PEETIC
PLAN DE TEST

2025

PROJET WINSTON

TABLE DES MATIÈRES

Contact

PABLO PERNOT
pablo@projetwinston.fr
06 32 43 78 03

Licence

1. INTRODUCTION

1.1. OBJECTIF DU DOCUMENT

Ce plan de tests présente la stratégie d'expérimentation pour valider les hypothèses du framework Thoughtful Execution appliqué à Peetic. Il détaille les tests prioritaires du Q1 2026, leurs protocoles, métriques et critères de succès.

1.2. MÉTHODOLOGIE

Nous utilisons une approche Lean Startup avec :

- Tests progressifs du plus simple au plus complexe
- Critères de succès définis avant chaque test
- Kill criteria clairs pour éviter l'effet sunk cost
- Documentation systématique des apprentissages

1.3. ALIGNEMENT STRATÉGIQUE

Ce plan de tests s'aligne avec :

- La métrique North Star : Rappels honorés par semaine
- La stratégie Rumelt Phase 1 : Ancrage utilitaire
- Les OKRs Q1 2026 : Créer habitude quotidienne

2. TESTS PRIORITAIRES Q1 2026

2.1. VUE D'ENSEMBLE

Test	Hypothèse testée	Type	Durée	Statut
T1.1.C	Timeline visuelle vs liste	A/B	2 sem	A LANCER
T1.2.B	Timing notification optimal	A/B	3 sem	EN COURS
T1.2.C	Personnalisation rappels	A/B	2 sem	A LANCER
T1.2.D	Wizard vs config manuelle	A/B	2 sem	A LANCER
T1.3.A	Pilote clinique vétérinaire	Bêta	4 sem	EN COURS
T1.3.C	Retention sync vs non-sync	Cohorte	90j	A LANCER
T2.1.B	Placement bouton urgence	A/B	1 sem	A LANCER
T2.1.C	Appel one-tap vs numéro	A/B	1 sem	A LANCER
T2.2.B	Format PDF préféré	Qual	1 sem	EN COURS
T2.2.C	Performance génération PDF	Tech	3j	A LANCER
T3.1.A	Wireframe dashboard	Qual	1 sem	A LANCER
T3.1.B	Priorisation actions	A/B	2 sem	A LANCER
T3.3.A	Wizard onboarding auto	A/B	2 sem	EN COURS

3. OPPORTUNITÉ 1 : GÉRER SANTÉ PRÉVENTIVE

3.1. TEST 1.1.C : VISUALISATION CARNET DE SANTÉ

3.1.1. HYPOTHÈSE

La visualisation en timeline visuelle réduit le temps de recherche d'une information de 40% par rapport à une liste chronologique classique.

3.1.2. SETUP EXPÉRIMENTAL

Participants : 60 utilisateurs bêta (30 par groupe)

Groupes :
- Groupe A : Liste chronologique classique (contrôle)
- Groupe B : Timeline visuelle avec icônes et codes couleur

Tâche de test :
1. Retrouver le dernier vaccin antirabique
2. Vérifier si le vermifuge est à jour
3. Consulter le poids de l'animal il y a 6 mois

Métriques :
- Temps moyen de complétion de chaque tâche (primaire)
- Taux de réussite (secondaire)
- Score satisfaction UX 1-10 (secondaire)

3.1.3. CRITÈRES DE SUCCÈS

Validation : Groupe B réduit temps recherche >= 35% ET score satisfaction >= 8/10

Itération : Groupe B réduit temps 20-34% OU satisfaction 6-7/10

Abandon : Groupe B réduit temps < 20% OU satisfaction < 6/10

3.1.4. PROTOCOLE

Semaine 1 :
- Recrutement 60 bêta-testeurs (critère : >= 3 documents uploadés)
- Assignation aléatoire aux groupes
- Test utilisateurs en remote (enregistrement écran + voice-over)

Semaine 2 :
- Analyse quantitative (temps, taux réussite)
- Analyse qualitative (verbatims, points de friction)
- Décision Go/NoGo/Iterate

3.1.5. RISQUES ET MITIGATIONS

Risque 1 : Effet nouveauté fausse les résultats
- Mitigation : Période acclimatation 48h avant mesure

Risque 2 : Taille échantillon insuffisanté
- Mitigation : Power analysis préliminaire (60 participants = 80% power)

3.2. TEST 1.2.B : TIMING NOTIFICATION RAPPELS

3.2.1. HYPOTHÈSE

Les notifications envoyées 15 jours avant échéance optimisent le taux d'honneur des rappels (ni trop tôt = oubli, ni trop tard = stress).

3.2.2. SETUP EXPÉRIMENTAL

Participants : 300 utilisateurs actifs avec >= 1 rappel configuré

Groupes :
- Groupe A : Notification 7 jours avant (n=100)
- Groupe B : Notification 15 jours avant (n=100)
- Groupe C : Notification 21 jours avant (n=100)

Métriques :
- Taux honneur rappel (action effectuée dans 7j suivant échéance) - primaire
- Taux ouverture notification - secondaire
- Score anxiété auto-déclaré post-rappel - secondaire

3.2.3. CRITÈRES DE SUCCÈS

Validation : Groupe gagnant >= 75% taux honneur ET significativement supérieur aux autres ($p < 0.05$)

Itération : Taux honneur 65-74% OU pas de différence significative

Abandon : Tous groupes < 65% taux honneur

3.2.4. PROTOCOLE

Semaines 1-3 :
- Envoi notifications selon timing groupe
- Tracking automatique des actions (marquage « fait » dans l'app)
- Survey post-action (J+1 après échéance)

Semaine 4 :
- Analyse statistique (chi-square test)
- Segmentation par type rappel (vaccin vs vermifuge vs autre)
- Recommandation timing optimal

3.2.5. RÉSULTATS INTERMÉDIAIRES (EN COURS)

Données partielles au 15/11/2025 (n=180 complétés) :

- Groupe A (7j) : 62% taux honneur
- Groupe B (15j) : 78% taux honneur
- Groupe C (21j) : 71% taux honneur

Tendance : Groupe B en tête mais pas encore significatif ($p=0.08$). Besoin 120 utilisateurs supplémentaires.

3.3. TEST 1.2.C : PERSONNALISATION RAPPELS

3.3.1. HYPOTHESE

Les rappels personnalisés par race et age augmentent le taux d'ouverture de 25% et le taux d'honneur de 15% vs rappels génériques.

3.3.2. SETUP EXPÉRIMENTAL

Participants : 200 utilisateurs actifs

Groupes :
- Groupe A : « Rappel : Vaccin antirabique de [Nom animal] » (generique)
- Groupe B : « Balto (Epagneul 4 ans) : Vaccin antirabique recommandé » (personnalisé)

Variables testées :
- Mention nom + race + age
- Ton personnalisé (« recommandé » vs « a faire »)
- Icone spécificité race dans notification

Metriques :
- Taux ouverture notification (primaire)
- Taux honneur rappel (primaire)
- Score pertinence perçue 1-10 (secondaire)

3.3.3. CRITERES DE SUCCÈS

Validation : Groupe B >= +20% ouverture ET >= +10% honneur vs Groupe A

Iteration : Entre +10-19% ouverture OU +5-9% honneur

Abandon : < +10% ouverture ET < +5% honneur

3.3.4. PROTOCOLE

Semaines 1-2 :
- Test A/B classique avec assignation aléatoire
- Tracking automatique ouverture + honneur
- Survey satisfaction J+7

3.3.5. LEARNINGS ATTENDUS

- Quel niveau de personnalisation optimal (nom seul vs nom+race vs nom+race+age)
- Impact de l'icone race sur perception
- Segments utilisateurs les plus réceptifs

3.4. TEST 1.2.D : WIZARD CONFIGURATION RAPPELS

3.4.1. HYPOTHESE

Un wizard de configuration guide augmente le taux de completion de 40% et le nombre moyen de rappels configurés de 2x vs configuration manuelle.

3.4.2. SETUP EXPÉRIMENTAL

Participants : 100 nouveaux utilisateurs en onboarding

Groupes :
- Groupe A : Configuration manuelle (page vierge + bouton « Ajouter rappel »)
- Groupe B : Wizard guide en 4 étapes avec suggestions automatiques

Etapes wizard Groupe B :
1. Race et age de l'animal (ou « Ne sait pas »)
2. Suggestions automatiques de rappels selon race/age
3. Validation en bloc ou ajustement individuel
4. Configuration préférences notifications

Metriques :
- Taux completion configuration (primaire)
- Nombre moyen rappels configurés (primaire)
- Temps passé dans configuration (secondaire)
- Taux ajustement postérieur (secondaire)

3.4.3. CRITERES DE SUCCÈS

Validation : Groupe B >= +35% completion ET >= 1.8x rappels configurés vs A

Iteration : +20-34% completion OU 1.4-1.7x rappels

Abandon : < +20% completion ET < 1.4x rappels

3.4.4. PROTOCOLE

Semaines 1-2 :
- Integration wizard dans flux onboarding
- A/B test avec assignation aléatoire nouveaux inscrits
- Tracking automatique + heatmaps Hotjar
- Interviews qualitatives 10 utilisateurs Groupe B

3.4.5. KILL CRITERIA SPÉCIFIQUES

Abandon immédiat si :
- Taux abandon wizard > 50%
- Temps configuration wizard > 5 min (vs 2 min manuel)
- NPS wizard < 5/10

3.5. TEST 1.3.A : PILOTE CLINIQUE VÉTÉRINAIRE

3.5.1. HYPOTHESE

La synchronisation automatique avec vétérinaire augmente le taux d'adoption Peetic a >= 50% des patients de la clinique et améliore la retention a 85% J+30.

3.5.2. SETUP EXPÉRIMENTAL

Partenaire : Clinique Dr Lemoine (Lyon, 800 patients actifs)

Type : Beta test avec integration manuelle puis API

Phase 1 (Semaines 1-2) : Integration manuelle
- Export CSV comptes-rendus consultation
- Import manuel dans Peetic
- Creation automatique rappels

Phase 2 (Semaines 3-4) : API beta
- Connexion API Peetic - VetoGestion
- Sync temps reel après chaque consultation
- Notification patient «Nouveau CR disponible»

Metriques :
- Taux adoption (% patients clinique avec compte Peetic actif) - primaire
- Taux activation sync (% comptes avec >= 1 CR synchronisé) - primaire
- Retention J+30 sync vs non-sync - primaire
- Satisfaction vétérinaires (NPS) - secondaire
- Nombre questions support utilisateurs - secondaire

3.5.3. CRITERES DE SUCCÈS

Validation Phase 1 : >= 40% adoption patients ET NPS veto >= 7/10

Validation Phase 2 : >= 60% adoption ET retention sync >= 80% J+30

Scale decision : Valide Phase 2 + CAC via veto < 15 euros = deploy 3 cliniques supplémentaires Q1

3.5.4. PROTOCOLE

Semaine 1 :
- Formation équipe Dr Lemoine (30 min)
- Affichage QR code salle attente + flyers
- Mention Peetic par secretaire a chaque prise RDV

Semaines 2-4 :
- Tracking adoption hebdomadaire
- Interviews 10 patients adoptants
- Interviews 10 patients non-adoptants (comprendre barrières)
- Survey satisfaction Dr Lemoine et assistantes

Semaine 5 :
- Analyse ROI (temps gagne vs temps integration)
- Apprentissages pour scale
- Go/NoGo extension autres cliniques

3.5.5. RESULTATS ACTUELS (EN COURS)

Données au 15/11/2025 (Semaine 3) :

- 34 patients avec compte Peetic actif (68% taux adoption sur 50 contactés)
- 28 comptes avec >= 1 CR synchronisé (82% activation)
- Retention J+7 : 94% (vs 68% baseline non-sync)
- NPS Dr Lemoine : 9/10 (« game changer pour suivi preventif »)

Barrieres identifiées :

- Patients séniors difficulté création compte (besoin aide famille)
- Manque notification push lors nouvelle sync (implementation S4)

3.6. TEST 1.3.C : RETENTION SYNC VS NON-SYNC

3.6.1. HYPOTHESE

Les utilisateurs avec synchronisation vétérinaire active ont une retention J+30 de 85% (vs 68% baseline) et J+90 de 75% (vs 52% baseline).

3.6.2. SETUP EXPÉRIMENTAL

Type : Etude de cohorte

Cohortes :
- Cohorte A : Utilisateurs avec sync vétérinaire active (n=100 cible)
- Cohorte B : Utilisateurs sans sync (controle, n=200)

Criteres inclusion :
- Inscription entre Dec 2025 et Jan 2026
- >= 1 document uploade
- >= 1 rappel configure

Metriques :
- Retention J+30 (connexion dans les 30 derniers jours) - primaire
- Retention J+90 - primaire
- Nombre sessions/mois - secondaire
- Taux conversion Premium - secondaire

3.6.3. CRITERES DE SUCCÈS

Validation : Cohorte A >= 80% retention J+30 ET >= +15 points vs Cohorte B

Iteration : 75-79% J+30 OU +10-14 points vs B

Abandon : < 75% J+30 OU < +10 points vs B

3.6.4. PROTOCOLE

Mois 1 (Dec 2025) :
- Constitution cohortes
- Tracking automatique métriques

Mois 2 (Jan 2026) :
- Analyse J+30
- Segmentation par profil utilisateur

Mois 3-4 (Fev-Mars 2026) :
- Suivi J+90
- Analyse survival curves
- Identification moments critiques churn

3.6.5. ANALYSES COMPLEMENTAIRES

- Correlation sync + autres features (rappels, urgence, premium)
- Impact qualité données sync sur retention
- Cout acquisition utilisateur via veto vs autres canaux

4. OPPORTUNITÉ 2 : URGENCE VÉTÉRINAIRE

4.1. TEST 2.1.B : PLACEMENT BOUTON URGENCE

4.1.1. HYPOTHESE

Un bouton urgence floating (omniprésent) augmente le taux de découverte de 3x vs placement dans menu.

4.1.2. SETUP EXPÉRIMENTAL

Participants : 150 utilisateurs actifs

Groupes :
- Groupe A : Bouton dans menu hamburger (controle)
- Groupe B : Bouton dans dashboard principal
- Groupe C : Floating action button rouge (toutes pages)

Metriques :
- Taux découverte feature (% utilisateurs cliquant >= 1x) - primaire
- Temps découverte moyen - primaire
- Taux utilisation en situation réelle (tracking via survey) - secondaire

4.1.3. CRITERES DE SUCCÈS

Validation : Groupe gagnant >= 70% découverte en 7j ET >= 2x vs controle

Iteration : 50-69% découverte OU 1.5-1.9x vs controle

Abandon : < 50% découverte ET < 1.5x vs controle

4.1.4. PROTOCOLE

Semaine 1 :
- A/B/C test avec assignation aléatoire
- Tracking clics + heatmaps
- Pas de communication proactive sur feature (test organique)

Semaine 2 :
- Analyse découverte par groupe
- Survey utilisateurs Groupe C : « Avez-vous remarqué le bouton urgence ? »
- Decision placement optimal

4.1.5. CONSIDERATIONS DESIGN

Groupe C floating button :
- Position : Bas droite
- Couleur : Rouge Winston (#F65e5e)
- Icone : Croix médicale + « SOS »
- Comportement : Visible toutes pages sauf onboarding

4.2. TEST 2.1.C : APPEL DIRECT ONE-TAP

4.2.1. HYPOTHESE

Un bouton d'appel direct augmente le taux de conversion «voir numero vétérinaire» vers «appeler» de 60% vs affichage numero seul.

4.2.2. SETUP EXPÉRIMENTAL

Participants : 100 utilisateurs ayant clique bouton urgence

Groupes :
- Groupe A : Numero telephone affiche (besoin copier-coller ou mémoriser)
- Groupe B : Bouton «Appeler [Nom vétérinaire]» (déclénchement appel direct)

Metriques :
- Taux conversion clic urgence vers appel effectué (primaire)
- Temps entre clic urgence et debut appel (secondaire)
- Satisfaction utilisateur post-urgence (secondaire)

4.2.3. CRITERES DE SUCCÈS

Validation : Groupe B >= +50% conversion ET temps réduit >= 40% vs A

Iteration : +30-49% conversion OU temps réduit 20-39%

Abandon : < +30% conversion ET < 20% temps réduit

4.2.4. PROTOCOLE

Semaine 1 :
- A/B test avec tracking appels (iOS CallKit / Android TelecomManager)
- Survey automatique J+1 après utilisation urgence

Considerations techniques :
- Permissions requises : Acces telephonie
- Fallback Android ancien (< v6) : Copy numero + toast «Collez dans app telephone»

4.3. TEST 2.2.B : FORMAT PDF PRÉFÈRE

4.3.1. HYPOTHESE

Les vétérinaires préfèrent un format PDF chronologique inverse (plus récent en haut) avec sections bien délimitées.

4.3.2. SETUP EXPÉRIMENTAL

Participants : 5 vétérinaires partenaires

Type : Test qualitatif avec 3 formats

Formats testes :
- Format A : Chronologique classique (ancien vers récent)
- Format B : Chronologique inverse (récent vers ancien)
- Format C : Par categories (Vaccins / Vermifuges / Consultations / Autres)

Metriques :
- Preference déclarée (classement 1-2-3) - primaire
- Temps localisation info clé (« Dernier vaccin antirabique ? ») - primaire
- Score lisibilité 1-10 - secondaire

4.3.3. CRITERES DE SUCCÈS

Validation : Format gagnant préfère par >= 4/5 vétérinaires ET temps localisation < 15 sec

Iteration : Format préfère 3/5 OU temps 15-30 sec

Abandon : Aucun format préfère majoritairement ET temps > 30 sec

4.3.4. PROTOCOLE

Semaine 1 :
- Preparation 3 PDFs fictifs (même données, formats différents)
- Sessions 30 min individuelles avec chaque vétérinaire
- Tache : « Trouvez rapidement ces 5 infos dans chaque format »
- Interview semi-structurée préférence

Learnings attendus :
- Structure optimale PDF
- Niveau détail vs synthèse
- Utilite QR code vérification
- Informations manquantes actuellement

4.3.5. RESULTATS ACTUELS (EN COURS)

Données partielles 3/5 vétérinaires :

- Format B (chronologique inverse) : 3 votes préférence
- Temps localisation moyen : Format B = 12 sec, Format A = 24 sec, Format C = 18 sec
- Feedbacks : « Format B plus intuitif, cherche toujours info récente en priorité »

Ameliorations suggérées :
- Ajouter photo animal en header

- Highlight allergies en rouge
- Tableau synthèse vaccins (nom + dernière date + prochaine échéance)

4.4. TEST 2.2.C : PERFORMANCE GÉNÉRATION PDF

4.4.1. HYPOTHESE

La génération PDF avec optimisations backend peut atteindre < 5 sec au p95 même avec 50+ documents.

4.4.2. SETUP EXPÉRIMENTAL

Type : Test technique load testing

Scenarios :
- Scenario 1 : 5 documents (p50 utilisateurs)
- Scenario 2 : 15 documents (p75)
- Scenario 3 : 50 documents (p95)
- Scenario 4 : 100 documents (p99)

Metriques :
- Temps génération PDF p50, p95, p99 (primaire)
- Taux erreur génération (primaire)
- Consommation mémoire serveur (secondaire)

4.4.3. CRITERES DE SUCCÈS

Validation : p95 < 5 sec ET p99 < 8 sec ET taux erreur < 1%

Iteration : p95 5-7 sec OU p99 8-12 sec

Abandon : p95 > 7 sec OU p99 > 12 sec

4.4.4. PROTOCOLE

Phase 1 (Jour 1) :
- Implementation cache Redis pour templates
- Optimisation requêtes DB (eager loading)
- Parallelisation traitement images

Phase 2 (Jour 2) :
- Load testing avec Apache JMeter
- 100 générations concurrentes par scenario
- Monitoring APM (NewRelic)

Phase 3 (Jour 3) :
- Analyse bottlenecks
- Optimisations ciblées
- Nouveau test validation

4.4.5. OPTIMISATIONS CANDIDATES

- Lazy loading images (charge uniquement si nécessaire)
- Pre-compilation template PDF au lieu compilation runtime
- CDN pour assets statiques PDF (logo, icônes)
- Queue asynchrone si génération > 3 sec

5. OPPORTUNITÉ 3 : RÉDUIRE CHARGE MENTALE

5.1. TEST 3.1.A : WIREFRAME DASHBOARD ACTIONS

5.1.1. HYPOTHESE

Un dashboard structuré «Urgent / Cette semaine / Ce mois» est compris immédiatement (< 10 sec) par >= 80% utilisateurs.

5.1.2. SETUP EXPÉRIMENTAL

Participants : 20 utilisateurs (10 actuels + 10 nouveaux)

Type : Test qualitatif wireframe

Wireframes testes :
- Version A : Structure temporelle (Urgent / Semaine / Mois)
- Version B : Structure catégorielle (Sante / Rendez-vous / Documents)
- Version C : Structure mixte (Urgent en haut + Categoriel en bas)

Tache :
- «Ou verriez-vous un rappel de vaccin prévu dans 5 jours ?»
- «Ou verriez-vous un rendez-vous vétérinaire aujourd'hui ?»
- «Quelle action feriez-vous en premier ?»

Metriques :
- Temps compréhension structuré (primaire)
- Taux réponse correcte localisation info (primaire)
- Preference déclarée (secondaire)

5.1.3. CRITERES DE SUCCÈS

Validation : Version gagnante >= 80% compréhension < 10 sec ET >= 85% localisation correcte

Iteration : 70-79% compréhension OU 70-84% localisation

Abandon : < 70% compréhension ET < 70% localisation

5.1.4. PROTOCOLE

Semaine 1 :
- Sessions 30 min remote (Zoom + Figma)
- Think-aloud protocol
- 3 wireframes présentés ordre aléatoire
- Interview post-test préférences

5.1.5. LEARNINGS ATTENDUS

- Structure mentale utilisateurs pour organiser tâches
- Vocabulaire optimal («Urgent» vs «Aujourd'hui» vs «A faire»)
- Densite information optimale par section
- Besoin filtres ou tout visible

5.2. TEST 3.1.B : PRIORISATION ACTIONS DASHBOARD

5.2.1. HYPOTHESE

Une priorisation automatique « smart » réduit le score anxiété de 30% vs affichage chronologique simple.

5.2.2. SETUP EXPÉRIMENTAL

Participants : 120 utilisateurs actifs avec >= 3 actions en cours

Groupes :
- Groupe A : Ordre chronologique simple (controle)
- Groupe B : Priorisation par importance (vaccins > vermifuges > autres)
- Groupe C : Priorisation smart (ML basique : urgence + importance + historique utilisateur)

Metriques :
- Score anxiété auto-déclaré 1-10 avant/après 2 semaines (primaire)
- Taux completion actions suggérées (secondaire)
- Temps passé sur dashboard (secondaire)

5.2.3. CRITERES DE SUCCÈS

Validation : Groupe gagnant reduction anxiété >= 25% ET significativement supérieur ($p < 0.05$)

Iteration : Reduction 15-24% OU tendance positive non significative

Abandon : Reduction < 15% ET pas difference vs controle

5.2.4. PROTOCOLE

Jour 0 :
- Survey baseline anxiété : « Sur échelle 1-10, quel est votre niveau anxiété concernant santé animal ? »
- Assignation aléatoire aux groupes

Semaines 1-2 :
- Utilisation normale app avec priorisation selon groupe
- Tracking completion actions

Jour 14 :
- Survey post-test anxiété (même question)
- Calcul delta anxiété par utilisateur
- Analyse statistique (ANOVA)

5.2.5. ALGORITHME PRIORISATION GROUPE C

Facteurs scores :
- Urgence temporelle (poids 40%) : échéance < 3j = urgent
- Importance type (poids 30%) : vaccin > vermifuge > toilettage
- Historique honneur (poids 20%) : utilisateur honore rappels = moins urgent
- Preferences utilisateur (poids 10%) : actions ignorées = baisse priorité

5.3. TEST 3.3.A : WIZARD ONBOARDING AUTO

5.3.1. HYPOTHESE

Un wizard de configuration automatique augmente le taux de completion de 92% (vs 54% config manuelle) et réduit anxiété immédiate de 2 points.

5.3.2. SETUP EXPÉRIMENTAL

Participants : 100 nouveaux utilisateurs

Groupes :
- Groupe A : Configuration manuelle (controle)
- Groupe B : Wizard avec suggestions automatiques («pilote auto»)

Etapes wizard Groupe B :
1. «Parlez-nous de [Nom animal]» (race, age, mode vie)
2. «Voici les rappels que nous recommandons» (liste pré-cochée)
3. «Validez ou ajustez» (possibilité modifier)
4. «C'est pret!» (récapitulatif + badge «Protege»)

Metriques :
- Taux completion configuration (primaire)
- Nombre moyen rappels configurés (primaire)
- Score anxiété immédiat après config (primaire)
- Taux ajustement suggestions (secondaire)

5.3.3. CRITERES DE SUCCÈS

Validation : Groupe B >= 90% completion ET >= 4 rappels moyenne ET anxiété −2 points vs A

Iteration : 80-89% completion OU 3-3.9 rappels OU anxiété −1 point

Abandon : < 80% completion ET < 3 rappels ET pas reduction anxiété

5.3.4. PROTOCOLE

Semaines 1-2 :
- A/B test sur nouveaux inscrits
- Tracking completion + temps passé
- Survey post-config : «Score anxiété 1-10» + «Score satisfaction wizard 1-10»

5.3.5. RESULTATS ACTUELS (EN COURS)

Données partielles au 15/11/2025 (n=62) :

- Groupe A : 54% completion, 2.1 rappels moyenne, anxiété 6.8/10
- Groupe B : 92% completion, 4.3 rappels moyenne, anxiété 4.9/10

Delta anxiété Groupe B : −1.9 points (proche objectif −2)

Insights qualitatifs Groupe B :
- «Rassurant de voir que l'app sait ce dont j'ai besoin»
- «J'aurais oublie le traitement antiparasitaire»
- «Gagner 10 min vs tout configurer manuellement»

Taux ajustement suggestions : 28% modifient >= 1 rappel (acceptable)

6. CALENDRIER EXPÉRIMENTAL Q1 2026

6.1. JANVIER 2026

Semaine	Tests lancés	Tests complétés
S1	▪ T1.2.B (suite EN COURS) ▪ T3.3.A (suite EN COURS)	
S2	▪ T1.3.A (suite EN COURS) ▪ T2.2.B (suite EN COURS)	▪ T3.3.A (résultats)
S3	▪ T1.1.C (lancement) ▪ T2.1.B (lancement)	▪ T1.2.B (résultats) ▪ T2.2.B (résultats)
S4	▪ T1.2.C (lancement) ▪ T3.1.A (lancement)	▪ T1.1.C (résultats) ▪ T2.1.B (résultats)

6.2. FEVRIER 2026

Semaine	Tests lancés	Tests complétés
S1	▪ T1.2.D (lancement) ▪ T2.1.C (lancement)	▪ T1.2.C (résultats) ▪ T3.1.A (résultats)
S2	▪ T3.1.B (lancement) ▪ T2.2.C (lancement)	▪ T1.2.D (résultats)
S3	▪ T1.3.C (lancement)	▪ T2.1.C (résultats) ▪ T2.2.C (résultats)
S4	▪ Review mid-Q1 ▪ Ajustements roadmap	▪ T3.1.B (résultats) ▪ T1.3.A (résultats finaux)

6.3. MARS 2026

Semaine	Tests lancés	Tests complétés
S1-S2	▪ Tests secondaires selon learnings ▪ Validations techniques solutions	▪ Compilation learnings Q1
S3	▪ Preparation tests Q2	▪ Decisions Go/NoGo features
S4	▪ Review complète Q1 ▪ Présentation stakeholders	▪ T1.3.C (résultats J+30)

7. GESTION DES APPRENTISSAGES

7.1. TEMPLATE DOCUMENTATION TEST

Pour chaque test complète, documenter :

1. Hypothese testée

[Enonce clair hypothèse]

2. Setup expérimental

[Participants, groupes, métriques]

3. Resultats quantitatifs

[Données chiffrées + significativité statistique]

4. Resultats qualitatifs

[Verbatims utilisateurs, observations, insights]

5. Decision

[VALIDE / INVALIDE / ITERATE + justification]

6. Apprentissages clés

[3-5 learnings actionnables]

7. Recommendations next steps

[Actions concrètes selon decision]

7.2. PARTAGE DES LEARNINGS

Hebdomadaire :
- Update fichier OST (marquer tests VALIDES/INVALIDES)
- Slack channel tests-apprentissages
- Standups équipe produit

Mensuel :
- Présentation all-hands avec highlights
- Documentation Notion centrale
- Update stakeholders externes (investisseurs si applicable)

Trimestriel :
- Retrospective expérimentations Q1
- Rapport learnings complet
- Ajustement stratégie Q2

8. PRINCIPES D'EXPÉRIMENTATION

8.1. VELOCITY VS LEARNING

Privilegier qualité apprentissage sur quantité tests :

- Mieux 5 tests bien conçus avec insights clairs
- Que 20 tests baclés sans apprentissages exploitables

Cadence cible :
- 3-4 tests lancés par semaine
- 2-3 tests complétés par semaine

8.2. RIGUEUR SCIENTIFIQUE

Avant chaque test :
- Hypothese claire et falsifiable
- Metriques définies a priori
- Criteres succès/échec explicites
- Taille échantillon calculée (power analysis)

Pendant test :
- Pas de modification setup en cours
- Tracking automatique prioritaire sur déclaratif
- Monitoring anomalies (bug, biais échantillon)

Apres test :
- Analyse statistique rigoureuse (significativité)
- Separation correlation et causalité
- Documentation complète même si échec

8.3. KILL CRITERIA SYSTÉMATIQUES

Chaque test définit critères abandon AVANT lancement :

- Metrique primaire < seuil minimum
- Cout implementation > budget
- NPS feature < 5/10 (rejet utilisateurs)
- Impact négatif autres métriques clés
- Delai réalisation > 2x estimé

Si kill criteria atteint : STOP immédiat, documenter, pivoter.

8.4. COLLABORATION CROSS-FONCTIONNELLE

Product : Hypothèses, métriques, decisions Go/NoGo

Design : Wireframes, prototypes, tests utilisateurs

Engineering : Faisabilite, implementation rapide, instrumentation

Data : Analytics, significativité statistique, dashboards

Marketing : Messaging, acquisition participants tests

Rituels :

- Planning tests hebdomadaire (Lundi)
- Review résultats hebdomadaire (Vendredi)
- Retrospective tests mensuelle

9. METRIQUES DE SUCCÈS PLAN DE TESTS

9.1. NIVEAU TESTS INDIVIDUELS

Objectif Q1 2026 :
- 15 tests lancés (100% objectif)
- 12 tests complétés (80% objectif)
- >= 8 tests VALIDES (67% taux validation)

Qualite apprentissages :
- 100% tests avec documentation complète
- >= 3 learnings actionnables par test
- Temps moyen decision Go/NoGo < 48h après résultats

9.2. NIVEAU OPPORTUNITES

Opportunité 1 (Gérer santé) :
- >= 4/6 tests valides
- >= 1 solution déployée production Q1
- Contribution >= 60% North Star Metric

Opportunité 2 (Urgence veto) :
- >= 3/5 tests valides
- Feature urgence en beta publique
- NPS feature >= 9/10

Opportunité 3 (Charge mentale) :
- >= 2/4 tests valides
- Reduction anxiété moyenne >= 20%
- Taux adoption dashboard >= 70%

9.3. NIVEAU NORTH STAR METRIC

Contribution plan tests :
- Features validées contribuent cumulativement a 2 500 rappels honores/semaine Q1
- Chaque solution validée impacte >= 10% métrique
- Pas de régression autres métriques clés (retention, NPS)

10. RISQUES ET MITIGATIONS

10.1. RISQUES EXPERIMENTAUX

Risque 1 : Taille échantillon insuffisanté
- Impact : Resultats non significatifs statistiquement
- Mitigation : Power analysis avant chaque test + recrutement anticipé
- Contingence : Extension duree test ou elargissement critères inclusion

Risque 2 : Effet Hawthorne (comportement modifie par observation)
- Impact : Resultats tests ne reflètent pas usage reel
- Mitigation : Tests A/B aveugles + période acclimatation + tracking passif
- Contingence : Validation avec cohorte non-informée

Risque 3 : Biais selection participants
- Impact : Resultats non généralisables
- Mitigation : Assignation aléatoire + diversité personas
- Contingence : Segmentation analyses par profil utilisateur

10.2. RISQUES OPERATIONNELS

Risque 4 : Retards implementation technique
- Impact : Decalage calendrier tests
- Mitigation : Buffer 20% roadmap + priorisation MVP features
- Contingence : Tests papier/prototype avant implementation complète

Risque 5 : Surcharge équipe
- Impact : Baisse qualité tests ou burnout
- Mitigation : Max 4 tests paralleles + rotation roles
- Contingence : Report tests P2 au Q2

10.3. RISQUES BUSINESS

Risque 6 : Tests négatifs consécutifs = perte confiance stakeholders
- Impact : Remise en cause stratégie produit
- Mitigation : Communication proactive valeur apprentissages échecs
- Contingence : Quick wins avec tests a faible risque

Risque 7 : Feature validée mais chère a scaler
- Impact : Impassé entre validation et contraintes techniques/budget
- Mitigation : Estimation coûts scale avant validation finale
- Contingence : Version dégradée ou pivot implementation

11. CONCLUSION

11.1. SYNTHESE DU PLAN

Ce plan de tests Q1 2026 structuré l'expérimentation autour de 13 tests prioritaires couvrant les 3 opportunites principales du framework Thoughtful Execution Peetic.

Principes directeurs :
- Rigueur scientifique avec hypothèses falsifiables
- Apprentissages systématiques même en cas échec
- Collaboration cross-fonctionnelle
- Alignement permanent avec North Star Metric

11.2. PROCHAINES ÉTAPES IMMÉDIATES

Semaine du 18 Novembre 2025 :
- Finalisation protocoles tests T1.1.C et T2.1.B
- Recrutement participants (objectif 60 + 150)
- Preparation instrumentation tracking

Semaine du 25 Novembre 2025 :
- Lancement T1.1.C
- Analyse résultats finaux T1.2.B (EN COURS)
- Decision Go/NoGo timing notification 15j

Decembre 2025 :
- Acceleration cadence tests (objectif 6 complétés)
- Premier bilan learnings
- Ajustements roadmap Q1 selon premiers résultats

11.3. ENGAGEMENT ÉQUIPE

L'expérimentation n'est pas un frein a la vélocité, c'est un accélérateur de certitude.

Chaque test réduit le risque, augmente les chances de succès produit, et renforce la culture data-driven de l'équipe Peetic.

12. ANNEXES

12.1. ANNEXE A : TEMPLATE FICHE TEST

```
FICHE TEST [ID]

Opportunité : [Opp X]
Hypothese : [Enonce]
Type : [A/B / Qualitatif / Technique / Cohorte]
Duree : [X semaines]
Participants : [n=X]

SETUP
- Groupes : [Description]
- Metriques : [Liste]
- Criteres succès : [VALIDE si...]

PLANNING
- S1 : [Actions]
- S2 : [Actions]

RESULTATS
[A complèter post-test]

DECISION
[VALIDE / INVALIDE / ITERATE]

LEARNINGS
1. [Learning 1]
2. [Learning 2]
3. [Learning 3]

NEXT STEPS
[Actions concrètes]
```

12.2. ANNEXE B : CHECKLIST LANCEMENT TEST

Avant lancement :

- [ ] Hypothese validée par PM + Design + Eng
- [ ] Metriques instrumentées et testées
- [ ] Participants recrutés et segmentés
- [ ] Brief équipe support (anticipér questions users)
- [ ] Plan B si bug ou anomalie
- [ ] Documentation test créée (Notion)
- [ ] Slack channel test créé
- [ ] Go final PO

12.3. ANNEXE C : CHECKLIST CLOTURE TEST

Après test complet :

- [ ] Données exportées et sauvegardées
- [ ] Analyse statistique terminée
- [ ] Insights qualitatifs documentés
- [ ] Decision Go/NoGo prise et argumentée
- [ ] Fiche test complétée
- [ ] Présentation équipe réalisée
- [ ] Update OST (statut solutions)

- [ ] Communication stakeholders
- [ ] Archivage prototype/code test

13. RÉFÉRENCES

Méthodologie expérimentation :
- Stefan Thomke, « Experimentation Works » (2020)
- Ronny Kohavi, « Trustworthy Online Controlled Experiments » (2020)
- Dan Siroker, « A/B Testing: The Most Powerful Way to Turn Clicks Into Customers » (2013)

Application au cas Peetic :
- Document Thoughtful Execution Framework Peetic
- Document OST-JTBD Peetic
- Document JTBD Peetic
- OKRs et Product Bets Peetic

Outils recommandés :
- A/B testing : Optimizely, LaunchDarkly, Firebase Remote Config
- Analytics : Mixpanel, Amplitude
- Qual testing : Hotjar, FullStory, Maze
- Stat analysis : R, Python (scipy.stats)