**گزارش فنی و پژوهشی پروژه «آزمون تطبیقی رایانه‌ای زبان فارسی مبتنی بر IRT»**

**چکیده**

این پژوهش به طراحی و پیاده‌سازی یک سامانهٔ آزمون تطبیقی رایانه‌ای (CAT) برای سنجش مهارت‌های زبان فارسی بر پایهٔ نظریهٔ پاسخ به سؤال (IRT) می‌پردازد. سامانه با استفاده از مدل سه‌پارامتری 3PL، توانایی آزمون‌شوندگان (θ) را در جریان آزمون به‌صورت برخط برآورد می‌کند و با بیشینه‌سازی اطلاعات آزمون، آیتم بعدی را انتخاب می‌نماید. معماری نرم‌افزار با Flask و SQLite پیاده‌سازی شده و ماژول‌های تکمیلی شامل ثبت فرا‌دادهٔ‌ آزمون (session logging)، اندازه‌گیری زمان پاسخ‌دهی (RT)، پرسشنامهٔ «راهبردهای آزمون‌دادن» (post-test) و دریافت پاسخ‌های صوتی برای نقش «مدیر مرکز آموزش زبان فارسی» را پوشش می‌دهد. طرح تحلیل داده‌ها شامل کالیبراسیون آیتم‌ها، سنجش دقت (SEM)، شاخص‌های برازش شخصی/آیتمی، تحلیل زمان پاسخ‌دهی، و کنترل مواجههٔ آیتم‌ها ارائه شده است. نتایج این کار زیرساختی پایدار برای اجرای CAT فارسی و تحلیل‌های سنجشی سطح پیشرفته فراهم می‌آورد.

**واژگان کلیدی:** آزمون تطبیقی (CAT)، نظریهٔ پاسخ به سؤال (IRT)، 3PL، برآورد توانایی، اطلاعات آزمون، زمان پاسخ‌دهی، پرسشنامهٔ راهبردها.

**۱. مقدمه و بیان مسئله**

گسترش آموزش زبان فارسی به غیر فارسی‌زبانان، نیاز به ابزارهای سنجش دقیق، کارآمد و انعطاف‌پذیر را افزایش داده است. آزمون‌های سنتی با دفترچه‌ٔ ثابت، از نظر کارآمدی زمان، امنیت آیتم‌ها و دقت در نواحی هدف توانایی محدودیت دارند. آزمون تطبیقی رایانه‌ای (CAT) با انتخاب پویای آیتم‌ها بر اساس برآورد لحظه‌ای θ، آزمون کوتاه‌تر، دقیق‌تر و شخصی‌تر ارائه می‌کند. هدف این پژوهش، طراحی و پیاده‌سازی یک سامانهٔ CAT فارسی مبتنی بر IRT و آماده‌سازی آن برای تحلیل‌های میدانی و پژوهشی است.

**سؤالات پژوهش:**

1. چگونه می‌توان مدل 3PL را در یک سامانهٔ CAT عملیاتی کرد؟
2. چه معیارهایی برای توقف آزمون دقت و کارآمدی بهتری دارند؟
3. چگونه می‌توان با ثبت فرا‌داده‌ها (RT، اطلاعات آیتم، دلایل توقف) تحلیل‌های سنجشی را غنی کرد؟
4. پرسشنامهٔ راهبردها چه تصویری از رفتار آزمون‌دادن ارائه می‌دهد و چگونه با داده‌های CAT همبسته می‌شود؟

**۲. مبانی نظری (IRT و CAT)**

**۲.۱. مدل سه‌پارامتری 3PL**

برای هر آیتم با پارامترهای تبعیض aa، دشواری bb و حدس cc، احتمال پاسخ درست فردی با توانایی θ\theta به‌صورت زیر است:

P(correct∣θ,a,b,c)  =  c+(1−c) σ ⁣(a(θ−b)),σ(x)=11+e−x.P(\text{correct} \mid \theta,a,b,c) \;=\; c + (1 - c)\,\sigma\!\big(a(\theta - b)\big),\quad \sigma(x)=\frac{1}{1+e^{-x}}.

**۲.۲. برآورد توانایی و دقت**

برآورد θ\theta به دو صورت انجام شده است:

* **MAP** در ابتدای آزمون (۲–۳ سؤال اول) برای پایداری: θ∼N(0,1)\theta \sim \mathcal{N}(0,1).
* **MLE** پس از تثبیت پاسخ‌ها.

اطلاعات آزمون در θ\theta برابر I(θ)=∑iIi(θ)\displaystyle I(\theta)=\sum\_i I\_i(\theta) است و خطای استاندارد برآورد:

SE(θ)=1I(θ).SE(\theta) = \frac{1}{\sqrt{I(\theta)}}.

**۲.۳. انتخاب آیتم و معیار توقف**

در هر گام، آیتمِ بیشترین اطلاعات در θ\theta فعلی انتخاب می‌شود (با محافظ‌های عددی). معیار توقف ترکیبی است:

* دستیابی به دقت هدف (مثلاً SE≤0.30SE \le 0.30) پس از حداقل تعداد آیتم؛
* قطعیت نسبت به مرز طبقه‌بندی (بازهٔ اطمینان ۹۵٪ به‌طور کامل بالای/پایین مرز)؛
* پایداری تغییر ∣Δθ∣|\Delta\theta| در چند سؤال پیاپی؛
* سقف ایمنی تعداد آیتم‌ها.

**۳. معماری و پیاده‌سازی سامانه**

**۳.۱. فناوری‌ها**

* **Backend:** Python/Flask
* **DB:** SQLite (WAL, busy\_timeout)
* **Front-end:** قالب‌های Jinja2، راست‌به‌چپ، فونت Vazirmatn، کارت‌های سبک
* **کتابخانه‌ها:** NumPy/Matplotlib/Pandas/Docx

**۳.۲. ساختار پایگاه‌داده (هسته)**

* participants(participant\_id, ... , role)
* questions(id, text, option1..4, correct\_option, a,b,c)
* answers(user\_id, question\_id, selected\_option, is\_correct, ...)
* user\_results(user\_id, theta)
* **ویژه مدیر:** manager\_questions، voice\_answers (ذخیرهٔ مسیر فایل صوتی، MIME، مدت)
* **پرسشنامهٔ پس‌آزمون:**
  + strategies(id, strategy, category, frequency) – بانک سؤالات
  + strategy\_answers(participant\_id, strategy\_id, choice, ...) – پاسخ‌ها (۱..۵، UPSERT)

برای تحلیل‌های پیشرفته، جداول پیشنهادی زیر نیز در طراحی پژوهش پیش‌بینی شده‌اند:  
test\_sessions(session\_id, user\_id, started\_at, ended\_at, stop\_reason, theta\_final, تنظیمات CAT, نسخهٔ پارامترها)،  
answers\_meta(session\_id, order\_idx, theta\_before/after, se\_after, info\_at\_theta, rt\_ms)،  
item\_param\_versions و item\_parameters برای نسخه‌بندی پارامتر آیتم‌ها.

**۳.۳. منطق اصلی آزمون**

* **شروع:** انتخاب آیتم آغازین با بیشترین اطلاعات حول prior (θ=0\theta=0).
* **ثبت پاسخ:** دریافت گزینهٔ ۱..۴، محاسبهٔ درستی/نادرستی، ثبت در DB، و در صورت فعال بودن، ثبت RT و فرا‌داده در answers\_meta.
* **برآورد θ:** MAP برای ابتدای آزمون؛ سپس MLE.
* **توقف:** بر مبنای SE، قطعیت طبقه‌بندی، پایداری ∣Δθ∣|\Delta\theta|، و سقف ایمنی.
* **نتایج:** نمایش θ\theta، SE، CI ۶۸٪/۹۵٪، تعداد پاسخ‌ها، دقت، نمودار ICC و اطلاعات آزمون، و «علت توقف».

**۳.4. رابط‌های کاربری**

* **ثبت‌نام/انتخاب نقش:** شرکت‌کننده/یادگیرنده/مدرس/مدیر مرکز.
* **آزمون تطبیقی:** صفحهٔ تک‌سؤالی با چهار گزینه (واکنش‌گرا)، نوار پیشرفت.
* **نتایج:** خلاصهٔ عددی، تفسیر کیفی (پایین/متوسط/بالا)، نمودارهای ICC/اطلاعات، لینک «پس‌آزمون».
* **پس‌آزمون (پرسشنامهٔ راهبردها):** مقیاس لیکرت ۱..۵، گروه‌بندی موضوعی بر اساس category.
* **مدیر مرکز (ویس):** نمایش سؤالات و کنترل ضبط/ارسال صوت (WebM/OGG/MP4)، ذخیرهٔ امن مسیر فایل.

**۳.۵. امنیت و حریم خصوصی**

* کلید محرمانهٔ Flask و محدودیت حجم آپلود (۱۵MB) فعال است.
* مسیر فایل‌های صوتی نسبی و در static/voices/{participant\_id}/ نگهداری می‌شود.
* پیشنهاد اجرایی: جداسازی PII از دادهٔ تحلیلی (کلید ناشناس)، خط‌مشی نگهداری و حذف دوره‌ای فایل‌های صوتی، و اخذ رضایت آگاهانه.

**۴. الگوریتم‌های سنجشی و جزئیات محاسباتی**

* **پایداری عددی:** کلیپ‌کردن آرگومان سیگموید، جلوگیری از p∈{0,1}p\in\{0,1\} با ε\varepsilon.
* **اطلاعات آیتم در 3PL (فرم عملیاتی):** محاسبهٔ Ii(θ)I\_i(\theta) با درنظرگرفتن cc و پایدارسازی مخرج.
* **بهینه‌سازی:** نیوتن–رافسون روی θ\theta با گام‌های محدودشده؛ برای MAP افزودن گرادیان prior.
* **معیارهای توقف:**
  + **دقت:** SE(θ)≤0.30SE(\theta) \le 0.30 بعد از ≥8\ge 8 آیتم؛
  + **قطعیت طبقه‌بندی:** [θ±1.96 SE][\theta \pm 1.96\,SE] کاملاً یک‌سوی مرز؛
  + **پایداری:** ∣Δθ∣<0.03|\Delta\theta| < 0.03 برای ≥2\ge 2 گام پیاپی؛
  + **سقف:** حداکثر 22 آیتم.
* **کنترل مواجهه (پیشنهاد اجرایی):** انتخاب تصادفی از بین ۵ آیتم با اطلاعاتِ بالاتر (Randomesque) و توازن محتوا با content\_code.

**۵. گردآوری داده‌ها و پروتکل اجرا**

1. **نمونه:** زبان‌آموزان و مدرسان/مدیران مرکز؛ معیارهای ورود مشخص (سن، سطح زبان، تجربهٔ آزمون).
2. **رویه:** ثبت‌نام → اجرای CAT → مشاهدهٔ نتایج → تکمیل پس‌آزمون (راهبردها) → (در صورت نقش مدیر) پاسخ صوتی.
3. **داده‌های ثبت‌شده:** پاسخ‌های آیتمی، درستی/نادرستی، θ\theta و SE در هر گام (در صورت فعال بودن answers\_meta)، RT، علت توقف، پاسخ‌های پرسشنامه، و—در نقش مدیر—فایل‌های صوتی.
4. **رضایت آگاهانه:** اطلاع از اهداف پژوهش، نحوهٔ استفاده از داده‌ها، حق انصراف و محرمانگی.
5. **اخلاق پژوهش:** حذف شناسه‌های مستقیم، ذخیرهٔ امن و دسترسی محدود، استفادهٔ پژوهشی.

**۶. طرح تحلیل داده‌ها**

**۶.۱. تحلیل سنجشی (IRT)**

* **کالیبراسیون آیتم‌ها:** برآورد پارامترهای a,b,ca,b,c با روش‌های MML/EM بر روی داده‌های جمعی؛ ایجاد نسخهٔ پارامترها و ثبت در item\_param\_versions.
* **برازش:** شاخص‌های برازش آیتمی/شخصی (مانند lzl\_z، HtH\_t)؛ بررسی الگوهای پاسخ‌های ناسازگار.
* **دقت اندازه‌گیری:** تحلیل SEM در بازه‌های θ\theta، رسم تابع اطلاعات آزمون، مقایسهٔ معیارهای توقف.
* **تحلیل DIF:** بررسی عملکرد افتراقی آیتم‌ها بین گروه‌ها (مثلاً ملیت/زبان مادری) با سنجه‌های مبتنی بر IRT.
* **روایی و پایایی:** همبستگی با ملاک‌های بیرونی (در صورت وجود)، ثبات نتایج با اجرای تکراری.

**۶.۲. تحلیل رفتار آزمون (RT و راهبردها)**

* **RT:** توزیع زمان پاسخ‌دهی، پرچم‌گذاری «حدس سریع» (مثلاً RT<800ms)، ارتباط RT با درستی/دشواری.
* **راهبردها:** میانگین/انحراف معیار گزینه‌های ۱..۵ در strategy\_answers، تحلیل عاملی اکتشافی/تأییدی (در صورت کفایت نمونه)، همبستگی با θ\theta و دقت.
* **پیوند CAT-پرسشنامه:** رگرسیون‌های ساده/چندگانه برای پیش‌بینی θ\theta یا دقت بر اساس راهبردهای گزارش‌شده و RT.

**۷. ارزیابی سامانه و کیفیت اندازه‌گیری**

* **کارآمدی:** میانگین تعداد آیتم‌ها تا توقف، در مقابل دقت هدف.
* **پایداری برآورد:** مقایسهٔ θ\theta نهایی با اجراهای تکراری یا آزمون مرجع.
* **پوشش دشواری:** توزیع پارامتر bb آیتم‌های منتخب برای آزمون‌شوندگان با θ\theta متفاوت.
* **تجربهٔ کاربر:** نرخ تکمیل، زمان کل آزمون، رضایت‌سنجی سبک.

**۸. ملاحظات اخلاقی و حقوقی**

* **حریم خصوصی:** جداسازی PII از دادهٔ تحلیلی؛ استفاده از کلیدهای ناشناس.
* **فایل‌های صوتی:** اطلاع‌رسانی هدف و مدت نگهداری؛ امکان حذف به درخواست.
* **کاربرد پژوهشی:** محدودیت دسترسی، عدم انتشار دادهٔ خام شناسایی‌پذیر.
* **رضایت آگاهانه:** فرم‌های شفاف با حق انصراف.

**۹. محدودیت‌ها**

* اندازهٔ بانک آیتم‌ها و کیفیت کالیبراسیون اولیه می‌تواند دقت را محدود کند.
* SQLite برای استقرارهای بزرگ مناسب نیست (پیشنهاد مهاجرت به PostgreSQL).
* پیامدهای رفتاری (خستگی، انگیزش) بر RT و دقت اثر می‌گذارند و باید در تحلیل کنترل شوند.
* تحلیل‌های DIF/مدل‌های چندبعدی نیازمند حجم نمونهٔ کافی است.

**۱۰. جمع‌بندی و مسیر آینده**

این سامانهٔ CAT فارسی با مدل 3PL، انتخاب آیتمِ اطلاعات‌محور، توقف هوشمند و لایه‌های تکمیلی (RT، پرسشنامهٔ راهبردها، پاسخ‌های صوتی نقش مدیر) پیاده‌سازی و آمادهٔ استقرار پژوهشی است. گام‌های آینده: کالیبراسیون گستردهٔ آیتم‌ها با دادهٔ میدانی، کنترل مواجهه با سیاست‌های پیشرفته، توازن محتوایی خودکار، تحلیل‌های DIF و روایی ملاکی، و مهاجرت به DB رابطه‌ای مقیاس‌پذیر جهت استقرار سازمانی.

**پیوست A: طرح پایگاه‌دادهٔ عملیاتی (گزیده)**

**participants**(participant\_id PK, name, nationality, mother\_tongue, official\_language, age, major, education\_level, job, role)  
**questions**(id PK, text, option1..4, correct\_option, a,b,c)  
**answers**(id PK, user\_id FK, question\_id FK, selected\_option, is\_correct, created\_at)  
**user\_results**(user\_id PK, theta)  
**manager\_questions**(id PK, text, is\_active, display\_order)  
**voice\_answers**(id PK, participant\_id FK, role, question\_id FK, file\_path, mime\_type, duration\_ms, size\_bytes, created\_at)  
**strategies**(id PK, strategy, category, frequency)  
**strategy\_answers**(id PK, participant\_id FK, strategy\_id FK, choice, created\_at, updated\_at, UNIQUE(participant\_id, strategy\_id))

جداول پیشنهادی برای پژوهش: **test\_sessions**، **answers\_meta**، **item\_param\_versions**، **item\_parameters**.

**پیوست B: شبه‌کد انتخاب آیتم و توقف**

**انتخاب آیتم:**

inputs: theta, answered\_indices, all\_item\_params = [(a,b,c), ...]

candidates = []

for i,(a,b,c) in enumerate(all\_item\_params):

if i in answered\_indices: continue

info = item\_information(theta, a, b, c) # فرم پایدار 3PL

candidates.append((i, info))

sort candidates by info desc

pick next\_idx = random choice from top-5 (یا argmax برای حالت پایه)

return next\_idx

**توقف:**

if n >= MIN\_QUESTIONS:

if SE(theta) <= SE\_TARGET: stop("دقت کافی")

if CI95(theta) بالای/پایین CUT\_SCORE: stop("قطعیت نسبت به مرز")

if |Δθ| < DELTA\_TARGET برای STREAK\_NEED بار: stop("پایداری θ")

if n >= HARD\_MAX: stop("سقف ایمنی")

**پیوست C: فرمول‌های کلیدی**

* احتمال درست در 3PL:  
  P=c+(1−c) σ(a(θ−b))P = c + (1-c)\,\sigma(a(\theta-b))
* اطلاعات آزمون:  
  I(θ)=∑iIi(θ)I(\theta) = \sum\_i I\_i(\theta)
* خطای استاندارد:  
  SE(θ)=1/I(θ)SE(\theta) = 1/\sqrt{I(\theta)}
* MAP (اولیه):  
  بیشینه‌سازی log⁡L(θ∣resp)+log⁡N(θ;0,1)\log L(\theta \mid \text{resp}) + \log \mathcal{N}(\theta;0,1)

**یادداشت اجرایی**

نسخهٔ فعلی کد، همهٔ اجزای اصلی را پوشش می‌دهد و برای فاز میدانی آماده است. برای فصل «نتایج» پایان‌نامه، توصیه می‌شود پس از گردآوری دادهٔ واقعی:

* توصیف نمونه،
* توزیع θ\theta، SEM برحسب θ\theta،
* آمار توقف (میانگین آیتم‌ها تا توقف، دلایل توقف)،
* همبستگی θ\theta با نمرات ملاک (در صورت وجود)،
* خلاصهٔ پرسشنامهٔ راهبردها و ارتباط آن با عملکرد،
* نمودارهای ICC و تابع اطلاعات آزمون،  
  به‌همراه جداول خلاصه ارائه گردد.

این پژوهش به طراحی و پیاده‌سازی یک سامانهٔ آزمون تطبیقی رایانه‌ای (CAT) برای سنجش مهارت‌های زبان فارسی مبتنی بر نظریهٔ پاسخ به سؤال (IRT) می‌پردازد و می‌کوشد سنجه‌ای دقیق، کارآمد و شخص‌ـ‌سازگار فراهم کند. مسئلهٔ اصلی، محدودیت آزمون‌های سنتیِ ثابت در دقت اندازه‌گیری در ناحیه‌های هدف توانایی و کارآمدی زمانی است. در مقابل، CAT با انتخاب پویای آیتم‌ها بر اساس برآورد لحظه‌ای توانایی آزمون‌شونده (θ) و تخصیص سؤالات با بیشترین اطلاعات، هم طول آزمون را کاهش می‌دهد و هم دقت را در همان تعداد آیتم افزایش می‌دهد. پروژهٔ حاضر با پیاده‌سازی مدل سه‌پارامتری 3PL، طراحی معیار توقف دقیق‌محور، و ثبت فرا‌داده‌های تحلیلی (مانند زمان پاسخ‌دهی) زیرساخت لازم برای پژوهش‌های سنجشی و بهبود مستمر بانک آیتم‌های فارسی را فراهم کرده است.

از نظر نظری، مدل احتمال پاسخ درست در 3PL بر حسب پارامترهای تبعیض aa، دشواری bb و حدس cc تعریف می‌شود و احتمال موفقیت فردی با توانایی θ\theta را به‌صورت c+(1−c)σ(a(θ−b))c + (1-c)\sigma(a(\theta-b)) مدل می‌کند که در آن σ\sigma تابع لجستیک است. برآورد θ\theta در سامانه به‌صورت مرحله‌ای انجام می‌شود: در نخستین گام‌ها از روش MAP با پیشین نرمال N(0,1)N(0,1) برای پایداری استفاده شده و پس از چند پاسخ اولیه، برآورد به سمت MLE تغییر می‌کند تا از اطلاعات داده‌ها بیشترین بهره گرفته شود. برای انتخاب آیتم بعدی، تابع اطلاعات آیتم‌ها در θ\theta فعلی محاسبه و آیتم با بیشترین اطلاعات گزینش می‌شود؛ برای جلوگیری از ناپایداری عددی، محافظ‌های لازم برای نزدیک‌شدن pp به ۰ یا ۱ اعمال شده است. دقت برآورد در هر لحظه با SE(θ)=1/∑Ii(θ)SE(\theta)=1/\sqrt{\sum I\_i(\theta)} سنجیده می‌شود و مبنای اصلی توقف آزمون قرار می‌گیرد.

طراحی معیار توقف به‌گونه‌ای است که آزمون به‌محض رسیدن به «کفایت دقت» متوقف شود و وابسته به سقف دلخواه تعداد آیتم‌ها نباشد. بدین منظور، چند معیار مکمل در کنار هم به‌کار رفته است: کاهش خطای استاندارد برآورد تا آستانهٔ هدف (برای نمونه SE≤0.30SE \le 0.30) پس از حداقل تعداد آیتم، قطعیت طبقه‌بندی نسبت به نمرهٔ مرزی (قرارگرفتن کامل بازهٔ اطمینان ۹۵٪ در یک سوی مرز)، پایداری تغییر برآورد (∣Δθ∣|\Delta\theta| کمتر از آستانهٔ کوچک در چند گام پیاپی) و نهایتاً یک سقف ایمنی برای شمار آیتم‌ها در شرایط خاص. این ترکیب، میانگینی از طول آزمون را پایین می‌آورد، دقت را یکنواخت می‌کند و از طولانی‌شدن بی‌ثمر جلوگیری می‌نماید.

از منظر پیاده‌سازی، لایهٔ کاربردی با Flask و ذخیره‌سازی سبک با SQLite پیاده شده است. برای سازگاری با استفادهٔ آموزشی و پژوهشی، الگوی «اتصال کوتاه» و حالت WAL همراه با زمان‌انتظار مناسب به‌کار رفته تا تزاحم‌های متعارف SQLite کاهش یابد. معماری داده‌ها شامل جدول شرکت‌کنندگان و جدول سؤالات با پارامترهای aa، bb، cc، جدول پاسخ‌های آیتمی، و جدول نتایج نهایی کاربر است. در کنار هستهٔ CAT، نقش «مدیر مرکز آموزش زبان فارسی» پیش‌بینی شده که در آن پاسخ به پرسشنامهٔ نقش به‌صورت صوتی دریافت و فایل‌ها با MIME معتبر، حجم محدود و مسیر نسبی امن در ساختار static/voices/{participant\_id}/ ذخیره می‌شوند. پس از اتمام آزمون، صفحهٔ نتایج علاوه بر نمایش θ، SE و بازه‌های اطمینان ۶۸ و ۹۵ درصدی، تعداد پاسخ‌ها، شمار پاسخ‌های درست و نادرست، و تفسیر کیفی سطح توانایی را ارائه می‌کند. به‌منظور تکمیل داده‌های رفتاری، یک «پس‌آزمون» به‌شکل پرسشنامهٔ راهبردهای آزمون‌دادن نیز طراحی شده است؛ بانک سؤالات این پرسشنامه در جدول strategies نگهداری می‌شود و پاسخ‌ها در جدول strategy\_answers با مقیاس لیکرت ۱ تا ۵ و قید یکتای ترکیبی ‌(شرکت‌کننده–سؤال) ذخیره می‌شود تا تحلیل‌های بعدی از سازگاری راهبردهای خوداظهاری با عملکرد CAT ممکن شود.

برای تحلیل‌پذیری دقیق‌تر داده‌ها، لایهٔ ثبت فرا‌دادهٔ آزمون (session logging) پیشنهاد و در طراحی پژوهش پیش‌بینی شده است. هر اجرای CAT با یک شناسهٔ نشست (UUID) در جدول test\_sessions ثبت می‌شود و متغیرهایی مانند زمان آغاز و پایان، دلیل توقف، تعداد آیتم‌های ارائه‌شده، θ نهایی، نسخهٔ پارامترهای آیتم و پارامترهای عملیاتی آزمون (حداقل آیتم، سقف، آستانهٔ SE، آستانهٔ ∣Δθ∣|\Delta\theta|، مرز طبقه‌بندی) ذخیره می‌گردد. هم‌زمان، جدول سبک answers\_meta فرا‌دادهٔ هر پاسخ را نگه می‌دارد: ترتیب نمایش، گزینهٔ انتخابی، درستی/نادرستی، θ پیش از پاسخ، θ و SE پس از پاسخ، اطلاعات آیتم در θ و زمان پاسخ‌دهی (RT). به‌منظور اندازه‌گیری RT، فرم کاربر یک فیلد پنهان دارد که با استفاده از زمان‌سنج مرورگر مقداردهی می‌شود. این ثبت فرا‌داده، امکان تحلیل‌های پیشرفته‌ای مانند تشخیص حدس‌زدن سریع، محاسبهٔ شاخص‌های شخص‌ـ‌فیت و برآورد پایداری مسیر θ\theta را فراهم می‌کند و بازتولید نتایج را در نسخه‌های بعدی بانک آیتم تضمین می‌نماید.

در مسیر تحلیل داده‌ها، کالیبراسیون آیتم‌ها با روش‌های حداکثر درست‌نمایی حاشیه‌ای و EM بر مجموعه‌های دادهٔ بزرگ‌تر انجام خواهد شد و نسخهٔ پارامترها در جداول نسخه‌بندی اختصاصی ثبت می‌شود تا هر نشست آزمون دقیقاً به نسخهٔ پارامتر متناظر اشاره کند. پس از کالیبراسیون، برازش آیتمی و شخصی با شاخص‌هایی مانند lzl\_z و HtH\_t ارزیابی می‌شود، تابع اطلاعات آزمون و الگوی SEM برحسب θ\theta بررسی می‌شود و در صورت کفایت داده، عملکرد افتراقی آیتم‌ها (DIF) در گروه‌های زبانی/ملیتی مطالعه خواهد شد. تحلیل زمان پاسخ‌دهی نیز توزیع‌های RT، رابطهٔ RT با دشواری و درستی پاسخ و نیز اثر احتمالی خستگی یا انگیزش را روشن می‌کند. پیوند داده‌های CAT با پاسخ‌های پرسشنامهٔ راهبردها برآوردهای همبستگی و رگرسیون را ممکن می‌سازد و می‌تواند نشان دهد کدام راهبردهای خوداظهاری با عملکرد بهتر یا با ثبات بیشتر θ همبسته‌اند.

از نظر کنترل مواجهه و توازن محتوایی، افزودن یک کد محتوایی به جدول سؤالات پیشنهاد شده و انتخاب آیتم با رویکرد «randomesque» بین چند آیتم با بیشترین اطلاعات، از فرسایش آیتم‌های پرتکرار جلوگیری می‌کند و تنوع محتوایی را افزایش می‌دهد. در سطح تجربهٔ کاربر، قالب‌های راست‌به‌چپ، فونت استاندارد فارسی، کارت‌های سبک و پیام‌های خطای شفاف به کارگرفته شده تا نرخ تکمیل افزایش یابد و دادهٔ پاک‌تری به دست آید. از نظر امنیت و حریم خصوصی، کلید محرمانهٔ برنامه، محدودیت حجم، اعتبارسنجی MIME، و ذخیره‌سازی نسبی فایل‌ها رعایت شده و توصیه می‌شود برای تحلیل‌های بیرون از محیط عملیاتی، شناسه‌های شخصی از داده‌ها جدا و تنها کلیدهای ناشناس در دسترس پژوهش قرار گیرد. همچنین اخذ رضایت آگاهانهٔ شرکت‌کنندگان، اطلاع از کاربری پژوهشی داده‌ها و تعریف دورهٔ نگهداری فایل‌های صوتی از ملاحظات اخلاقی ضروری است.

محدودیت‌های پروژه عمدتاً به اندازهٔ بانک آیتم‌ها و کیفیت کالیبراسیون اولیه وابسته است؛ تا پیش از گردآوری داده‌های میدانیِ کافی، دقت نهایی و تحلیل‌های DIF محدود خواهد بود. SQLite برای نمونه‌های بزرگ توصیه نمی‌شود و در صورت استقرار در مقیاس سازمانی، مهاجرت به سامانهٔ پایگاه‌دادهٔ رابطه‌ای مانند PostgreSQL ضروری است. افزون‌براین، عوامل رفتاری چون خستگی یا تفاوت‌های انگیزشی می‌تواند بر RT و دقت اثر بگذارد و لازم است در تحلیل‌ها کنترل یا دست‌کم گزارش شود.

در جمع‌بندی، سامانهٔ حاضر نمونه‌ای عملیاتی از CAT فارسی بر پایهٔ 3PL است که انتخاب آیتم اطلاعات‌محور، توقف هوشمند مبتنی بر دقت، ثبت فرا‌دادهٔ پاسخ و ابزارهای مکمل مانند پرسشنامهٔ راهبردها و دریافت پاسخ‌های صوتی را در خود ادغام کرده است. این زیرساخت برای فازهای میدانی آماده است و مسیر آیندهٔ پژوهش را هموار می‌کند: کالیبراسیون گستردهٔ آیتم‌ها و نسخه‌بندی پایدار پارامترها، کنترل مواجههٔ پیشرفته و توازن محتوایی خودکار، تحلیل‌های DIF و روایی ملاکی، و در نهایت مهاجرت به زیرساخت‌های مقیاس‌پذیر برای استقرار آموزشی/پژوهشی. انتظار می‌رود در فصل نتایج پایان‌نامه، با داده‌های واقعی، توزیع θ و SEM، آمار توقف، ارتباط با ملاک‌های بیرونی و خلاصهٔ پرسشنامهٔ راهبردها همراه با نمودارهای ICC و تابع اطلاعات آزمون گزارش و تبیین شود تا شواهد روایی سازه و اعتبار عملی سامانه به‌صورت منسجم ارائه گردد.