

# テスト開発の現場マネジメント

—制約の中でどう目的にかなったものを作るか—

藤田彩子

(株)リクルートキャリア 測定技術研究所

まず、「テスト開発の成功」を、①(品質や利便性などの)目的にかなったテストを②予算・時間の制約の中で完成させ③提供のための仕組みを用意すること、と置く。筆者の経験から、テスト開発での各ステップを様々な制限の中でうまく進めていくために重要と考えることをいくつか提示したい。

## 1. 問題項目の開発・作成

新しいテストには新しい問題が必要である。時にはこれまで測定したことがない「〇〇力」を測定するテストの開発ということもある。テストにおいて一番肝心なのは問題項目だが、「〇〇力」を受験者から引き出す問題項目の開発は試行錯誤となることも多い。したがって、問題項目の開発・作成にまずは早く着手すべきである。小社でも新しいテストを作ろうとした場合には、プロジェクトが本格化する数か月～半年くらい前には問題項目の開発に着手する。プロトタイプをいくつか作成し簡単に検証した上で、「この問題ならテストにできそう(尺度化できそう)」という感触を得てから、テスト開発が本格始動する。

開発が本格始動するあたりから、問題項目の作成体制を整える。具体的には、作成するのはどのような問題項目かを可能な限り言語化し、ガイドラインや例題などを用意し、作成スタッフを複数名(5～6人以上)集めてチームを作る。スタッフが問題案を作成して持ち寄り、皆で集まって問題案の検討会を行う。小社の開発の社員も必ず加わる。検討会では、問題案が誤解なく理解できるか、「〇〇力」の高低によって正誤が分かれるか、難易度は適切か、などを皆で議論する。多くの人の目を通すことで、作成時には気が付かなかった問題点が明らかになるため、やはり5～6人程度は必要である。率直な議論が求められるから人選も重要である。社員は会議のファシリテーター・書記として加わり、議論を整理しながら「〇〇力」のテスト項目としての適否を最終的に判断する。このような体制を早期に作るメリットとしては、安定的に問題項目が作られテストを編集する際に項目を選ぶことができるようになる、項目が大量に必要となる商品化・汎用化に対応できるようになる、作成ノウハウが維持・向上される、などがある。社員も問題作成の現場に加わることで、分析業務と合わせて「よいテスト問題」に対する具体的な知見が蓄積され、あらゆるテスト開発業務に生きる経験を積むことができる。

## 2. 尺度化・テスト化

問題項目がある程度できたところで、テストとして編集し、データを集める。十分な信頼性を得ようとする、問題項目数は多いに越したことはないが、一方で現実的な検査時間におさめなければならない。信頼性が上がらないとき、テスト側の要因としては、①項目数が少なすぎる②項目の“質”がよくない、の2点に集約されると考えられる。①については、検査時間を延長し問題を増やすことも考えられるが、短時間で回答できる問題を多く入れる、1つの題材に対して問題を複数つける(IRTを用いる場合などは特に、問題の局所独立に配慮する必要があるが)など、時間の割に多くの回答情報が得られる工夫が有効である。特に、短時間で回答できる問題は、大概シンプルな問題であり、「〇〇力」を端的に測定し、質も良くなりやすい。②については、一般的には問題が複雑になればなるほど、そ

の問題の独自性が高くなるため、何問かまとめてテストにしても測定の”軸”が定まらず、失敗しやすい。測定領域の都合上、どうしても複雑な問題が必要な場合も、その領域に関連するできるだけシンプルな問題を一定数含めておけば(例えば国語のテストに文章読解だけでなく語彙の単問を含めるなど)、そのシンプルな問題達が”軸”を作ってくれて、複雑な問題についても、その性質を判断しやすくなる。こうしたことを考慮しながらテストを編集し、データを集める。

### 3. データ収集

データ収集の手段としては、身近な人達に協力してもらい、調査会社などを使ってモニターを集めたりネット調査をしたりする、すでに運用されているテストに紛れ込ませる、など様々考えられる。データ収集は多くの場合、お金をかけるか手間と時間をかけるかのせめぎ合いだが、データは数だけでなくモニターとして適切な集団かという質の観点が重要である。また、データ収集の目的(使える問題かざっくり判断したいのか、項目パラメータを求めたいのかなど)によっても、データの量・質に対する考え方・割り切りが異なる。開発の早い段階でのデータ収集は、問題項目の質を判断したり、改善・修正の方向性を見極めたりするためのことが多いので、人数もそこまで多くなくても、ある程度の役立つ情報を得ることができる。問題の質や尺度としてのまとまりを判断するなら、(測定領域やデータの質にもよるが)100程度でも分かることは多い。ただし、例えばIRTのパラメータを求めたい場合などは、通常最低500くらいのデータは必要で、モニター集団の性質によってもパラメータが変わってしまうことを考えると、お金をかけ、工夫を凝らすならココと考える。データ収集は、お金や手間暇をかけたとしても、理想的なデータが集まることはまずない。企画段階での工夫、実施の工夫はもちろん、得られたデータの質の判断や、分析方法の選択、データから得られた情報をどう解釈・活用するかも、テスト開発の成功に直接関係する。

### 4. システム構築

テストを汎用化していく上では、出題・採点・運用をシステム化する必要がある。コンピュータシステムを開発する場合は、世の中の様々な開発プロジェクトと同様、パートナーの開発会社との密なコミュニケーションが重要である。適応型テストなど、凝った出題・採点の仕組みをシステムに載せる場合は、仕様を丁寧に伝えるだけでなく、要所所でロジックの検証ができるようなデータを渡すなど、限られた時間の中でスムーズに開発が進むよう工夫する必要がある。また、そもそも、後からシステム検証が可能なような仕様におかなくてはならない。様々な観点からユーザーテストを行い、テストを世に送り出す立場として、間違いのない品質のシステムを完成させる責任がある。要するに、システムの要件定義・開発・検証をパートナーのシステム開発会社に任せきりにしないということである。どんなに素晴らしいテストもシステムがコケたら元も子もない。出題・回答画面(紙筆版の場合は問題冊子・回答用紙)に間違いがなく適切に表示されるか、正答やスコア化のためのパラメータは正しく登録されているかなどを念入りに確認することも言うまでもない。

### 5. 最後に～どこまでがテスト開発か?

長い目で見たテストの成功を、「多くの人に利用され、信頼され、役に立つこと」と置いたとき、テスト開発は、問題を作り、尺度化し、データを分析し、システム化するだけでは終わらない。多くの人に利用されるためには、新しい問題項目を継続的に作り続けてゆかなければならず、信頼され役に立つためには、テストの実施データを蓄積し分析し、妥当性を検証し、世の中に発信していく必要があるだろう。そのための項目管理システムや分析システム、それらを活用しての様々な面からの改善がそのテストが続いていく限り求められる。