R

Technical Proposal:

*Request for Proposal 25-78912*

Additional Psychometric Services for ILEARN, I AM, and IREAD-3

|  |  |
| --- | --- |
| Prepared for:  Image result for indiana doe | Prepared by:    Nathan Thompson, PhD  CEO |

June, 2024

Table of Contents

[Introduction 1](#_Toc169098593)

[SOW Responses 1](#_Toc169098594)

[1.1 (3.0.a) Technical Requirements and Key Deliverables 1](#_Toc169098595)

[1.2 (3.0.b) Technical Requirements and Key Deliverables 2](#_Toc169098596)

[1.3 (3.0.c) Technical Requirements and Key Deliverables 3](#_Toc169098597)

[1.4 (3.0.d) Technical Requirements and Key Deliverables 5](#_Toc169098598)

[1.5 (4.0) Timeline of Key Deliverables 6](#_Toc169098599)

[1.6 (5.0) Communication and Scheduled Meetings 6](#_Toc169098600)

[1.7 (6.0) Staff Qualifications 7](#_Toc169098601)

Introduction

The Indiana Department of Education (IDOE) delivers assessments in schools throughout the state, including the Indiana Learning Evaluation Assessment Readiness Network (ILEARN). These assessments are professionally developed and delivered by a third party (Cambium). Assessment Systems Corporation is pleased to submit this proposal to provide quality assurance services on these, namely the replication of psychometric analyses and statistical reporting. This document suggests a workflow to accomplish IDOE’s goals, per Attachment K.

**Section 1: Scope of Work Questions**

|  |  |
| --- | --- |
| **Question #** | **Response Page #** |
| **1.1** | **1** |
| **1.2** | **2** |
| **1.3** | **3** |
| **1.4** | **5** |
| **1.5** | **6** |
| **1.6** | **6** |
| **1.7** | **7** |

SOW Responses

This section provides narrative responses to the points outlined in Attachment F: Technical Proposal.

1.1 (3.0.a) Technical Requirements and Key Deliverables

*The Respondent must provide a description of the process/feedback loop to ensure timely reporting of any scoring issues.*

There are two aspects of the feedback loop: example data and production data.

For the example data, we recommend that IDOE/CAI provide the raw data and intended results, so that we can engineer the processes and software to replicate. We will have a kickoff meeting to discuss, and occasional meetings during these months of development to clarify methodologies and discuss any discrepancies in results that are found. We also recommend email for much of this, since it is sometimes easier to provide lengthy tables that take a half hour to digest, and not feasible to discuss in real time until after everyone has had a chance to digest. By having the intended results, we will have immediate feedback during the development phase of the project.

For the production data, the RFP recommends weekly status meetings to ensure timely progress. We will meet with IDOE in this schedule, providing written weekly progress reports and noting any discrepancies to discuss. This will ensure timely dealing of any issues.

There will also likely be a meeting after the critical time period, as a retrospective to that period and to discuss any requirements for the summary reports to be provided later.

ASC places great emphasis on data and IT security standards. We have achieved SOC2 Type II and FedRAMP accreditation audits for our online assessment platform. This necessitates extensive documentation and training of employees.

1.2 (3.0.b) Technical Requirements and Key Deliverables

*The Respondent must provide a description of the process of data cleaning.*

We recommend the following steps. These would be first performed based on example files before the tight schedule with actual student data, so that we are sure that we can recreate the rules used by Cambium.

**Step 1: Receive Database Files from Vendor and Format Input Files**

The first step of work is to download the large files provided by Cambium and prepare them for our software. More specifically, this stage includes work to extract relevant data from the large Config files and format it in data structures that are more easily accessible. Also, specifications are set for the primary algorithms regarding input, such as determining which column in a 1000-column file contains the student response. We intend to use the secure file transfer system employed by Cambium.

**Step 2: Parse Database Files**

In the previous project, Cambium delivered data in database extracts with the “stacked” format rather than matrix format typically used by psychometricians, as this provides more data than the matrix method, such as item sequence. The first algorithms of our software focus on extracting all the necessary information and storing it in data structures that are more easily accessible for subsequent analysis. The first is the commonly used person (row) x item (column) matrix of student responses. Additional data structures contain item sequence data, item metadata and parameter input, and test blueprints. This stage requires much of the software development time and software run time.

**Step 3: Codify Data Cleaning Rules**

We will work with IDOE and Cambium to establish a formal list of data cleaning rules, perhaps using pseudocode to get as close as possible to the enactment of the rules. We will also establish the dependent variables that serve as the benchmarks for confirmation.

**Step 4: Implement Data Cleaning**

We will implement the rules from Step 3 onto the simplified files from Step 2, typically to exclude students based on some criteria such as not attempting a sufficient number of items. This will be done programmatically, using Python, R, or Pascal.

**Step 5: Confirmation**

We will compare our results on the established dependent variables (e.g., N counts per school) to the results from Cambium, and investigate any discrepancies. Once confirmed, we will provide a brief report to IDOE for the results of this stage, serving as the key deliverable.

1.3 (3.0.c) Technical Requirements and Key Deliverables

The Respondent must provide a description of the process to meet and address the requirements of the replication of psychometric analyses as shared in Section 3.0 Technical Requirements and Key Deliverables from the Scope of Work document (Att K) for this RFP.

**Establishing Specifications**

As noted in Attachment K, the RFP does not include details about the psychometric analyses, making it difficult to describe how we would perform them. So the first step in this project is to review the specifications, which IDOE states will be shared in the kick-off meeting. We will evaluate the specifications and recommend approaches to calculating each of the indices, such as commercial vs. custom software.

**Classical Analysis**

We will take the cleaned data sets and run them through our item analysis software. This produces all the common indices with classical test theory. Examples:

* Test level (and subscore level)
  + Student scores
  + N of students in each classification, if used
  + Mean score
  + SD of scores
  + Internal consistency reliability (alpha/KR-20)
  + Standard error of measurement
* Item level
  + P (proportion-correct, a difficulty index)
  + Rpbis (point-biserial correlation, a discrimination index)
  + N responded vs. omit
* Option/Distractor level
  + P
  + Rpbis
  + N

These statistics do not have alternative methodologies; there is only one way to calculate each of these.

We will flag any statistics that do not agree with CAI, or are outside the acceptable criteria specified by DOE, such as Rpbis<0.1. We will attempt to resolve any discrepancies with CAI via email and/or weekly meeetings.

**IRT Analysis**

ASC will run an IRT analysis if IRT parameters remain to be calibrated; if all items have been field tested and calibrated before the actual testing window and CAI is not performing an IRT analysis on this data, we will not either.

We typically use marginal maximum likelihood estimation for item parameters. For dichotomous models, this includes 1PL/Rach, 2PL, and 3PL, where we estimate the a, b, and c parameters. We can also calibrate polytomous models such as the generalized partial credit model. If CAI uses proprietary models (see below) we are not able to reproduce those for calibration.

Some of the recommended results for comparison

* Test level
  + Student scores (theta and SEM)
  + Mean theta
  + SD of theta
* Item level
  + *a*
  + *b*
  + *c*
  + Polytomous thresholds (e.g., category *b* with GPCM)

An example of our software is shown below.

A screenshot of a computer

Description automatically generated

**Science Performance Tasks**

We will analyze the Science performance tasks with the raw data, and with the all correct/incorrect transformation as defined by IDOE. Though the description provided is unclear, as it suggests that a score of 0.5 is possible but IRT models such as the partial credit model assume whole numbers as observable scores.

**Equating**

The RFP states that we should also replicate equating but does not provide information about the approach that is used. We assume that there are anchor items from previous years that have fixed item parameters. We can equate using fixed-anchor calibration methods, or with post hoc conversion methods such as Stocking-Lord or Haebera. Ideally, all items will be “anchors” in that they have existing IRT parameters from previous usage and field testing. If we find that equating might be a potential source of disagreement between our results and CAI, we will investigate and report on it.

**Comparisons and documentation**

ASC will provide a formal report as a key deliverable for this portion of the project, for each of the points listed above. This will list our result and CAI’s result for each of the benchmark indices, preferably as a spreadsheet. We will also provide a narrative summary covering these topics:

* Data cleaning
* Psychometric methodologies
* Comparisons of CTT results
* Comparisons of IRT results
* Comparisons of scoring
* Discrepancies, and their resolution or potential causes

**Software for the Project**

The RFP stated that we should use the same software as Cambium if feasible, but did not state what this is. We recommend the use of ASC’s software Xcalibre for this work. In the past IDOE project, Xcalibre’s theta estimations (IRT scoring) were closely aligned with the production results. Xcalibre also provides both classical and IRT analysis, including item difficulty, item discrimination, reliability, IRT parameters, and standard error functions. If we are required to use the same software as Cambium, such as IRTPRO or FlexMIRT, that would be an additional expense in the contract; we do not know the price but it would be charged at-cost to IDOE.

A related concern is that in the past IDOE project, Cambium used a proprietary scoring model; if recalled correctly, it was an expanded Testlet Response Theory model, which extended IRT to new dimensions. This was extremely difficult to replicate with the sparse documentation that was provided. If we are to replicate the calibration of this, Cambium must provide a copy of their internal software. If no calibration is needed, we will reproduce the scoring with this model utilizing the documentation provided by Cambium.

1.4 (3.0.d) Technical Requirements and Key Deliverables

*The Respondent must provide a description of the process to meet and address the requirements of the replication of reported student scores at multiple levels as shared in Section 3.0 Technical Requirements and Key Deliverables from the Scope of Work document for this RFP*

This is the process we would use to replicated scores.

1. Raw scores

We would calculate raw scores as part of the psychometric analysis. For classically scored tests, this is typically number-correct, while IRT scored tests will include theta estimation. For IRT we utilize maximum likelihood estimation scoring with standard IRT models shown above.

If there are subscores, we will calculate those too. These will be confirmed against CAI results before proceeding.

2. Scaled scores

We would implement scaled scoring rules as defined by IDOE. This is typically a normative or linear transformation, with min/max limits. Our item analysis software is also able to implement this automatically. An example screen from our software is shown below.

A screenshot of a computer

Description automatically generated

3. Performance levels

If a test assigns examinees into performance levels based on cutscores, we will assign the students. This is typically done with the scaled scores.

4. Aggregate reporting

IDOE likely calculates aggregate values of the previous results, such as the number of students in each performance level per school/corporation, or average scaled score. We will calculate such descriptive statistics.

Later in the year, we will also confirm summary statistics tables in final technical reports, as noted in Section 4.0. Finally, we will provide a summary report of all our activities at the end of the year.

1.5 (4.0) Timeline of Key Deliverables

*The Respondent must provide their proposed timeline for the data replication work, and provide an explanation for any deviations from the IDOE timeline provided in Section 4.0 Timeline of Key Deliverables from the Scope of Work document for this RFP. The Respondent should also refer to the tables for each assessment found in Section 3.0 for additional information/key dates.*

Given the tight schedule in Attachment K, the most critical portion of this project will be to establish sample data sets with CAI and IDOE well ahead of time, and use those to develop our processes, and if necessary, custom software. For example, ILEARN 2025 Spring Grades 3-8 has only May 14-17 for all the activities describe in this report. This cannot be met manually. So, while we have existing software for psychometric analysis, we will need months of time to develop the processes and software for data cleaning and custom scoring. There remain some open questions, such as what is involved with the linking and equating and how much of that is put on ASC to calculate.

The suggested timeline in Attachment K sets the handoff of all documentation and sample files as Feb 20, leaving three months to establish the processes and software. We will use this time to replicate the sample files and confirm the results with IDOE. This will allow us to move quickly during May 14-17. A similar approach applies to other assessments and the final replications.

We do not have objections to the timelines in the RFP, such as the one below.

**ILEARN Timeline for Spring 2025**

|  |  |  |
| --- | --- | --- |
| **Task** | **Begin** | **End** |
| Kickoff for ILEARN Spring 3-8 | 2/13/2025 | 2/13/2025 |
| Handoff of Test Design and Reporting Documentation, Sample Files | 2/20/2025 | 2/22/2025 |
| Handoff of Data Sets | 5/9/2025 | 5/13/2025 |
| Replication Activities | 5/14/2025 | 5/17/2025 |
| Initial Sign Off | 5/20/2025 | 5/20/2025 |
| Delivery of Final Data Files | 6/16/2025 | 6/20/2025 |
| QC of Final Data Files | 6/23/2025 | 6/29/2025 |

Based on the timelines provided, here are suggested key deliverables and dates.

|  |  |  |
| --- | --- | --- |
| **Project portion** | **Deliverables** | **Date** |
| Project launch | * Kickoff meeting * Set up file transfer access * Receive documentation and sample files | 2/22/2025 |
| Development phase | * Spreadsheet comparing ASC vs CAI results; a draft of what final comparison spreadsheet would look like | 4/10/2025 |
| IREAD-3 Spring | * Spreadsheet of replicated results and comparison to CAI * Report with description of process, summary of discrepancies, and narrative explanations | 4/22/2025 |
| I AM Spring | * Spreadsheet of replicated results and comparison to CAI * Report with description of process, summary of discrepancies, and narrative explanations | 6/13/2025 |
| ILEARN Spring | * Spreadsheet of replicated results and comparison to CAI * Report with description of process, summary of discrepancies, and narrative explanations | 6/29/2025 |
| IREAD-3 Summer | * Spreadsheet of replicated results and comparison to CAI * Report with description of process, summary of discrepancies, and narrative explanations | 7/22/2025 |
| Confirmation of tables in technical reports | * Summary report comparing tables in IDOE report to our results | Nov 2025 |
| Final summary report | * Summary report comparing tables in IDOE report to our results | Dec 2025 |
|  |  |  |

1.6 (5.0) Communication and Scheduled Meetings

*The Respondent must provide a description regarding how it will meet and address the requirements shared in Section 5.0 Communication and Scheduled Meetings from the Scope of Work document for this RFP.*

We will schedule a kickoff meeting before any of the timelines in Attachment K begin. This will provide an opportunity for introductions, overview of work, setting expectations and roles, and discussing specifics. This will be helpful for us to begin the development of processes and software as early as possible. We will provide an agenda beforehand, and notes afterwards.

During the setup timeline, we will have periodic meetings with IDOE and CAI to discuss specifics, such as how equating is done so that we can accurately replicate it.

During the critical period of data coming in and being replicated, we will have weekly status calls with IDOE and CAI, perhaps daily if warranted. Here, we will provide update on our progress as well as discuss issues such as unclear specifications or disagreement with CAI results. This will also be summarized in written progress reports as part of the meeting minutes.

We can utilize Microsoft Teams if IDOE or CAI sets up the calls. Typically, we use Zoom or Google Meet, which can be accessed via browser.

We will provide summary reports after the critical work period, and meet with IDOE at that time to ensure that they meet the required level of documentation for validity evidence. This includes documentation of our processes, results, and summary information.

1.7 (6.0) Staff Qualifications

*The Respondent must provide a description regarding how it will meet and address the expectations of staff experiences in psychometrics and data replication shared in Section 6.0 Staff Qualifications from the Scope of Work document for this RFP.*

ASC has an extensive team of assessment experts, including five with PhDs and five more with Masters degrees. For this project, we recommend the personnel below. Belinda Brunner would be the project manager and primary contact. Dr. Tammy Trierweiler would lead all work, while Dr. Natasha Wilson and Dr. Fernando Austria Corrales serving as key staff. Nathan Thompson would provide advisory services based on his previous experience with IDOE, and also provide the source code from that project which would start us ahead on this project.

**A person smiling at the camera

Description automatically generatedNathan Thompson, PhD – CEO**

Nathan Thompson earned his PhD in Psychometrics from the University of Minnesota, with a focus on computerized adaptive testing. His undergraduate degree was from Luther College with a triple major of Mathematics, Psychology, and Latin. He is primarily interested in the use of AI and software automation to augment and replace the work done by psychometricians, which has provided extensive experience in software design and programming. Dr. Thompson has published over 100 journal articles and conference presentations. His experience includes roles as a psychometrician, test development manager, item writer, essay scorer, product officer, and executive. He’s worked with more than 100 assessment organizations to improve their assessments, including national groups like the Singapore Examinations and Assessments Board, Nepal Exams Board, UAE Ministry of Education, and national civil service exams in Colombia.

**A person with blonde hair smiling

Description automatically generatedTammy Trierweiler, PhD - VP of Psychometrics**

Dr. Trierweiler is an expert in Psychology, Quantitative methods, Psychometrics, and Computer Science, simplifies complex problems with data-driven or human-centered solutions. With 16+ years' experience, she excels in assessment design, data analysis, and clear communication. Skilled in statistical methodologies like IRT/CTT Models and computational methods like AI/NLP, Tammy is your go-to problem solver. With a PhD in psychometrics from Fordham University and a Master’s in computer science from the University of Pennsylvania, she offers expertise in statistical analysis software such as SAS, SPSS, R, and Python.

A close-up of a person smiling

Description automatically generated**Natasha Wilson, PhD – Test Development Manager**

A methodical, reliable researcher, supportive leader and team player with 12 years of total working experience, with 5 years experience in process engineering and manufacturing, 3 years of nonprofit experience, and 4 years of work in education. Prior to ASC, Dr. Wilson worked as director of test development for a university admissions test. At ASC, she primarily works in data engineering to clean/manage data files and move them through the report generation process. Holds a PhD in Chemical and Biochemical Engineering from the University of Maryland, Baltimore County.

A person with short brown hair wearing a necklace

Description automatically generated

**Belinda Brunner, MS - Digital Assessment Project Manager**

Belinda has two decades of experience in managing the assessment process. This includes leadership of the certification exams at three nonprofit boards and a decade at Pearson VUE where she helped dozens of organizations to improve their assessments.

A person wearing glasses and smiling

Description automatically generated**Fernando Austria Corrales, PhD – Psychometrician**

Fernando has a PhD in educational assessment and more than a decade of experience in test development and psychometrics. He served as a team lead at the National Institute of Educational Assessment in Mexico, lead psychometrician for the State of Aguascalientes department of education, and now as a psychometrician for ASC. In all three roles, he has worked in test development, CTT/IRT analysis, and scoring. He has extensive experience in R and Python.