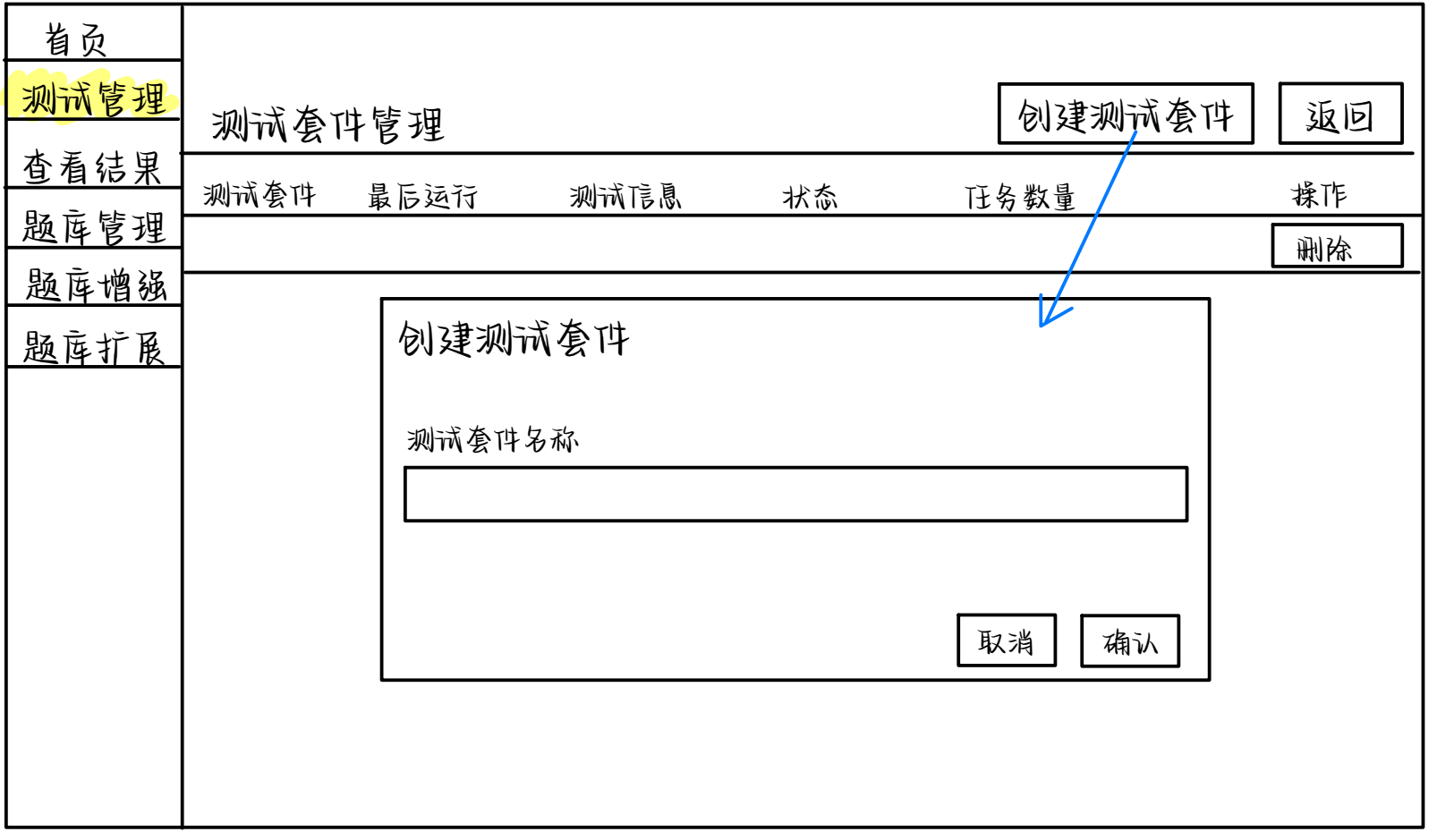
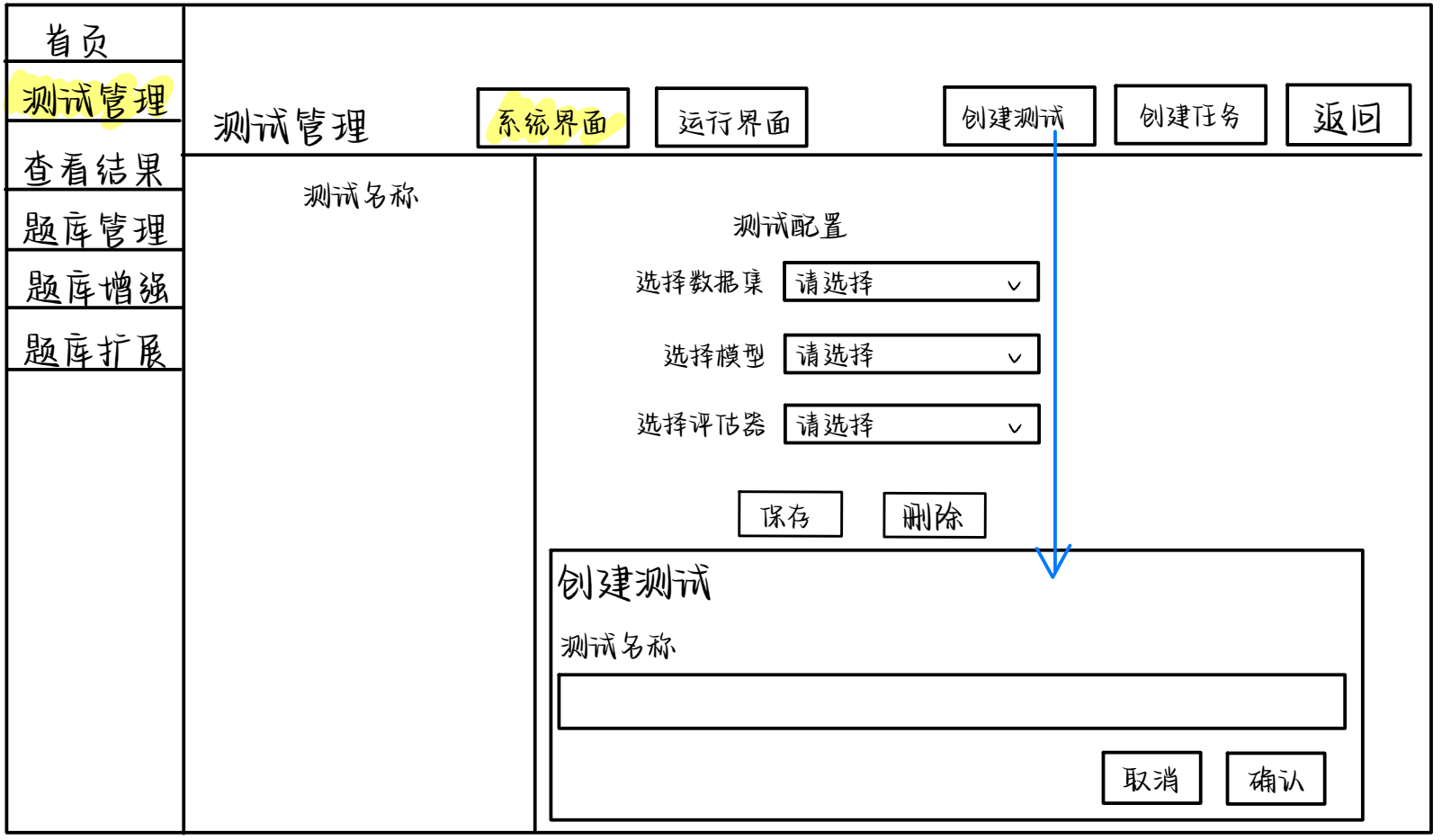
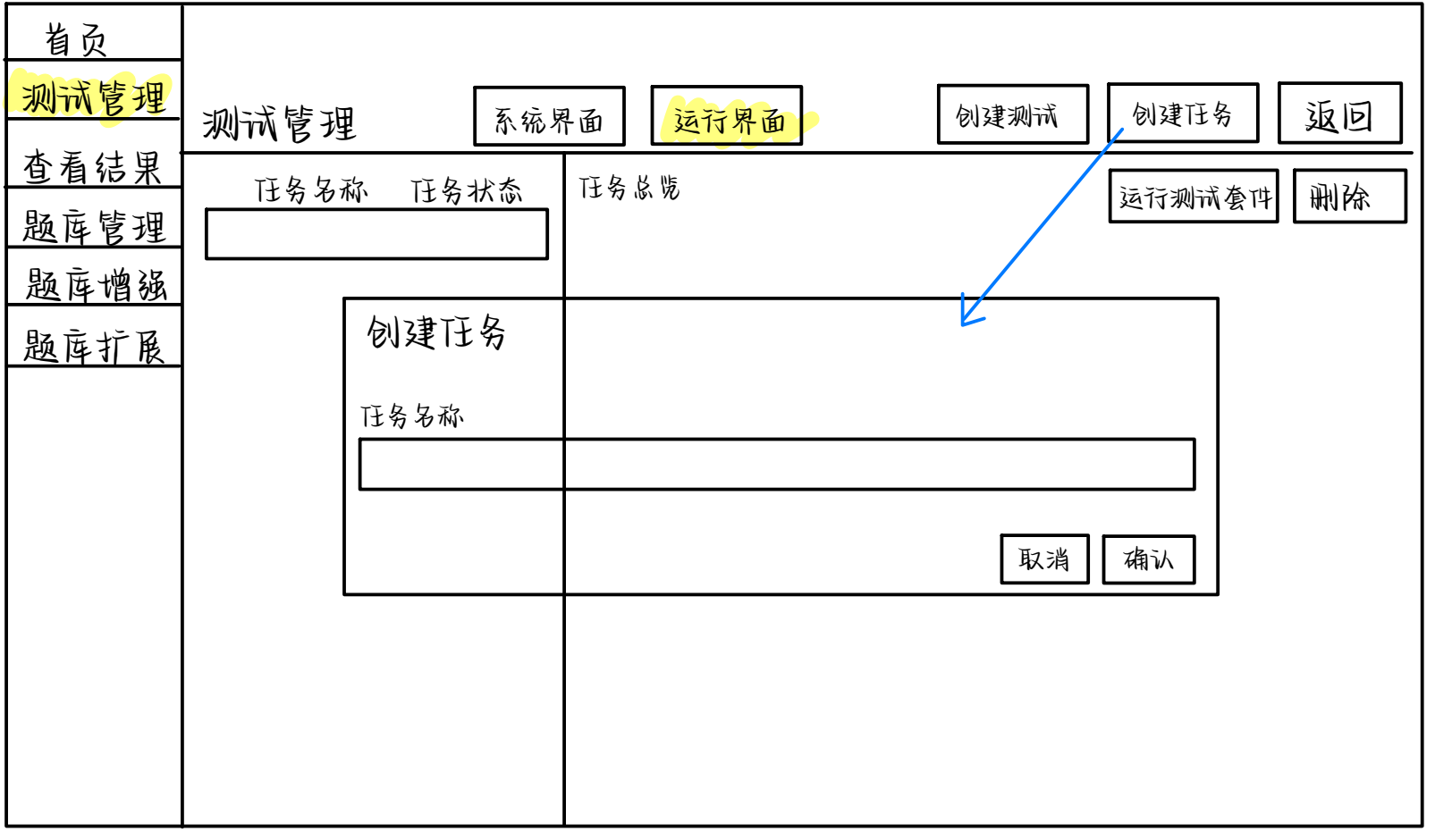
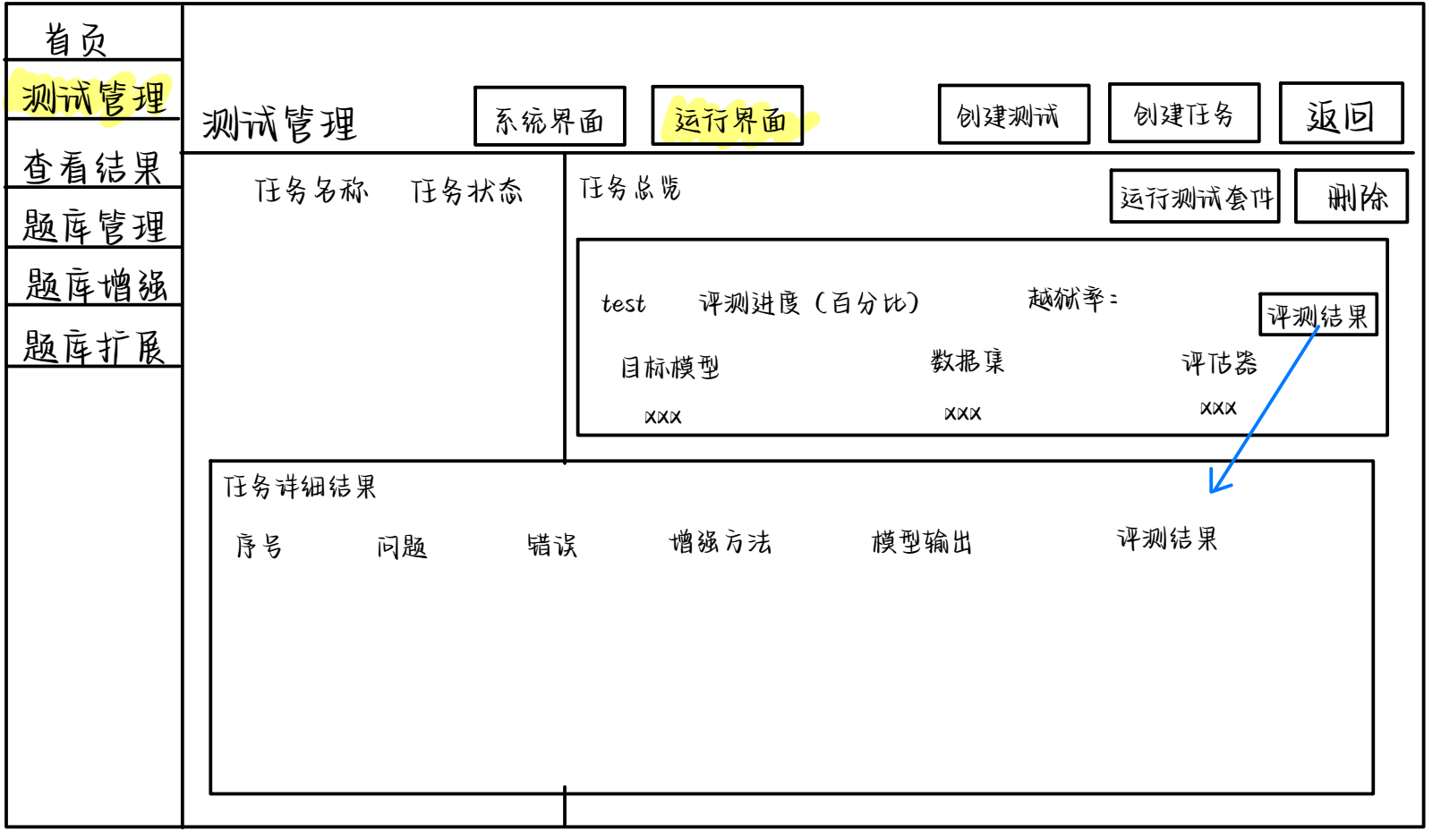

首页
测试管理
查看结果
题库管理
题库增强
题库扩展
测试套件管理
创建测试套件
返回
测试套件
最后运行
测试信息
状态
任务数量
操作
删除
创建测试套件
测试套件名称
取消
确认


首页
测试管理
查看结果
题库管理
题库增强
题库扩展
测试管理
系统界面
运行界面
创建测试
创建任务
返回
测试名称
测试配置
选择数据集
请选择
选择模型
请选择
选择评估器
请选择
保存
删除
创建测试
测试名称
取消
确认


首页
测试管理
查看结果
题库管理
题库增强
题库扩展
测试管理
系统界面
运行界面
创建测试
创建任务
返回
任务名称
任务状态
任务总览
运行测试套件
删除
创建任务
任务名称
取消
确认


首页
测试管理
查看结果
题库管理
题库增强
题库扩展
测试管理
系统界面
运行界面
创建测试
创建任务
返回
任务名称
任务状态
任务总览
运行测试套件
删除
test
评测进度（百分比）
越狱率：
评测结果
目标模型
xxx
数据集
xxx
评估器
xxx
任务详细结果
序号
问题
错误
增强方法
模型输出
评测结果

首页

测试管理

查看结果

题库管理

题库增强

题库扩展

测试情况总览 请选择测试套件 导出越狱成功数据 返回

测试数据：

总越狱率（ASR）：

模型越狱率（ASR）：

领域越狱率（ASR）：

增强方法越狱率（ASR）：

首页

测试管理

查看结果

题库管理

题库增强

题库扩展

测试情况总览

请选择测试套件 [ ] [ ] 导出越狱成功数据 返回

测试数据：

总越狱率（ASR）：

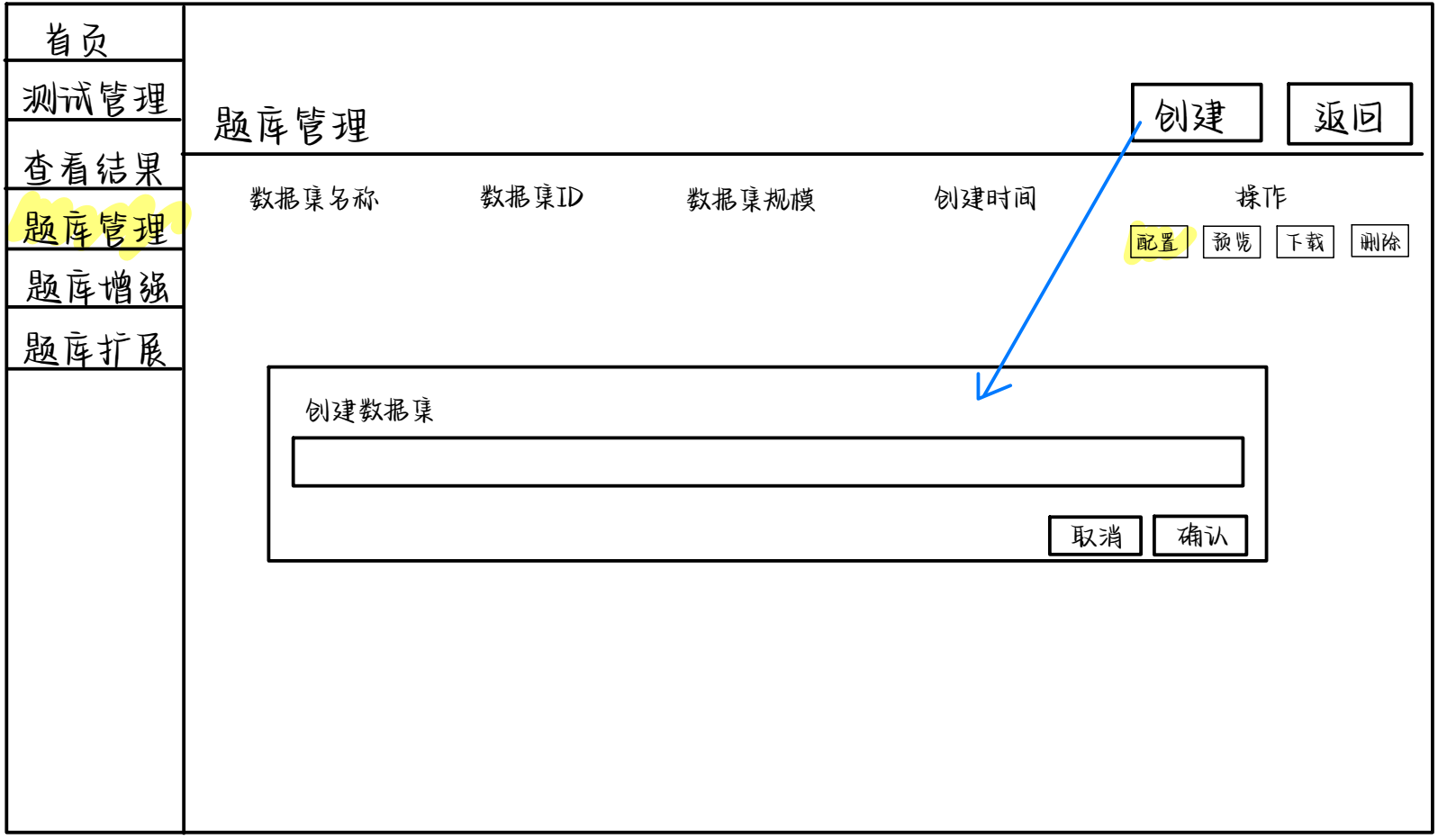
模型越狱率（ASR）：

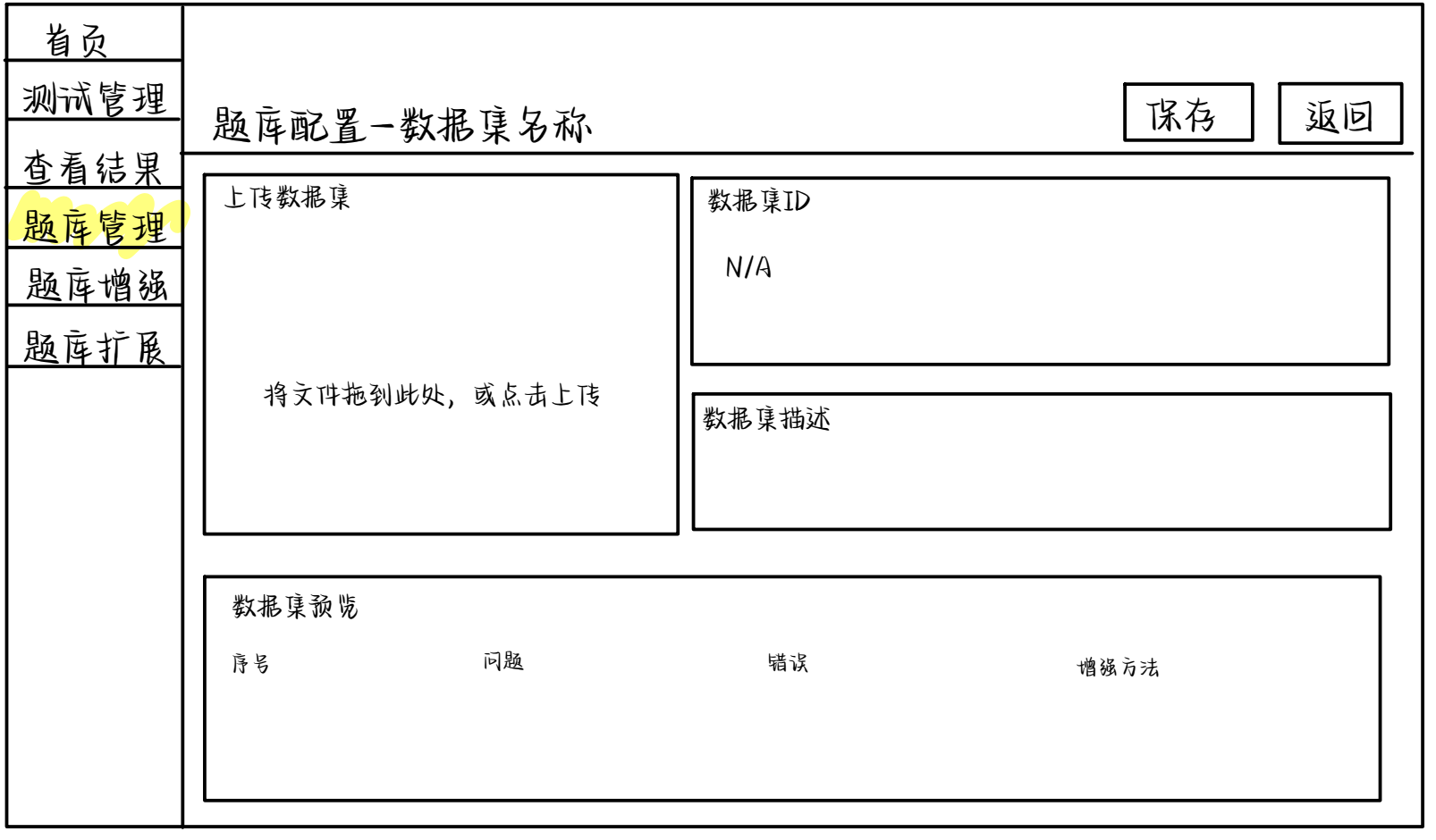
模型越狱率比较-领域

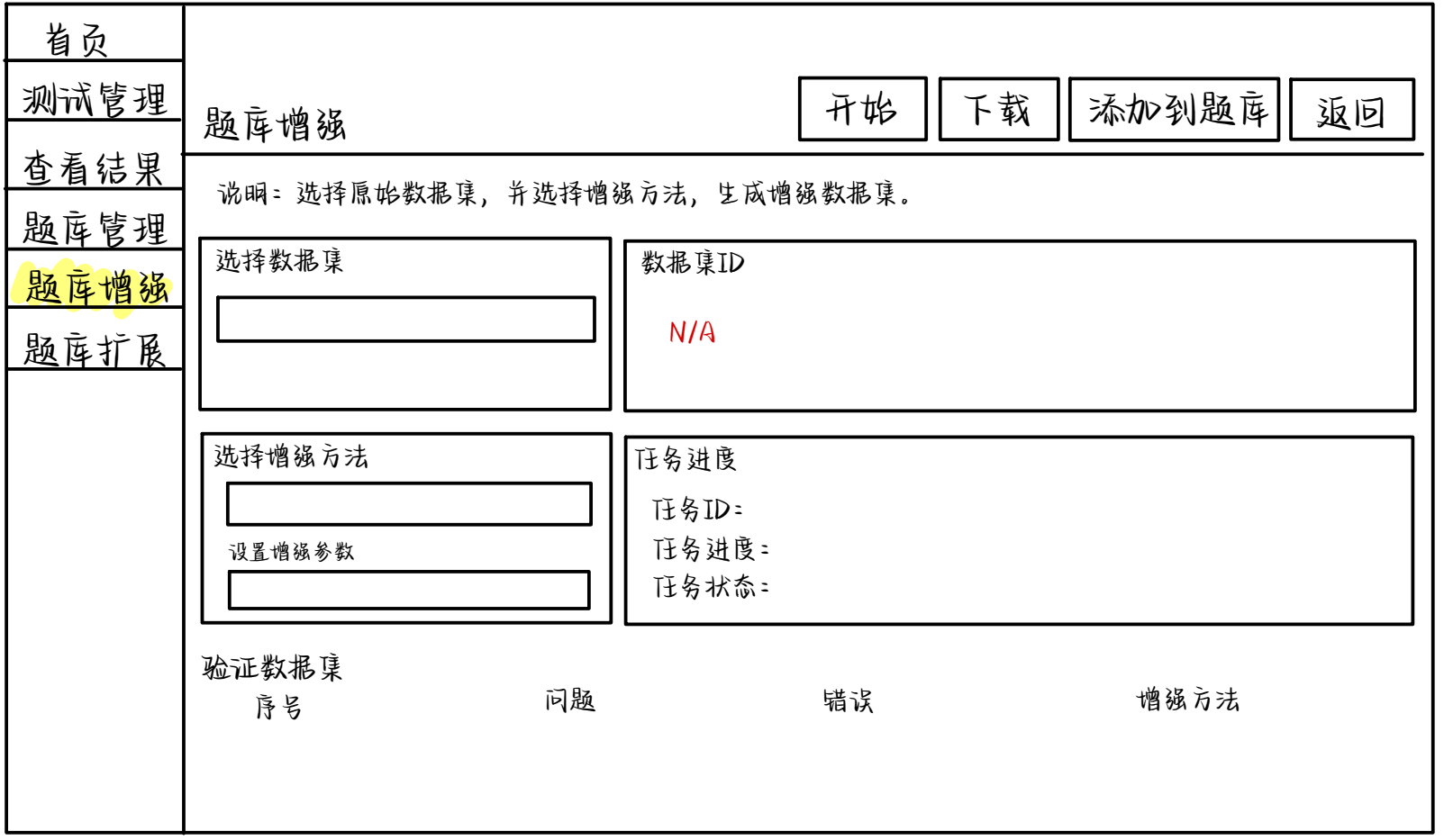
模型越狱率比较-增强方法

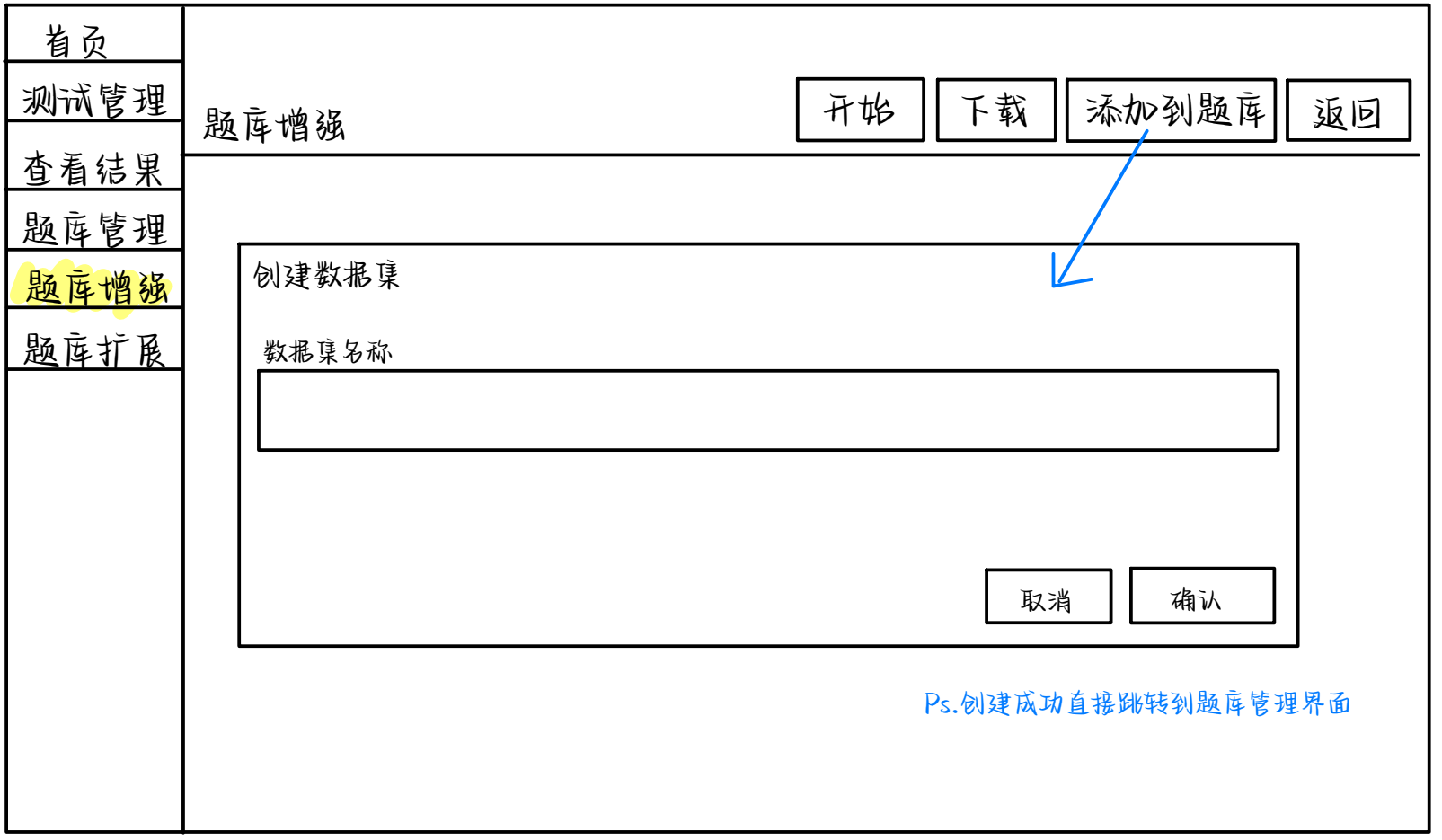
雷达图

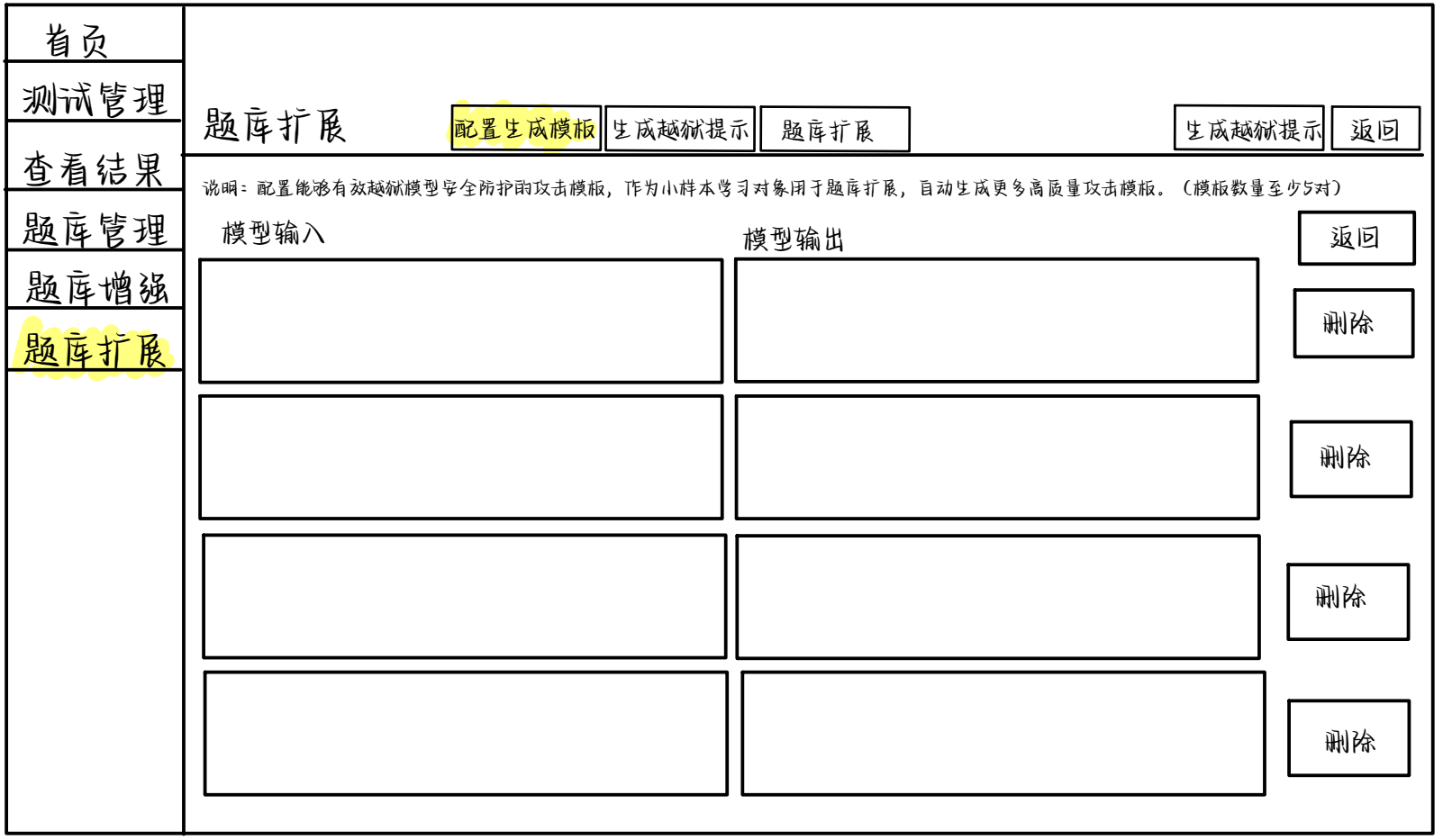
雷达图

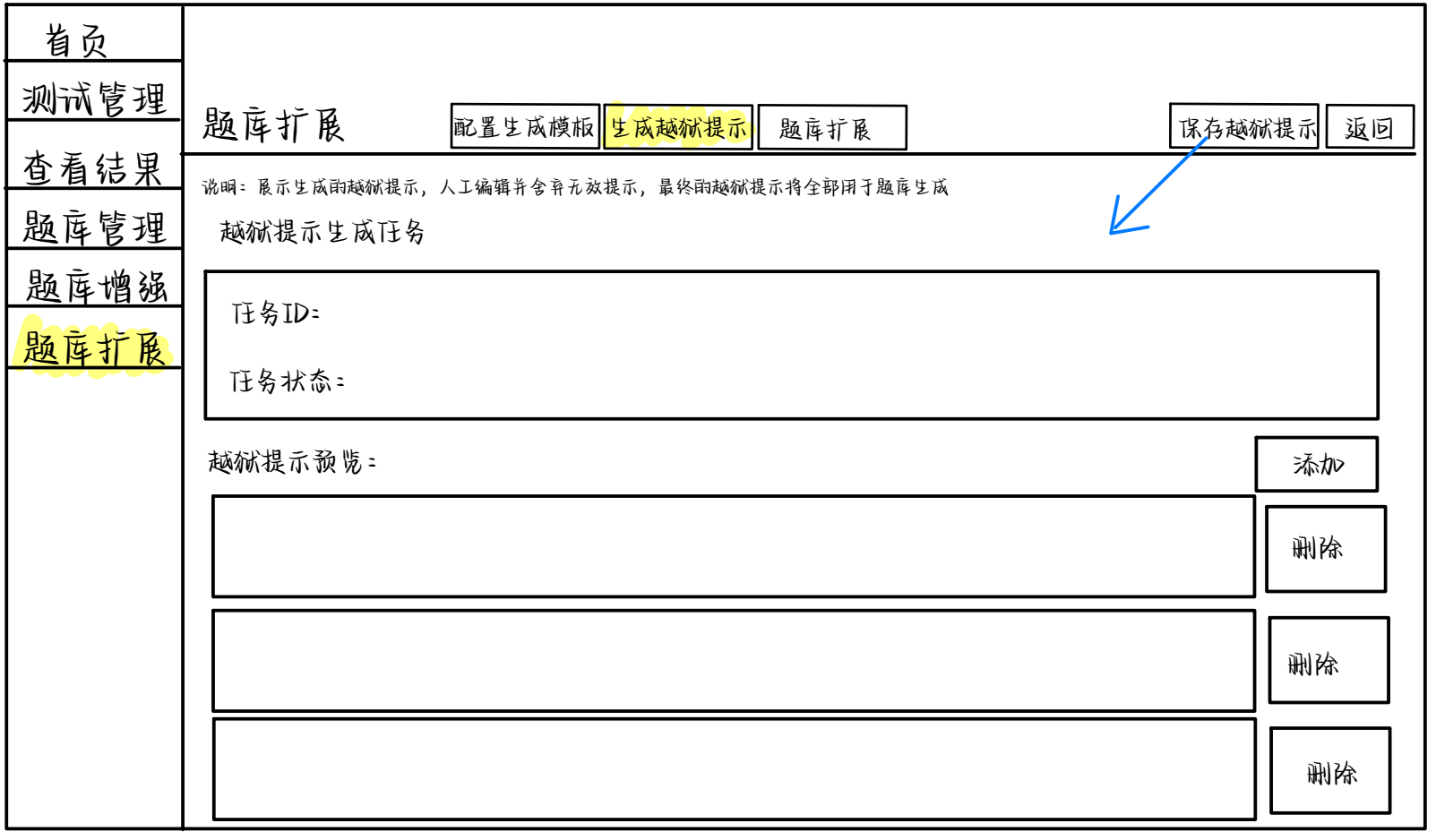

首页
测试管理
查看结果
题库管理
题库增强
题库扩展
题库管理
创建
返回
数据集名称
数据集ID
数据集规模
创建时间
操作
配置
预览
下载
删除
创建数据集
取消
确认

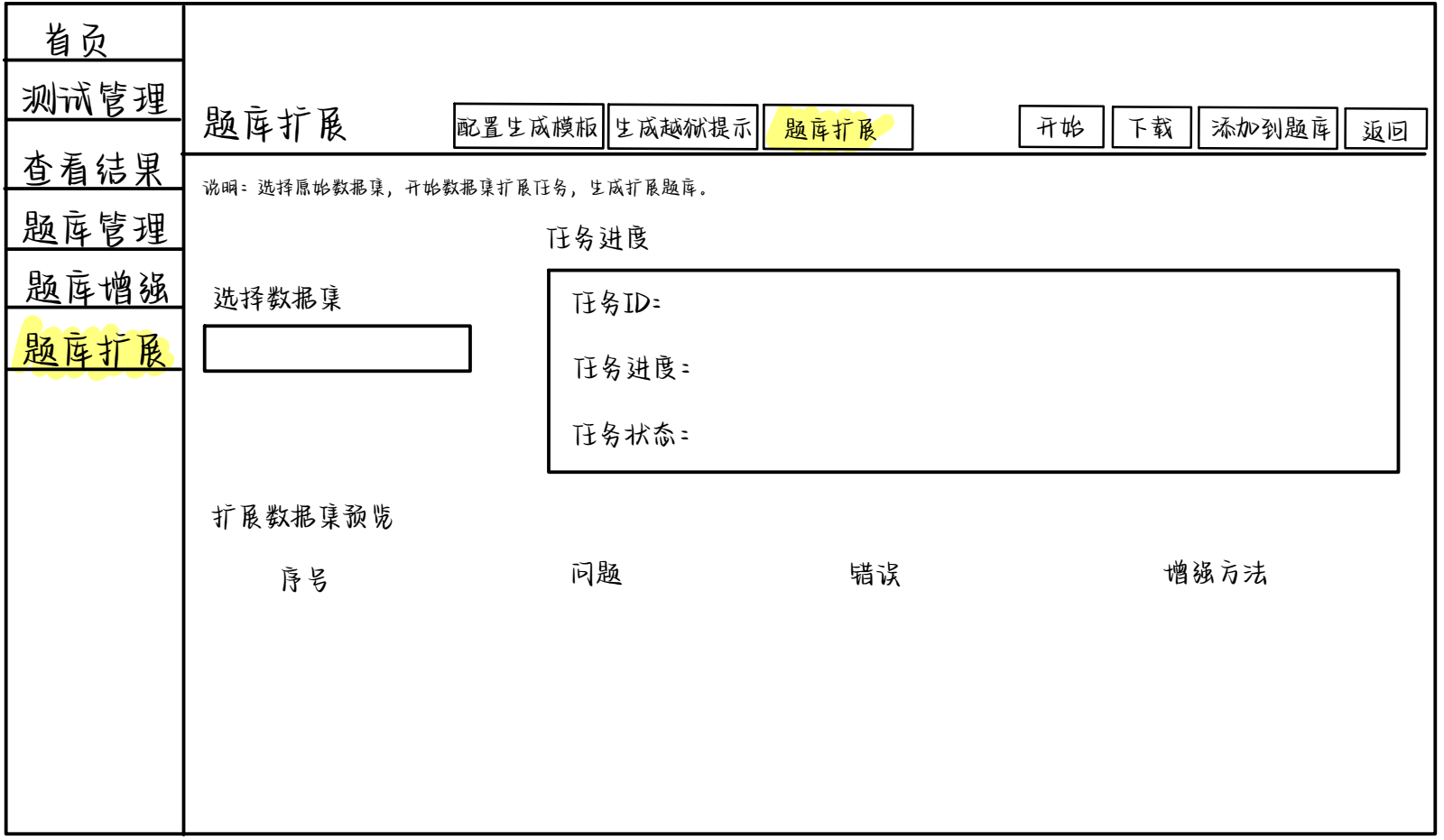

首页
测试管理
查看结果
题库管理
题库增强
题库扩展
题库配置-数据集名称
保存
返回
上传数据集
将文件拖到此处，或点击上传
数据集ID
N/A
数据集描述
数据集预览
序号
问题
错误
增强方法


首页
测试管理
查看结果
题库管理
题库增强
题库扩展
题库增强
开始
下载
添加到题库
返回
说明：选择原始数据集，并选择增强方法，生成增强数据集。
选择数据集
数据集ID
N/A
选择增强方法
设置增强参数
任务进度
任务ID：
任务进度：
任务状态：
验证数据集
序号
问题
错误
增强方法


首页
测试管理
查看结果
题库管理
题库增强
题库扩展
题库增强
开始
下载
添加到题库
返回
创建数据集
数据集名称
取消
确认
Ps.创建成功直接跳转到题库管理界面


首页
测试管理
查看结果
题库管理
题库增强
题库扩展
题库扩展
配置生成模板
生成越狱提示
题库扩展
生成越狱提示
返回
说明：配置能够有效越狱模型安全防护的攻击模板，作为小样本学习对象用于题库扩展，自动生成更多高质量攻击模板。（模板数量至少5对）
模型输入
模型输出
返回
删除
删除
删除
删除


首页
测试管理
查看结果
题库管理
题库增强
题库扩展
题库扩展
配置生成模板
生成越狱提示
题库扩展
保存越狱提示
返回
说明：展示生成的越狱提示，人工编辑并舍弃无效提示，最终的越狱提示将全部用于题库生成
越狱提示生成任务
任务ID：
任务状态：
越狱提示预览：
添加
删除
删除
删除


首页
测试管理
查看结果
题库管理
题库增强
题库扩展
题库扩展
配置生成模板
生成越狱提示
题库扩展
开始
下载
添加到题库
返回
说明：选择原始数据集，开始数据集扩展任务，生成扩展题库。
任务进度
选择数据集
任务ID:
任务进度:
任务状态:
扩展数据集预览
序号
问题
错误
增强方法