second monitor was used, but a larger sample size and longer duration is required for statistical validation.

Project Scope:
- Implement a comprehensive study involving customer service agents across all channels (email, chat, phone).
- Distribute, troubleshoot, and setup monitors across sites with the help of IT and Facilities.
- Coordinate with Finance for the capital expenditure request needed to procure extra monitors.
- Ensure proper auditing of interactions with the help of QA to maintain quality standards.

Goals:
- Improve speed of task completion.
- Improve quality of tasks.
- Improve agent satisfaction.
- Standardize the equipment setup across the organization.

Critical Success Factors:
- Achieve statistically significant results that validate the impact of dual monitors on agent performance.
- Successful implementation of a standardized dual monitor setup across all agents.

Stakeholder Involvement:
- Regular bi-weekly updates on the project's progress, next steps, completed tasks, and decision items for stakeholders.
- Foster a culture of continuous communication, participation, and progress sharing to keep everyone engaged.

Risks & Mitigation Strategies:
- Variability in agent setups: Careful design of the study to ensure fair comparison between single and dual monitor setups.
- Resistance to change: Regular communication and engagement with agents to help them understand the benefits of dual monitors.
- Technical difficulties in monitor setup: Assistance from IT and Facilities teams to ensure smooth setup.

Project Timeline:
On other slide

As part of the discovery phase, this charter provided a clear understanding of what the Dual Monitor Study project aimed to achieve, however, the final results, and executive summary was written up as a White Paper.

Preliminary Study

The first study illustrates that there isn't enough data to confidently reject the null hypothesis. I will need more data since this small difference is within Std. Dev., to determine whether a relationship exists between two monitors and better performance.

Time Period	AHT* - Test Group	AHT - All Others
BASELINE "Before test period": 10 Days Before Test	401	387
DAY 1 "Calibrate Day": Agents in lab setting using single monitor	410	382
DAY 2 "Train dual monitor day": Agents practicing with 2 monitors	422	387
DAY 3 "Test Dual monitor day": Agents testing using 2 monitors	398	284

Hypothesis

H_0 = There is no significant AHT decrease using two monitors over one ($P >= 0.05$)

H_a = There is a significant decrease using two monitors over one in AHT ($P < 0.05$)

Statistical Test

Paired t-Test of Test Group Agents, One-tailed

- 12 second decrease in AHT
- p = .33

We failed to reject the H_0 hypothesis (in other words we "accept" the H_0 Hypothesis).

Conclusion

Agents in the test group had **12** second lower AHT in aggregate on 3/23 vs. the 3/21 calibration date. These results are favorable, but not statistically significant, with a p-value of **.33.**

We failed to reject the H_0 hypothesis (in other words we "accept" the H_0 Hypothesis). This high p-value indicates we have weak evidence that H_a is true. More data will help us determine with more certainty that this not a false negative.

Dealing with Resistance to Change

- **Education and Communication**
  - Shared results of initial studies
  - Demonstrated how potential benefits outweigh costs
- **Engagement and Involvement**
  - Involved agents and representatives from IT, Finance, and Facilities in decision-making
- **Support and Facilitation**
  - Provided necessary training and troubleshooting for dual monitor setup
- **Negotiation and Agreement**
  - Provided regular progress updates
  - Addressed concerns and reached compromises when necessary
- **Manipulation and Co-option**
  - Emphasized potential benefits in specific areas to gain support, such as improved performance metrics for CS managers and cost savings for Finance

END