Standard Setting Suite: User Documentation

1. Introduction & Purpose

The Standard Setting Suite is an all-in-one, browser-based application designed for faculty and assessment committees at the College of Pharmacy, Gulf Medical University. Its primary purpose is to streamline and automate three common, evidence-based methods for setting defensible cut-scores (passing marks) for student examinations.

This suite consolidates the following three distinct standard-setting tools into a single, user-friendly interface:

1. **EZ Method**
2. **Simple Angoff Method**
3. **Modified Cohen Method**

By providing instant calculations and visual feedback (where applicable), this application removes the need for complex manual spreadsheets, reduces the chance of calculation errors, and allows committees to focus on the judgmental and deliberative aspects of standard setting.

2. Target Audience

This application is intended for:

- Faculty members
- Assessment coordinators
- Any institutional body responsible for establishing or validating the passing standards for student assessments.

3. How to Use the Suite

The suite is designed to be intuitive and requires no installation. It runs entirely within a modern web browser (like Chrome, Firefox, Edge, or Safari).

3.1. The Home Page

When you first open the application, you are greeted with the "Standard Setting Suite" home page. This page serves as the central navigation hub. It displays three interactive cards, one for each of the available methods.

3.2. Selecting a Tool

To use a specific method, simply click on its corresponding card:

- **EZ Method**
- **Simple Angoff Method**
- **Modified Cohen Method**

3.3. Using a Tool

After clicking a card, you will be taken to the dedicated calculator page for that method. Each page contains a form where you will input the required data (e.g., judge ratings, student scores, exam parameters).

3.4. Returning to the Home Page

On each of the three tool pages, a **"Back to Home"** button is located at the top. Clicking this button will return you to the main home page at any time, allowing you to easily switch between methods.

4. Detailed Tool Descriptions

Below is a detailed breakdown of each method, its underlying principle, the data you need to provide, and the results it will calculate.

4.1. EZ Method

- **Purpose**: The EZ Method is a "holistic" standard-setting method. The equal Z method (henceforth: EZ method, pronounced “easy method”) aimed to generate cut scores that are placed between the average minimum passing score and the averaged maximum failing score for the en- tire examination as determined by a panel of experts.

- **Methodology:** Judges estimate a cut-score based on their overall impression of the exam's difficulty and the expected performance of a borderline student. Each expert separately provides answers to the following 2 questions: first, what would be the lowest score that indicates, without any doubt, that an examinee is competent in the topics assessed?; second, What would be the highest score that indicates, without any doubt, that an examinee is incompetent in the topics assessed?. So for each station, we define L as the highest failing score below which an examinee is incompetent; and we define H as the lowest passing score above which an examinee is competent. From the collated scores (L and H), the means of L and H (XL and XH, re-spectively) and standard errors of the means (SEL and SEH, re- spectively) are calculated.
- The following equation is used to identify the same Z score (Z) that would apply to both confidence intervals of XL and XH when they interface:
  Z = (XH‑XL)/(SEL+SEH)
- The cut score is then set at XL+Z*SEL, which is also equal to XH Z*SEH.
- **Required Inputs:**
  - **Institutional Cut Score (E.g., 65%)**
  - **Number of Judges:** The total number of faculty participating.
  - **Total Exam mark** (E.g. 40 marks)
  - **Number of OSCE/OSPE Stations/Experiments**
  - H: the lowest passing score above which an examinee is competent (e.g., 75).
  - L: the highest failing score below which an examinee is incompetent (e.g., 50).
- **Calculated Outputs:**
  - **Calculated Cut-Score:** The final recommended passing mark, determined by the graph intersection.

- **Standard Deviation:** A measure of agreement (or disagreement) among the judges' estimates.

Reference: Yang, Y.-Y., Shulruf, B., Huang, P.-H., et al. (2022). *Journal of Educational Evaluation for Health Professions*, 19, 27.

4.2. Simple Angoff Method

- **Purpose:** The Angoff method is a widely-used, "analytical" or item-by-item method. It is considered highly defensible as it requires judges to review every single question on the exam.
- **Methodology:** Judges (faculty) examine each individual test item and estimate the probability (from 0.0 to 1.0) that a "minimally competent" or "borderline" student would answer that specific item correctly. The final cut-score is the sum of these probabilities, averaged across all judges.
- **Required Inputs:**
  - **Number of Judges:** The total number of faculty participating.
  - **Number of Items:** The total number of questions on the exam.
  - **Judge-Item Probability Table:** The tool will dynamically generate a table (matrix) with a row for each judge and a column for each item. You must enter each judge's probability estimate (a value between 0.0 and 1.0) for every item in this table.

- **Calculated Outputs:**
  - **Overall Recommended Cut-Score:** The final passing mark, calculated as the average of all judges' individual cut-scores.
  - **Standard Deviation of Judges' Scores:** A measure of inter-rater reliability. A low SD indicates high agreement among judges.

- **References**:
  - **Angoff, W. H. (1971)**. Scales, norms, and equivalent scores. *Educational Measurement: Issues and Practice*, 1(1), 35-45.
  - **Ricker, K. L. (2006).** Setting cut-scores: A critical review of the Angoff and Modified Angoff methods. *Alberta Journal of Educational Research*, 52(1), 53-64.

4.3. Modified Cohen Method

- **Purpose:** The Cohen method is a "test-centered" or "hybrid" method. It is not used to create a standard from scratch, but rather to *adjust* a pre-existing or traditional cut-score (e.g., 65%) based on the actual difficulty of a *specific* exam administration.
- **Methodology:** This method finds a compromise between the traditional pass mark (e.g., 65%) and a standard derived from the actual performance of the *best* students on that exam (The score of the student at the top 10%). It effectively lowers the cut-score for unusually difficult exams and raises it for unusually easy ones, making it fairer for all student cohorts. Adjusted Cut Score is calculated using the following formula: PM = 0:65 * P90
- **Required Inputs:**
  - **Number of Examinees:** The total number of students who took the exam.
  - **Institutional Cut-Score (%):** The default pass mark your college normally uses (e.g., 65).
  - **Total Exam Points**
  - **Total Exam Score**
  - **All Student Scores (comma-separated):** A complete list of *every* student's score on the exam, separated by commas (e.g., 72, 58, 91, 64, ...).

- **Calculated Outputs:**
    - **90 Percentile (P90): The score of the student at the top 10%.**
    - **Adjusted Cut-Score:** The final, adjusted passing mark calculated using the Cohen formula. This is the "compromise" score.

Reference: Celia A. Taylor (2011) Development of a modified Cohen method of standard setting, Medical Teacher, 33:12, e678-e682, DOI: 10.3109/0142159X.2011.611192

5. Technical Specifications

- **Format:** Single HTML file.
- **Core Logic:** Vanilla JavaScript.
- **Styling:** Tailwind CSS.
- **Visualizations:** Chart.js library.
- **Environment:** Runs 100% in the browser (client-side). No internet connection is required after the page is loaded. No data is ever transmitted to a server.
- **Concept & Logic:** Dr. Muhammad AlShorbagy, Dean, College of Pharmacy, GMU.
- **Technical Implementation:** AI-Assisted Development (Code generation).
- **Methodology:** "This single-file HTML application demonstrates a 'No-Code/Low-Code' development approach. The domain expertise, algorithm logic, and user experience design were provided by Dr. Muhammad AlShorbagy, while the source code was generated via prompt engineering using Large Language Models (LLMs)."