

PSP AI Agent Mentor Application: Comprehensive Testing Document

Table of Contents

1. [Introduction](#)
2. [General PSP & FBS Knowledge](#)
3. [DIVE Process Overview](#)
4. [PSP Culture & Team Structure](#)
5. [Problem Solving Report \(PSR\)](#)
6. [DEFINE Stage Testing](#)
7. [INVESTIGATE Stage Testing](#)
8. [VERIFY/IMPLEMENT Stage Testing](#)
9. [ENSURE SUSTAINMENT Stage Testing](#)
10. [Cross-Functional Testing](#)
11. [Validation Criteria](#)
12. [Advanced PSP Concepts & Common Pitfalls](#)

Introduction

This comprehensive testing document outlines the evaluation framework for the PSP AI Agent Mentor Application. The document contains 300 test questions organized by topic area, along with detailed test scenarios derived from the PSP Mentor Testing Plan. The testing approach follows the DIVE methodology (Define, Investigate, Verify/Implement, Ensure Sustainment) to ensure the AI mentor can effectively guide users through the complete problem-solving process.

General PSP & FBS Knowledge

Foundational Concepts

1. What is the Problem Solving Process at Fortive?
2. What does PSP stand for?
3. What is the Fortive Business System (FBS)?
4. How does PSP relate to the Fortive Business System?
5. What are the key fundamentals of FBS?
6. When should we use the Problem Solving Process?

7. What is the relationship between PSP and Kaizen methodologies?
8. How does PSP help deliver extraordinary results at Fortive?
9. What percentage of time should teams spend in Define and Investigate stages?
10. Why is PSP considered "what we do when we don't know what to do"?

DIVE Process Overview

Core Framework

1. What does DIVE stand for in the problem-solving process?
2. What are the four stages of the DIVE process?
3. Is DIVE a linear or iterative process?
4. Why are there arrows drawn both ways between DIVE stages?
5. What is the recommended time allocation across DIVE stages?
6. Why should teams spend 70% of their time in the first two stages?
7. What happens if we skip stages in the DIVE process?
8. Can teams go back to previous stages during DIVE?
9. What are the main deliverables from each DIVE stage?
10. How does DIVE ensure we address root causes, not just symptoms?

PSP Culture & Team Structure

Cultural Elements

1. What are the key elements of Fortive's problem-solving culture?
2. What does "Extraordinary Teams" mean in the PSP context?
3. What is meant by "Relentless Curiosity" in problem solving?
4. How does transparency support the PSP culture?
5. What is the "Deep Dive" mindset?
6. Why is diversity important in problem-solving teams?
7. What does "beginner's mindset" mean in PSP?
8. How do we create a safe space for problem solving?
9. Why should we avoid blame in problem solving?
10. What does it mean to be a "problem solver" at Fortive?

Team Structure

1. What roles are needed in a problem-solving team?
2. What is the role of the Team Leader?
3. What does a Sponsor do in PSP?
4. What are the responsibilities of a Facilitator?
5. Who should be Team Members?
6. Can one person serve multiple roles in a PSP team?
7. What is the Circle of Concern?
8. What is the Circle of Control?
9. How do we align Circle of Concern with Circle of Control?
10. What happens when Circle of Control doesn't cover Circle of Concern?
11. How do we determine if we have the right team for a problem?
12. When should we expand the team?
13. When should we reduce the problem scope?
14. What skills should PSP team members have?
15. How do we match the correct team to the problem?

Problem Solving Report (PSR)

PSR Fundamentals

1. What is a Problem Solving Report (PSR)?
2. What is the purpose of the PSR?
3. Is the PSR the actual work or just documentation?
4. When should the PSR be created?
5. How does the PSR evolve during the process?
6. What are the main sections of a PSR?
7. Why is PSR considered strictly a communication tool?
8. How often should the PSR be updated?
9. Who should have access to the PSR?
10. What artifacts are consolidated into the PSR?

DEFINE Stage Testing

Problem Definition

1. What is the first stage of DIVE?
2. What are the three specific steps in the DEFINE stage?
3. What is a problem statement?
4. What does TAGS stand for?
5. What are the four elements of TAGS?
6. What is the difference between Trend, Actual, Gap, and Standard?
7. How do we formulate an effective problem statement?
8. What is a Run Chart?
9. Why do we need Run Charts with problem statements?
10. What is the difference between caused and created problems?
11. What are characteristics of caused problems?
12. What are characteristics of created problems?
13. How do we determine if a problem is caused or created?
14. What is Mount Fortive analogy about?
15. What is the Point of Recognition?
16. How do we determine problem scope?
17. What is the Question Burst methodology?
18. How long should a Question Burst session last?
19. What are the rules for Question Burst?
20. What are catalytic questions?
21. What common challenges do teams face in DEFINE?
22. What is meant by "jumping to conclusions"?
23. How do we avoid poor problem definitions?
24. What deliverables come from the DEFINE stage?
25. Why is rationale important in problem definition?

Test Case 1.1: Problem Statement Creation (TAGS Format)

- Test scenario input: "Our AMR channel is underperforming. We have a goal of \$110M but are only projected to hit \$85M. The trend is worsening month over month."
- Expected AI mentor behavior: Guide user to structure this into TAGS format, ask clarifying questions for each component, validate completeness before proceeding

- Expected output validation: Trend (Worsening), Actual (\$84,889k projected), Gap (\$25,513k), Standard (\$110,402k goal)

Test Case 1.2: Run Chart Generation

- Test input: Monthly data from PSR (Jan-Dec projections vs goals)
- Expected output: Visual run chart with time-denominated x-axis (months), goal line (dashed blue), actual projection line, gap visualization, and glide path (dashed gray)

Test Case 1.3: Caused vs Created Classification

- Test scenario: Present the AMR POS problem
- Expected AI response: Guide user to classify as "Created Problem" with rationale that it's a new higher target (\$110M vs previous performance) in Policy Deployment/Strategic initiative context

Test Case 1.4: Circle of Control/Concern Assessment

- Test input: Team composition from PSR
- Expected validation: Assess if team (Joel Mann, Joe Wang, Boris Ilic, Emma Deye, Petula Kelson, Ray Roman, Nate Otto, Matt Ochs, Chris Bohn, Adrian Pierce) can control identified problem scope

Gemba Concepts

1. What is Gemba?
2. What does "Go to Gemba" mean?
3. What are the three reals of Gemba?
4. Where is Gemba in manufacturing?
5. Where is Gemba in service industries?
6. When should teams visit Gemba?
7. What are the 4W's to ask at Gemba?
8. How do we make Gemba visits safe?
9. Why should we avoid asking "why" initially at Gemba?
10. What facts should we gather at Gemba?
11. How many times should teams visit Gemba during PSP?
12. What does "fact-based discovery" mean?
13. Why is first-hand information important?
14. How do teams assess facts at Gemba?

15. What's the difference between Gemba and conference room problem solving?

Tree Concept

1. What is the Tree Concept in problem solving?
2. What do leaves represent in the Tree Concept?
3. What do roots represent in the Tree Concept?
4. What does the trunk represent in the Tree Concept?
5. How do we separate problems from symptoms and causes?
6. What happens first - causes or symptoms?
7. How do we apply the Tree Concept?
8. What is the Point of Recognition in the tree?
9. What are the risks of picking problems too high in the tree?
10. What are the risks of picking problems too deep in the tree?
11. How do we choose the right perspective for the problem?
12. Why do symptoms distract from root causes?
13. How do we sort facts using the Tree Concept?
14. What questions help identify symptoms vs causes?
15. How do we save insights from the tree analysis?

Run Charts & Trends

1. What are the six common ways to describe trends?
2. What does a flat trend indicate?
3. What questions should we ask about gradually increasing trends?
4. What does an abrupt change in trend suggest?
5. How do we interpret cyclical trends?
6. What do intermittent outliers tell us?
7. What does an erratic trend mean?
8. What is a cumulative run chart?
9. When should we use cumulative vs regular run charts?
10. What is a glide path on a run chart?
11. What color should the goal line be on run charts?
12. How do we use callouts on run charts?

13. What scale should we use for run charts?
14. How do run charts show systemic vs special causes?
15. What are best practices for building run charts?

Consensus & Convergence

1. What is consensus in problem solving?
2. Does consensus mean complete agreement?
3. How do we reach consensus in PSP?
4. What is buy-in versus agreement?
5. When does consensus occur in diverge-converge cycles?
6. How do we test for consensus?
7. What is a "thumbs up" test?
8. What if there's a stalemate between ideas?
9. Is consensus a permanent commitment?
10. How do we handle unaddressed concerns in consensus?

INVESTIGATE Stage Testing

Investigation Fundamentals

1. What is the second stage of DIVE?
2. What is the main goal of the INVESTIGATE stage?
3. What is the Causal Chain?
4. What are the 3W's in finding causes?
5. What is the Point of Occurrence?
6. What is the Point of Cause?
7. How do we walk backwards through the causal chain?
8. What is a Pareto Chart?
9. What is the 80/20 principle?
10. What is a Level 1 (L1) Pareto Chart?
11. What is a Level 2 (L2) Pareto Chart?
12. What's the difference between WHERE and WHY Paretos?
13. How many Pareto Charts should we create?

14. What is the Pareto Challenge?
15. How do we overcome hurdles in building Pareto Charts?
16. What are stacked Pareto Charts?
17. How do we break down problems by WHERE?
18. How do we break down problems by WHY?
19. What is BOB (Best of Best)?
20. What is WOW (Worst of Worst)?
21. What are "actual causes"?
22. How many actual causes should we identify?
23. What percentage of the gap should actual causes account for?
24. What deliverables come from INVESTIGATE stage?
25. What are common challenges in INVESTIGATE?

Test Case 2.1: Level 1 Pareto Generation (WHERE)

- Test input data: Top 2 (Newark/TE): \$10,205k gap (40%), Broad Business: \$7,654k gap (30%), Federal Sales: \$5,103k gap (20%), New Partners: \$2,551k gap (10%)
- Expected AI functionality: Generate visual Pareto chart, 80/20 principle validation, guide user to focus on top contributors

Test Case 2.2: Level 2 Pareto Generation (WHY)

- Test input data: Root causes across channels - Competition/Discount Levels: 54% of gap, Partner Led Demand Gen: 40% additional, Unique GTM Experiment: Contributing factor, Direct AM Engagement: Remaining gap
- Expected output: Stacked Pareto showing WHY breakdown by WHERE categories

Test Case 2.3: PSP Tree Creation

- Test scenario: Upload/describe the AMR channel problem
- Expected AI behavior: Guide user to separate symptoms (Revenue miss, booking gaps, OP delivery issues), problems (trunk: \$25M POS gap in AMR channel), and causes (roots: Pricing competitiveness, partner demand gen, GTM alignment)

Test Case 2.4: 5-Whys Root Cause Analysis

- Test starting point: "Competition and discount levels impacting sales"
- Expected AI guidance through 5-Whys process to reach root cause: "For specific product families, soft market pushes business toward heightened price sensitivity. We did not move aggressively to win those customers."

Test Case 2.5: VOC Integration Testing

- Test input: Gemba evidence from PSR - "RS said our margin blend makes selling our brand hard to justify. Newark used Keithley as example for YoY margin compression. Most partners say we are 2-3% below standard discount levels."
- Expected AI processing: Incorporate VOC into root cause analysis, link evidence to specific causes, guide follow-up questions for deeper investigation

Voice of Customer (VOC)

1. What is Voice of Customer (VOC)?
2. Why is VOC critical in problem solving?
3. Who should we talk to for VOC?
4. What makes a good VOC target?
5. How many customers should we talk to?
6. What is the "rule of seven" in VOC?
7. What are customer personas?
8. What are customer process maps?
9. What are "jobs-to-be-done"?
10. What are the three categories of jobs-to-be-done?
11. What are emotional jobs-to-be-done?
12. What are functional jobs-to-be-done?
13. What are social jobs-to-be-done?
14. How do we capture VOC insights?
15. What questions should we ask customers?
16. How much time should we spend listening vs talking?
17. What are best practices for questioning?
18. What are best practices for observing?
19. How do we synthesize VOC findings?
20. When should we synthesize VOC data?
21. What is affinity diagramming for VOC?
22. How do we prioritize VOC insights?
23. Why should we record customer interviews?
24. What is the minimum team size for interviews?

25. How do we ask open-ended questions?

Fishbone Diagram

1. What is a Fishbone Diagram?
2. What is another name for the Fishbone Diagram?
3. What does the Fishbone Diagram visualize?
4. Where are causes located on the Fishbone?
5. Where are effects located on the Fishbone?
6. What are the splines in a Fishbone Diagram?
7. How do we develop a Fishbone Diagram?
8. What are affinities in the Fishbone?
9. How do we identify possible causes using Fishbone?
10. How do we prioritize causes from the Fishbone?

5 Whys & Root Cause

1. What is a Root Cause?
2. What is the 5 Whys technique?
3. Why is it called "5 Whys"?
4. How many whys should we actually ask?
5. What happens if we ask too few whys?
6. What happens if we ask too many whys?
7. How do we know when to stop asking why?
8. What is the step-by-step process for 5 Whys?
9. Can we ask questions other than "why"?
10. How do we test our 5 Whys logic?
11. What is the "walk-back" method?
12. What does "therefore" testing mean?
13. How do we avoid blame in 5 Whys?
14. Why use Gemba evidence in 5 Whys?
15. How do we keep 5 Whys within our control?
16. What if 5 Whys goes outside our Circle of Control?
17. What are examples of good 5 Whys analysis?

18. What makes a strong root cause statement?

19. Can there be multiple root causes?

20. How do we document 5 Whys analysis?

VERIFY/IMPLEMENT Stage Testing

Countermeasure Development

1. What is the third stage of DIVE?
2. What are countermeasures?
3. What's the difference between temporary and permanent countermeasures?
4. When do we use temporary countermeasures?
5. When do we use permanent countermeasures?
6. What makes an effective countermeasure?
7. What are weak countermeasures?
8. What are strong countermeasures?
9. What is Countermeasure Storming?
10. How many countermeasures should each team member generate?
11. How long should Countermeasure Storming take?
12. What is an Impact Matrix?
13. How do we prioritize countermeasures?
14. What does "Immediate" mean in the Impact Matrix?
15. What are assumptions in countermeasures?
16. What are uncertainties in countermeasures?
17. How do we identify assumptions?
18. What are IF...THEN statements?
19. How do we assess uncertainty levels?
20. What is experimentation in PSP?
21. Why do we experiment with countermeasures?
22. What's the difference between experimentation and Trystorming?
23. What are the steps for effective experimentation?
24. What is a prototype?

25. What are low fidelity prototypes?
26. What are high fidelity prototypes?
27. What is an Experiment Card?
28. What are the three Ps (Pivot, Pursue, Pass)?
29. When do we Pivot?
30. When do we Pursue?
31. When do we Pass?
32. How many experiments should we run?
33. What are WOW scores?
34. What are examples of prototypes?
35. What deliverables come from VERIFY/IMPLEMENT?

Test Case 3.1: Countermeasure Generation

- Test input: Root cause from 5-Whys analysis
- Expected AI output options: Temporary (Deploy scheduled promotions calendar - 12M impact), Permanent (Focus on effective EDU strategy - 4M impact), Experimental (Innovative promotions hard for competitors to match - 1M impact)

Test Case 3.2: Impact Matrix Creation

- Test scenario: Multiple countermeasures with effort/impact ratings
- Expected visualization: High Impact/Low Effort (Immediate), High Impact/High Effort (Plan), Low Impact/Low Effort (Consider), Low Impact/High Effort (Avoid)

Test Case 3.3: Assumption Testing

- Test input: "Deploy scheduled promotions calendar"
- Expected AI analysis: IF-THEN statements with uncertainty assessment (High/Medium/Low) for each assumption

Test Case 3.4: Experimentation Framework

- Test scenario: High uncertainty assumption needs testing
- Expected AI guidance: Prototype development (low to high fidelity), success criteria definition, experiment card creation, Build-Measure-Learn cycle

ENSURE SUSTAINMENT Stage Testing

Sustainment Planning

1. What is the fourth stage of DIVE?
2. What is the purpose of Ensure Sustainment?
3. What do we monitor during sustainment?
4. How do we know if countermeasures are working?
5. What is a sustainment plan?
6. What questions does a sustainment plan answer?
7. What are the 3 W's and 2 H's of sustainment?
8. How often should we have sustainment meetings?
9. What is covered in sustainment meetings?
10. How long should we monitor sustainment?
11. Who decides when to stop monitoring?
12. What is the 80/20 rule in sustainment?
13. Which countermeasures need most attention?
14. How do we handle countermeasures that aren't working?
15. Can we alter or drop countermeasures?
16. What role does leadership play in sustainment?
17. How do we prevent process reversion?
18. What are sustainment best practices?
19. What deliverables come from ENSURE SUSTAINMENT?
20. Why do processes tend to revert without monitoring?

Test Case 4.1: Sustainment Plan Creation

- Test input: Implemented countermeasures from PSR
- Expected AI output: What (Promotion standard work and performance review), Who (Performance review team/Promo planning team), When (Monthly reviews), How (Standard work and check-ins), How Long (Until gap is closed and sustained)

Test Case 4.2: Progress Tracking

- Test input: Status updates (0%, 20%, 50% completion)
- Expected AI monitoring: Track countermeasure implementation, flag at-risk items, suggest corrective actions

Cross-Functional Testing

File Integration & Session Management

- Test Case CF.1: File Upload Integration - Testing scenarios for Pareto data upload, PSR critique, and PSP Tree image upload
- Test Case CF.2: Session Management - Testing thread creation, conversation continuity, file association, and context preservation
- Test Case CF.3: Azure Integration - Testing authentication, CogSearch integration, and Prompt Flow endpoint

Validation Criteria

Output Quality Metrics

- TAGS Completeness: All 4 elements present and accurate
- Pareto Accuracy: 80/20 principle applied, visual clarity