Psychometrika (2026), **91**, 279–311
doi:10.1017/psy.2025.10056

PSYCHOMETRIC SOCIETY

THEORY AND METHODS

Bayesian Selection Policies for Human-in-the-Loop Anomaly Detectors with Applications in Test Security

Michael Fauss[1,2], Xiang Liu[1], Chen Li[1], Ikkyu Choi[1] and H. Vincent Poor[2]

[1]ETS Research Institute, USA; [2]Department of Electrical and Computer Engineering, Princeton University, USA

Corresponding author: Michael Fauss; Email: mfauss@ets.org

(Received 4 November 2024; revised 23 June 2025; accepted 29 September 2025)

Abstract

This article investigates the problem of automatically flagging test takers who exhibit atypical responses or behaviors for further review by human experts. The objective is to develop a selection policy that maximizes the expected number of test takers correctly identified as warranting additional scrutiny while maintaining a manageable volume of reviews per test administration. The selection procedure should learn from the outcomes of the expert reviews. Since typically only a fraction of test takers are reviewed, this leads to a semi-supervised learning problem. The latter is formalized in a Bayesian setting, and the corresponding optimal selection policy is derived. Since calculating the policy and the underlying posterior distributions is computationally infeasible, a variational approximation and three heuristic selection policies are proposed. These policies are informed by properties of the optimal policy and correspond to different exploration/exploitation trade-offs. The performance of the approximate policies is assessed via numerical experiments using both synthetic and real-world data and is compared with procedures based on off-the-shelf algorithms as well as theoretical performance bounds.

Keywords: anomaly detection; human in the loop; test security; variational Bayes

1. Introduction

A common task when administering and evaluating tests is to identify behaviors or responses that are atypical and may require special attention to ensure the integrity of the test (Bulut et al., 2024; Cizek & Wollack, 2016; He et al., 2022; Kingston & Clark, 2014; Wongvorachan, 2023). For example, in writing tests, some responses might need closer inspection to establish if they were plagiarized (Foltýnek et al., 2019; Gomaa & Fahmy, 2013) or, more recently, AI-generated (Jiang et al., 2024; Yan et al., 2023). In language tests, test takers might be reading off scripts or repeating after a hidden *souffleur* instead of speaking freely (Evanini & Wang, 2014; Wang et al., 2016, 2019). Similar tasks also occur outside a fraud detection context. For example, one might want to identify test takers that did not properly engage with the items (Booth et al., 2023; Cocea & Weibelzahl, 2010), that should have been provided with certain accommodations (Kettler, 2012; Sireci et al., 2005), or are not using the provided equipment as intended (Taylor et al., 1998).

In many such scenarios, the respective test takers or responses are first *flagged* by automated systems. For example, given the large amount of potential source material, initial plagiarism checks are almost

© Educational Testing Service and the Author(s), 2025. Published by Cambridge University Press on behalf of Psychometric Society.
This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

exclusively done via automated systems (Jiffriya et al., 2021). However, especially in high-stakes testing, relying on automated flags alone is risky because of inevitable false positives. Therefore, once a test taker has been flagged by an automated system, their case is typically reviewed in more detail before a final decision is made. The subject of this article is the design of an automated system that flags test takers for review and learns from their outcomes. Since in a testing context such reviews are typically carried out by one or more human experts, we refer to this scenario as *human-in-the-loop* anomaly detection. However, the proposed procedure is, in principle, agnostic to how the reviews are conducted.

Before going into more detail, it is useful to fix some terms. In what follows, we refer to the group of test takers that we seek to identify as *critical group*. All test takers that are not members of the critical group are referred to as the *reference group*. Moreover, we refer to test takers that are identified by the automated system as *flagged* or *selected* for review. If the review outcome is positive, that is, the test taker is indeed found to be a member of the critical group, we refer to the corresponding flag as a *detection* or *true positive*. If the review outcome is negative, that is, the test taker is found *not* to be a member of the critical group, we refer to the flag as a *false alarm* or *false positive*. The rule by which test takers are selected for review is referred to as a *flagging* or *selection policy*. Finally, depending on the context, we will say that a *test taker* is selected for review or that a *response* is selected for review. For the purpose of this article, this difference is insubstantial.

We assume that a test is administered periodically over time. The flagging procedure we aim to design is assumed to work as follows: for each test taker in the current administration, the automated system is provided with a set of features that were extracted from their response and/or behavior. The system processes these features and outputs a binary flagging decision. The goal is to find a policy that maximizes the number of test takers correctly identified as needing special attention (true positives), while keeping the number of required reviews per test administration at a manageable level. The latter is important since, depending on what it entails, an expert review can require a significant amount of resources. The features based on which the flagging decision is made are not subject to this optimization, but are assumed to be given. We refer to this problem as the *flagging problem*. It will be defined more formally in Section 2.

Although the methods proposed in this article potentially cover a wide range of applications, we envision them to be most useful in settings where the outcomes of multiple detectors or classifiers need to be fused (Varshney, 2012). For example, returning to the task of plagiarism detection, instead of relying on a single checker, multiple checkers might be run on a submitted response, each of them returning a value that quantifies how likely the response was plagiarized. A decision whether or not to flag the response then needs to be made by fusing the results of the individual checkers into a single, binary outcome.

In order to position this article in the context of existing research, it is useful to discuss some of the characteristics and requirements of the system we seek to design:

- We do not assume the existence of a (large) training data set. This assumption is somewhat pessimistic, but often realistic. For example, when new item types are introduced or new ways of cheating emerge, little to no data will be available to train detectors.
- We assume that a test taker's group membership can be established via an expert review. However, typically only a fraction of test takers will undergo a review, thus rendering the flagging problem a *semi-supervised* learning problem (Van Engelen & Hoos, 2020).
- In contrast to many semi-supervised learning problems, we do not assume the subset of labeled data points to be given in advance. Instead, the selection policy itself determines which data points are reviewed and labeled; in a sense, the system can and must select its own training data—a concept commonly found in *active learning* (Felder & Brent, 2009).
- We consider data fusion the main use case of the system we seek to design. Hence, it should be able to handle "reasonably high-dimensional" feature spaces, say, on the order of tens or hundreds of features, but is not necessarily expected to process very large feature vectors commonly used in supervised machine learning (Caruana et al., 2008).

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

- We seek to avoid strong assumptions on the number of test takers or the number of reviews conducted. That is, the system should neither require a minimum sample size, nor should compute or memory requirements limit its application to small-scale tests.
- In order to learn and track the feature distributions of interest, the system should be adaptive and operate in a continuous feedback loop. That is, the outcomes of expert reviews are fed back into the system, which then uses this feedback to update its parameters and in turn improve the accuracy of its selection policy.
- Especially in a high-stakes testing context, it is crucial that important metrics of the automated selection system, such as its false positive and false negative rate, can be evaluated or at least estimated. In the scenario considered here, this is not straightforward since only selected test takers (positives) are reviewed. Hence, metrics, such as the false negative rate, cannot be evaluated empirically. This limits the applicability of commonly used classifiers whose outputs can be interpreted as probabilities, but are not based on probabilistic models (Bielza & Larrañaga, 2020).

In light of these requirements, we propose a Bayesian approach to the flagging problem. In particular, as will be shown in the course of the article, a Bayesian approach allows for a unified treatment of labeled and unlabeled data points, and in turn enables a seamless implementation of a feedback loop. Moreover, prior distributions make it possible to incorporate qualitative prior knowledge in cases where little or no training data is available. As more samples are collected, this prior knowledge is gradually overwritten by evidence learned from the data. A disadvantage of the Bayesian approach is that both the inference step, that is, the update of the posterior distribution, and the optimal selection policy are too complex to be implemented in practice. We address this problem by, first, using a variational approximation of the posterior distribution, and, second, proposing three heuristic selection policies corresponding to different exploration/exploitation trade-offs.

Naturally, Bayesian approaches have been used in the context of test integrity before (see, for example, Lu et al., 2023; Marianti et al., 2014; Van der Linden & Guo, 2008; van der Linden & Lewis, 2015; Zhang et al., 2022 to name just a few). Moreover, some Bayesian methods have been proposed for general semi-supervised learning problems (Adams & Ghahramani, 2009; Bruce, 2001; Rottmann et al., 2018). However, the specific scenario considered in this article, learning sequentially from self-selected subsets of batched data, is non-standard and non-trivial. To the best of our knowledge, it has not been investigated in detail before, and methods in the literature did not meet all of the requirements listed above.

Finally, we would like to highlight that, although they were motivated by a test integrity scenario, both the problem formulation and the inference and flagging procedures presented in this article are generic and extend beyond this specific application. In principle, the proposed methods apply whenever noisy measurements are used to select individuals or objects for expert inspection. This is the case, for example, in industrial quality control, where sensor data is used to monitor and flag defective products for manual inspection (Mitra, 2016), in condition-based maintenance, where predictive models are used to determine when machinery needs servicing based on wear and usage data (Prajapati et al., 2012), and in cybersecurity, where anomalous network activity might be flagged for further review by security experts (Ahmed et al., 2016).

The remainder of the article is organized as follows: In Section 2, we introduce our notation and discuss the problem formulation and the underlying assumptions in a more formal manner. The corresponding optimal selection policy is presented and discussed in Section 3. In Section 4, a variational approximation of the true posterior distribution is presented and three heuristic selection policies are proposed that are informed by properties of the optimal policy and correspond to different exploration/exploitation trade-offs. Two theoretical performance bounds are stated in Section 5. In Section 6, numerical examples are given that demonstrate the proposed procedure and compare its performance to that of off-the-shelf algorithms. Section 7 concludes the article and provides a brief outlook on open questions and possible extensions.

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

2. Problem formulation and assumptions

In this section, we introduce our notation, state our assumptions, discuss the underlying information flow, and give a formal definition of the flagging problem.

2.1. Notation

Random variables are denoted by uppercase letters, X, and their realizations by the corresponding lowercase letters, x. Analogously, uppercase letters, P, denote probability distributions and the corresponding lowercase letters, p, denote probability density functions (PDFs). Occasionally, subscripts are used to indicate a distribution or density of a certain random variable, say, P_X and p_X. We use Q_X and q_X to indicate that the corresponding distribution or density is closely related but not identical to P_X or p_X, respectively. For example, q_X might be an approximation or unnormalized version of p_X. The expected value of a random variable is written as $E[X]$. Again, a subscript is used to explicitly indicate the random variable, say, $E_X[c+X]$. Equality in distribution is denoted by $\overset{d}{=}$. Unless stated otherwise, collections of variables, such as vectors and matrices, are indicated by bold font. We write $\mathcal{A}^n$ to denote the n-fold Cartesian product of a set $\mathcal{A}$ with itself. The n-dimensional probability simplex is denoted by Δ^n. The set of Boolean vectors of length N whose elements sum to K is denoted by $\mathcal{S}_K^N$. The set of all positive semi-definite matrices of size M is denoted by $\mathbb{S}_+^M$. We further define the elliptope of correlation matrices as $\mathcal{E}_M \coloneqq \left\{\boldsymbol{\Sigma} \in \mathbb{S}_+^M : \operatorname{diag}(\boldsymbol{\Sigma}) = \boldsymbol{I}_M\right\}$, and the projection of a matrix $\boldsymbol{\Sigma} \in \mathbb{S}_+^M$ on $\mathcal{E}_M$ as $\mathcal{P}_{\mathcal{E}}(\boldsymbol{\Sigma}) \coloneqq \operatorname{diag}(\boldsymbol{\Sigma})^{-\frac{1}{2}}\,\boldsymbol{\Sigma}\,\operatorname{diag}(\boldsymbol{\Sigma})^{-\frac{1}{2}}$. Here, $\boldsymbol{I}_M$ denotes the identity matrix of size M and $\operatorname{diag}(\boldsymbol{X})$ denotes the diagonal matrix with the same main diagonal elements as $\boldsymbol{X}$. Additional symbols and notations will be defined when they occur in the text.

We assume that a test is administered periodically at time instances $t = 1, 2, \ldots, T$. Each administration is assumed to have $N \geq 1$ test takers.[1] An unknown share of test takers is assumed to belong to the critical group whose members we seek to flag. In order to formalize the flagging problem, we introduce the following random variables:

- $R \in [0,1]$ denotes the rate of critical group members, that is, the probability that a randomly selected test taker is a member of the critical group. R is assumed to be a latent variable.
- $C_{t,n} \in \{0,1\}$ indicates whether the nth test taker in the tth administration is a member of the critical group ($C_{t,n} = 1$) or not ($C_{t,n} = 0$). $C_{t,n}$ is assumed to be a latent variable.
- $\boldsymbol{X}_{t,n} \in \mathbb{R}^M$, $M \geq 1$, denotes the vector of features associated with the nth test taker in the tth administration. $X_{t,n}$ is assumed to be observable and provided by a given mechanism.
- $S_{t,n} \in \{0,1\}$ indicates whether the nth test taker in the tth administration was selected for review ($S_{t,n} = 1$) or not ($S_{t,n} = 0$). $S_{t,n}$ is set according to the chosen selection policy.
- $D_{t,n} \in \{0,1\}$ denotes whether the nth test taker of the tth administration was detected as being a member of the critical group ($D_{t,n} = 1$) or not ($D_{t,n} = 0$). $D_{t,n}$ is the outcome of an expert review and can be observed.

For the sake of a more compact notation, variables corresponding to the same administration are also written as column vectors:

$$\boldsymbol{X}_t \coloneqq \left[\boldsymbol{X}_{t,1}^\top\ \boldsymbol{X}_{t,2}^\top\ \ldots, \boldsymbol{X}_{t,N}^\top\right]^\top \in \mathbb{R}^{MN}, \qquad \boldsymbol{C}_t \coloneqq \left[C_{t,1}\ C_{t,2}\ \ldots, C_{t,N}\right]^\top \in \{0,1\}^N,$$
$$\boldsymbol{S}_t \coloneqq \left[S_{t,1}\ S_{t,2}\ \ldots, S_{t,N}\right]^\top \in \{0,1\}^N, \qquad \boldsymbol{D}_t \coloneqq \left[D_{t,1}\ D_{t,2}\ \ldots, D_{t,N}\right]^\top \in \{0,1\}^N.$$

Analogously, horizontal stacks of vectors of the first t administrations are written as

$$\boldsymbol{X}_{1:t} \coloneqq \left[\boldsymbol{X}_1\ \boldsymbol{X}_2\ \ldots, \boldsymbol{X}_t\right] \in \mathbb{R}^{t \times MN}, \qquad \boldsymbol{C}_{1:t} \coloneqq \left[\boldsymbol{C}_1\ \boldsymbol{C}_2\ \ldots, \boldsymbol{C}_t\right] \in \{0,1\}^{t \times N},$$
$$\boldsymbol{S}_{1:t} \coloneqq \left[\boldsymbol{S}_1\ \boldsymbol{S}_2\ \ldots, \boldsymbol{S}_t\right] \in \{0,1\}^{t \times N}, \qquad \boldsymbol{D}_{1:t} \coloneqq \left[\boldsymbol{D}_1\ \boldsymbol{D}_2\ \ldots, \boldsymbol{D}_t\right] \in \{0,1\}^{t \times N}.$$

[1] An extension to a random number of test takers is possible, but will not be attempted in this article. However, the heuristics presented in Section 4 are applicable in both cases since they do not require N to be known in advance.

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

2.2. Information flow

Before going into more details of the model underlying the random variables defined in the previous section, it is useful to highlight how the available information evolves over time. Before the administration at time instant $t+1$, the available data consist of the features of all previous test takers, $\boldsymbol{X}_{1:t}$, the subset of test takers selected for review, $\boldsymbol{S}_{1:t}$, and the outcomes of these reviews, $\boldsymbol{D}_{1:t}$. We denote the σ-algebra of events generated by these random variables by

$$\mathcal{F}_t \coloneqq \sigma\big(\boldsymbol{X}_{1:t}, \boldsymbol{S}_{1:t}, \boldsymbol{D}_{1:t}\big). \tag{1}$$

After the administration at time instant $t+1$ is completed and all responses are processed, the extracted features, $\boldsymbol{X}_{t+1}$, become available. We denote this refined σ-algebra by

$$\mathcal{F}'_t \coloneqq \sigma\big(\boldsymbol{X}_{1:t+1}, \boldsymbol{S}_{1:t}, \boldsymbol{D}_{1:t}\big). \tag{2}$$

Note that the decision about which test takers to select for review should take $\boldsymbol{X}_{t+1}$ into account, hence, it can use all information available in $\mathcal{F}'_t$, not just $\mathcal{F}_t$. Finally, the selected responses, $\boldsymbol{S}_{t+1}$, and the outcomes of the corresponding reviews, $\boldsymbol{D}_{t+1}$, are observed. This refines the σ-algebra to $\mathcal{F}_{t+1}$ and completes the cycle. Note that $\mathcal{F}_t \subset \mathcal{F}'_t \subset \mathcal{F}_{t+1}$ for all $t = 1, \ldots, T-1$.

2.3. Assumptions

Throughout the article, we make the following assumptions:

1. All group membership indicators, $C_{t,n}$, are independent and identically distributed Bernoulli random variables with success probability R so that $P(C_{t,n} = 1 \,|\, R = r) = r$.
2. All feature vectors, $\boldsymbol{X}_{t,n}$, are independent and identically distributed conditioned on the group membership indicator $C_{t,n}$. That is, there exist two random vectors, $\boldsymbol{X}_0$ and $\boldsymbol{X}_1$, such that $\boldsymbol{X}_{t,n} | (C_{t,n} = 0) \overset{d}{=} \boldsymbol{X}_0$ and $\boldsymbol{X}_{t,n} | (C_{t,n} = 1) \overset{d}{=} \boldsymbol{X}_1$ for all t and n. The distributions of $\boldsymbol{X}_0$ and $\boldsymbol{X}_1$ are denoted by P_0 and P_1, respectively.
3. P_0 and P_1 are assumed to be members of two, not necessarily identical, parametric families of distributions. That is, there exist two (vector) parameters in suitably defined spaces, $\theta_0 \in \mathcal{T}_0$ and $\theta_1 \in \mathcal{T}_1$, such that

$$\begin{aligned} P_0 &= P(\bullet \,|\, \theta_0), \\ P_1 &= P(\bullet \,|\, \theta_1). \end{aligned} \tag{3}$$

4. The selection process is modeled as follows: After each administration, $K \leq N$ test takers are selected for review. The corresponding selection policy is a function $\xi_t : \mathcal{S}^N \to \Delta^N$ that assigns a probability to every possible selection vector. Each ξ_t is assumed to be $\mathcal{F}'_{t-1}$-measurable, that is, the selection policy can depend only on events whose occurrence can be established based on knowledge of $\boldsymbol{X}_{1:t}$, $\boldsymbol{S}_{1:t-1}$, and $\boldsymbol{D}_{1:t-1}$. Note that this implies that $\boldsymbol{S}_t$ and $\boldsymbol{C}_t$ are conditionally independent since

$$\begin{aligned} P(\boldsymbol{S}_t = \boldsymbol{s}, \boldsymbol{C}_t = \boldsymbol{c} \,|\, \mathcal{F}'_{t-1}) &= P(\boldsymbol{S}_t = \boldsymbol{s} \,|\, \mathcal{F}'_{t-1})\, P(\boldsymbol{C}_t = \boldsymbol{c} \,|\, \boldsymbol{S}_t = \boldsymbol{s}, \mathcal{F}'_{t-1}) \\ &= \xi_t(\boldsymbol{s})\, P(\boldsymbol{C}_t = \boldsymbol{c} \,|\, \mathcal{F}'_{t-1}), \end{aligned} \tag{4}$$

   where the second equality holds since $\boldsymbol{S}_t$ can only provide information that is already contained in $\mathcal{F}'_{t-1}$.
5. The review process is modeled as follows: All flagged test takers are reviewed and the review resolves any ambiguity about the test taker's group membership. Test takers that did not get

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

flagged are automatically assumed to be members of the reference group. Unflagged members of the critical group remain undetected. This means

$$D_{t,n} = \begin{cases} 1, & S_{t,n} = 1, C_{t,n} = 1 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

or, using Hadamard's element-wise vector product notation,

$$\boldsymbol{D}_t = \boldsymbol{S}_t \odot \boldsymbol{C}_t. \tag{6}$$

Assumptions 1 and 2 are made to simplify the mathematical analysis and to keep the model general. Possible relaxations, such as allowing dependencies between test takers (and possibly re-takers), or considering sub-populations with different feature distributions or rates of critical group members, will significantly complicate the analysis and likely depend on what behavior or phenomenon underpins the definition of the critical group. A brief outlook on possible research avenues in this direction will be given in Section 7.

Assumption 3 enables us to formulate the flagging problem in a standard Bayesian setting. An extension to a non-parametric formulation is beyond the scope of the article.

Assumption 4 is common in sequential decision making and ensures that the policy does not use information that is not available at the time the selection decision needs to be made.

Assumption 5 is likely the most controversial, and it will not always hold in practice. However, at least in the context of test security, various standards and criteria have been established according to which an expert or a panel of experts can evaluate the evidence and arrive at a well-justified decision whether or not a certain response or behavior constitutes cheating. Alternatively, one can define the critical group as those test takers who an expert considers "sufficiently curious" to be reviewed in detail, irrespective of the outcome of the review. This criterion is less sharp, but can be more applicable in practice. In general, we assume that some, potentially expensive and time-consuming, process exists that assigns "true positive" and "false positive" labels to the selected cases. The proposed procedure is agnostic to the exact meaning of these labels and the process used to assign them.

2.4. Problem formulation

As mentioned before, we adopt a Bayesian perspective in this article. That is, we assume that unknown quantities are themselves random variables, and that the joint distribution of all random variables, latent and observable, is known. We denote this joint distribution by P, so that

$$\left(\boldsymbol{X}_{1:T}, \boldsymbol{S}_{1:T}, \boldsymbol{D}_{1:T}, \boldsymbol{C}_{1:T}, R, \Theta_0, \Theta_1, \right) \sim P, \tag{7}$$

where $\boldsymbol{X}_{1:T}, \boldsymbol{C}_{1:T}, \boldsymbol{S}_{1:T}, \boldsymbol{D}_{1:T}$, and R are defined in Section 2.1, and Θ_0 and Θ_1 denote the parameters of the conditional feature distributions in (3).

Our aim is to design a selection policy, $\boldsymbol{\xi}_{1:T}$, that maximizes the expected number of detected members of the critical group, while limiting the number of expert reviews per test administration to at most $K \leq N$. This leads to the following optimization problem:

$$\max_{\boldsymbol{\xi}_{1:T}} E\left[\sum_{t=1}^{T}\sum_{n=1}^{N} D_{t,n}\right] \quad \text{s. t.} \quad \sum_{n=1}^{N} S_{t,n} \leq K, \tag{8}$$

where the constraint needs to hold almost surely for all $t = 1, \ldots, T$.

Clearly, (8) is by no means the only way of formalizing the flagging problem. For example, one could relax the constraint to hold in expectation, constrain the false or true positive rate instead of the absolute number of reviews, or, in a more traditional Bayesian formulation, define costs of reviews, detections, false alarms, and so on. Nevertheless, we believe that the problem formulation in (8) strikes a good balance between transparency, tractability, and applicability. In particular, the question of how

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

to set K, that is, the number of reviews one is willing to afford, is well-defined and can be understood and discussed without deeper technical knowledge. By contrast, choosing Bayesian costs is typically less straightforward and can lead to confusion about what these cost should and should not reflect. Moreover, we believe that the optimal and approximate selection policies presented in this article provide a blueprint that can be adapted to variations of the problem in (8) in a relatively straightforward manner.

3. Optimal selection policy

This section states and discusses the selection policy that solves the problem in (8). While typically too complex for practical use, it offers conceptual insights and can guide the design of approximate or heuristic policies.

Theorem 1. *Let* $\left(r_t\right)_{0\leq t\leq T}$ *and* $\left(\rho_t\right)_{0\leq t\leq T}$ *with* $r_t\colon \mathcal{X}^{(t+1)N}\times\{0,1\}^{tN}\times\{0,1\}^{tN}\to\mathbb{R}_+$ *and* $\rho_t\colon \mathcal{X}^{tN}\times\{0,1\}^{tN}\times\{0,1\}^{tN}\to\mathbb{R}_+$ *be two sequences of functions that are defined recursively via*

$$r_{t-1}(\boldsymbol{X}_{1:t},\boldsymbol{S}_{1:t-1},\boldsymbol{D}_{1:t-1}) \coloneqq \max_{\boldsymbol{s}\in\mathcal{S}_K^N} E_{\boldsymbol{C}_t}\left[\boldsymbol{s}^\top\boldsymbol{C}_t+\rho_t\left(\boldsymbol{X}_{1:t},[\boldsymbol{S}_{1:t-1},\boldsymbol{s}],[\boldsymbol{D}_{1:t-1},\boldsymbol{s}\odot\boldsymbol{C}_t]\right)\middle|\mathcal{F}'_{t-1}\right] \tag{9}$$

and

$$\rho_t(\boldsymbol{X}_{1:t},\boldsymbol{S}_{1:t},\boldsymbol{D}_{1:t}) \coloneqq E_{\boldsymbol{X}_{t+1}}\left[r_t\left([\boldsymbol{X}_{1:t},\boldsymbol{X}_{t+1}],\boldsymbol{S}_{1:t},\boldsymbol{D}_{1:t}\right)\middle|\mathcal{F}_t\right], \tag{10}$$

with recursion base $\rho_T(\boldsymbol{x}_{1:T},\boldsymbol{s}_{1:T},\boldsymbol{d}_{1:T})=0$. *Let* $\mathcal{S}_t^*\subset\mathcal{S}_K^N$ *be the set of selection vectors that attain the maximum on the right-hand side of (9). Every selection policy whose probability is concentrated on the sets* $\mathcal{S}_1^*,\ldots,\mathcal{S}_T^*$, *that is,*

$$\xi_t(\mathcal{S}_t^*)=1 \tag{11}$$

for all $t=1,\ldots,T$ *is optimal in the sense of (8).*

Theorem 1 is proven in Appendix A. The functions r_t and ρ_t represent two types of expected rewards. Specifically, r_t denotes the expected number of critical group members that will be flagged in the remaining $T-t$ test administrations, given the data from all previous administrations and the feature vectors of the next one. The function ρ_t is then obtained by marginalizing r_t over the feature vectors of the next administration with respect to their current posterior predictive distribution. In what follows, we will occasionally refer to ρ_t as a *look-ahead* step or function.

The optimal selection policy will be discussed in more detail shortly. Before doing so, it is instructive to investigate how the underlying probability distributions evolve as more data become available and how the conditional expectations in (9) and (10) can be calculated.

3.1. Posterior update

As more tests are administered and more reviews are conducted, more information is gathered that needs to be incorporated into the underlying Bayesian model. In this section, we detail the corresponding updates and provide expressions for the respective distributions. Of particular interest is the joint posterior distribution of Θ_0, Θ_1, and R since the latter are assumed to persist between test administrations. That is, Θ_0, Θ_1, and R can be learned from the data, while the uncertainty about the group indicators, $\boldsymbol{C}_t$, will not be resolved completely in general.

Assume that the administration at time instant $t-1$ was completed, and let the corresponding conditional PDF of the persistent random variables be denoted by $p(r,\theta_0,\theta_1\,|\,\mathcal{F}_{t-1})$. Now, consider the joint distribution of $\boldsymbol{X}_t$ and $\boldsymbol{C}_t$ conditioned on the model parameters:

$$p(\boldsymbol{x},\boldsymbol{c}\,|\,r,\theta_0,\theta_1) = p(\boldsymbol{c}\,|\,r,\theta_0,\theta_1)\,p(\boldsymbol{x}\,|\,r,\theta_0,\theta_1,\boldsymbol{c}) \tag{12}$$

$$= p(\boldsymbol{c}\,|\,r)\,p(\boldsymbol{x}\,|\,\theta_0,\theta_1,\boldsymbol{c}) \tag{13}$$

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

$$= \prod_{n=1}^{N} \left[r\, p(x_n \mid \theta_1) \right]^{c_n} \left[(1-r)\, p(x_n \mid \theta_0) \right]^{1-c_n} \tag{14}$$

$$= r^{\|\boldsymbol{c}\|_1} (1-r)^{N-\|\boldsymbol{c}\|_1} \prod_{n=1}^{N} p(x_n \mid \theta_1)^{c_n} p(x_n \mid \theta_0)^{1-c_n}. \tag{15}$$

Since conditioning on the model parameters renders the distribution independent of the specific administration, we dropped the time index in the notation. By marginalizing out the unknown group membership, we obtain the feature distribution for a given set of model parameters:

$$p(\boldsymbol{x} \mid r, \theta_0, \theta_1) = \sum_{\boldsymbol{c} \in \mathcal{S}^N} p(\boldsymbol{x}, \boldsymbol{c} \mid r, \theta_0, \theta_1) \tag{16}$$

$$= \sum_{\boldsymbol{c} \in \mathcal{S}^N} r^{\|\boldsymbol{c}\|_1} (1-r)^{N-\|\boldsymbol{c}\|_1} \prod_{n=1}^{N} p(x_n \mid \theta_1)^{c_n} p(x_n \mid \theta_0)^{1-c_n}. \tag{17}$$

Averaging over the unknown model parameters yields the posterior predictive distribution of the features at time instant t:

$$p(\boldsymbol{x}_t \mid \mathcal{F}_{t-1}) = E\left[p(\boldsymbol{x}_t \mid R, \Theta_0, \Theta_1) \mid \mathcal{F}_{t-1} \right], \tag{18}$$

where the expected value is taken with respect to the current model posterior, $p(r, \theta_0, \theta_1 \mid \mathcal{F}_{t-1})$. Note that the posterior predictive in (18) is needed for the look-ahead step in (10).

Now, assume that the test administration at time instant t was completed and that $\boldsymbol{X}_t = \boldsymbol{x}_t$ was observed. In order to determine the optimal selection policy in (9), we require the posterior distribution of the group memberships, $\boldsymbol{C}_t$, conditioned on all previous observations. According to Bayes' rule, this distribution is given by

$$p(\boldsymbol{c}_t \mid \mathcal{F}'_{t-1}) = \frac{p(\boldsymbol{x}_t, \boldsymbol{c}_t \mid \mathcal{F}_{t-1})}{p(\boldsymbol{x}_t \mid \mathcal{F}_{t-1})} = \frac{E\left[p(\boldsymbol{c}_t, \boldsymbol{x}_t \mid R, \Theta_0, \Theta_1) \mid \mathcal{F}_{t-1} \right]}{E\left[p(\boldsymbol{x}_t \mid R, \Theta_0, \Theta_1) \mid \mathcal{F}_{t-1} \right]}, \tag{19}$$

where, again, all expected values are taken with respect to $p(r, \theta_0, \theta_1 \mid \mathcal{F}_{t-1})$.

Given (19), the optimal selection vector can, in principle, be determined by finding the maximum in (9), and the corresponding test takers can be reviewed. In our model, this corresponds to observing $\boldsymbol{S}_t = \boldsymbol{s}_t$ and $\boldsymbol{D}_t = \boldsymbol{d}_t$. This new information leads to the model update

$$p(r, \theta_0, \theta_1 \mid \mathcal{F}_t) = \frac{p(\boldsymbol{x}_t, \boldsymbol{s}_t, \boldsymbol{d}_t \mid r, \theta_0, \theta_1, \mathcal{F}_{t-1})}{p(\boldsymbol{x}_t, \boldsymbol{s}_t, \boldsymbol{d}_t \mid \mathcal{F}_{t-1})} p(r, \theta_0, \theta_1 \mid \mathcal{F}_{t-1}). \tag{20}$$

The likelihood in the numerator of (20) is given by

$$p(\boldsymbol{x}_t, \boldsymbol{s}_t, \boldsymbol{d}_t \mid r, \theta_0, \theta_1, \mathcal{F}_{t-1}) = \sum_{\boldsymbol{c}_t \in \mathcal{S}^N} p(\boldsymbol{x}_t, \boldsymbol{c}_t, \boldsymbol{s}_t, \boldsymbol{d}_t \mid r, \theta_0, \theta_1, \mathcal{F}_{t-1}) \tag{21}$$

$$= \sum_{\boldsymbol{c}_t \in \mathcal{S}^N} p(\boldsymbol{d}_t \mid \boldsymbol{s}_t, \boldsymbol{c}_t)\, \xi_t(\boldsymbol{s}_t)\, p(\boldsymbol{x}_t, \boldsymbol{c}_t \mid r, \theta_0, \theta_1) \tag{22}$$

$$= \xi_t(\boldsymbol{s}_t) \sum_{\boldsymbol{c}_t \in \mathcal{S}^N} \left[\boldsymbol{d}_t = \boldsymbol{s}_t \odot \boldsymbol{c}_t \right] p(\boldsymbol{x}_t, \boldsymbol{c}_t \mid r, \theta_0, \theta_1), \tag{23}$$

where we used Iverson bracket notation as a shorthand for the indicator function, meaning that the summands on the right-hand side of (23) are zero unless $\boldsymbol{d}_t = \boldsymbol{s}_t \odot \boldsymbol{c}_t$ holds. Now, define

$$q(\boldsymbol{x}, \boldsymbol{s}, \boldsymbol{d} \mid r, \theta_0, \theta_1) = \sum_{\boldsymbol{c} \in \mathcal{C}(\boldsymbol{d}, \boldsymbol{s})} p(\boldsymbol{x}, \boldsymbol{c} \mid r, \theta_0, \theta_1), \tag{24}$$

where $\mathcal{C}(\boldsymbol{d}, \boldsymbol{s}) \coloneqq \left\{ \boldsymbol{c} \in \mathcal{S}^N : \boldsymbol{d} = \boldsymbol{s} \odot \boldsymbol{c} \right\}$. Note that $q(\boldsymbol{x}, \boldsymbol{s}, \boldsymbol{d} \mid r, \theta_1, \theta_1)$ in (24) and $p(\boldsymbol{x} \mid r, \theta_1, \theta_1)$ in (17) are closely related. While the latter is obtained by summing over *all* possible group memberships, the former is obtained by summing only over those that are compatible with the observed review outcomes.

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

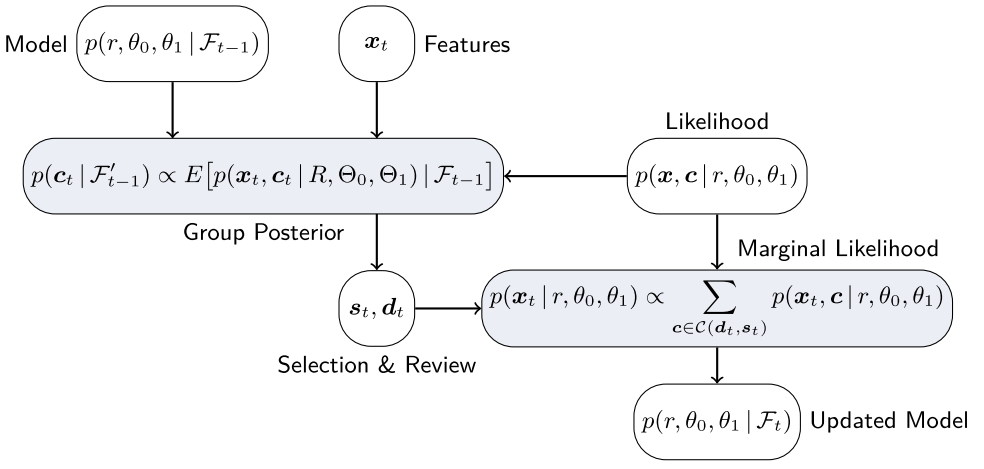


Figure 1. Graphical illustration of information flow and posterior update of the Bayesian model.

This implies that (24) needs to be normalized to be a valid PDF. The final posterior update is then given by

$$p(r,\theta_0,\theta_1 \mid \mathcal{F}_t) = \frac{q(\boldsymbol{x}_t,\boldsymbol{s}_t,\boldsymbol{d}_t \mid r,\theta_0,\theta_1)}{E\big[q(\boldsymbol{x}_t,\boldsymbol{s}_t,\boldsymbol{d}_t \mid R,\Theta_0,\Theta_1) \mid \mathcal{F}_{t-1}\big]}\, p(r,\theta_0,\theta_1 \mid \mathcal{F}_{t-1}). \tag{25}$$

The factor $\xi_t(\boldsymbol{s}_t)$ in (23) is $\mathcal{F}_{t-1}$-measurable and cancels out in the normalization. The update steps now repeat with $t \leftarrow t+1$ (see Figure 1 for a graphical illustration of the process).

3.2. Discussion

In this section, we briefly discuss the general structure and some properties of the optimal selection policy stated in Theorem 1. This discussion will also provide qualitative insights that can inform the design of approximate procedures.

By inspection, the optimal selection policy in (9) maximizes the expected value of a sum of two terms, $\boldsymbol{s}^\top \boldsymbol{C}_t$ and ρ_t. The first term corresponds to the number of critical group members detected in the current administration; the second term corresponds to the expected number of critical group members detected in all future administrations. This trade-off between immediate and future rewards is typical for sequential decision-making procedures. However, there are a few aspects that are non-standard:

- In general, the optimal policy at time t depends on all data observed so far and requires averaging over all future data. This results in extremely high complexity, even for small T and N, making strictly optimal policies virtually impossible to compute in practice.
- In line with the information evolving in two steps, compare Section 2.2, the calculation of the expected reward in Theorem 1 is split into two parts. First, in (9), the review outcomes are predicted based on the observed features, $\boldsymbol{X}_t$. Then, in (10), the features themselves are predicted based on the data from previous administrations. These two expectations correspond to two different sources of uncertainty: feature vectors need to be estimated to predict future rewards, but can in principle be observed. In contrast, the group memberships need to be estimated because they are unobservable to begin with.
- The optimization over the selection vector, $\boldsymbol{s}$, is "sandwiched" between the two expected values discussed in the previous bullet point. It happens inside the expectation over the feature vectors, which are available at the time the selection is made, but outside the expectation over the unknown group memberships.

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

- There always exists a deterministic optimal selection policy. This follows directly from the fact that in (A.13) a function that is linear in the optimization variable, ξ_t, is maximized over a probability simplex. Therefore, at least one vertex attains the maximum, and any vertex of a probability simplex corresponds to a single point mass on the respective outcome.
- The update of the posterior distribution detailed in Section 3.1 depends on the selection, but not on the *selection policy*. This is the case since the selected test takers only enter via the marginalization in (24). The latter requires knowledge of the selection, $\boldsymbol{s}$, but is independent of how this selection was made.
- In principle, it is not necessary to track $\boldsymbol{S}_t$ and $\boldsymbol{D}_t$ separately. Both can be combined into a single vector that indicates if the respective test taker is known to be a member of the critical group, of the reference group, or if their group membership is unknown. We decided to go with the slightly lengthier notation since we believe it to be conceptually simpler.

This concludes the discussion of the optimal selection policy. Based on the insights gained, we now turn to the design of approximate selection policies that are practical to implement.

4. Approximate methods

For the reasons discussed above, implementing the optimal selection policy is prohibitively complex, even for small N and T. In this section, we present heuristic policies that are significantly simpler yet sufficiently flexible and powerful to be useful in practice.

As discussed in Section 3.2, the update of the posterior distribution is independent of the selection policy. Therefore, we treat the two separately: first, we propose a variational approximation of the posterior distribution, and, second, we discuss three selection policies that can be used in conjunction with the approximate posterior update.

4.1. Approximate posterior update

The approximation method proposed in this section is essentially an expectation maximization (EM) algorithm (Dempster et al., 2018). However, it also incorporates ideas from the mixture-based extended auto-regressive model, see (Šmídl & Quinn, 2006, Chapter 8), and basic copula theory (Jaworski et al., 2010). First, each element of the feature vector, $\boldsymbol{X}$, is assumed to follow a distribution from the exponential family. That is,

$$p_{X_m}(x|\eta_m) = h_m(x)\exp\left(\eta_m^\top T_m(x) - A_m(\eta)\right), \tag{26}$$

with natural parameters $\eta_m \in \mathcal{T}_m$, base measure h_m, sufficient statistic T_m, log-partition A_m and $m = 1,\dots,M$. Correlations between the elements of $\boldsymbol{X}$ are modeled via a Gaussian copula. To this end, we define the random vector $\boldsymbol{Z} \in \mathbb{R}^M$ as

$$Z_m = \Phi^{-1}\left(F_{\boldsymbol{\eta}_m}(X_m)\right), \tag{27}$$

where $F_{\boldsymbol{\eta}_m}$ denotes the cumulative distribution function (CDF) of P_{X_m}, and Φ denotes the CDF of the standard normal distribution. The random vector $\boldsymbol{Z}$ is then modeled as normally distributed with mean zero and covariance $\boldsymbol{\Sigma}$, that is, $\boldsymbol{Z} \sim \mathcal{N}(0,\boldsymbol{\Sigma})$. It can be shown that (26) and (27) define a class of distributions with PDFs

$$p_{\boldsymbol{X}}(\boldsymbol{x}|\theta) = p_{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{\eta},\boldsymbol{\Sigma}) = p_{\boldsymbol{Z}}(\boldsymbol{z}|\boldsymbol{\Sigma})\prod_{m=1}^{M} g(x_m|\eta_m). \tag{28}$$

Here, $\boldsymbol{z}$ is calculated from $\boldsymbol{x}$ according to (27), $p_{\boldsymbol{Z}}(\bullet|\boldsymbol{\Sigma})$ denotes the PDF of a zero-mean multivariate normal distribution with covariance matrix $\boldsymbol{\Sigma}$, and we defined

$$g_m(x|\eta_m) \coloneqq \sqrt{2\pi}\exp\left(\operatorname{erfc}^{-1}\left(2F_{\eta_m}(x)\right)^2\right)p_{X_m}(x|\eta_m), \tag{29}$$

where erfc denotes the complementary error function.

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

The model in (28) is assumed to hold under both hypotheses, that is, for test takers in the critical group and in the reference group. The corresponding parameters and functions are denoted by an additional subscript $i \in \{0,1\}$, for example, $\eta_{i,m}$. Note that the exponential families of distributions do not need to be of the same type under both hypotheses.

A priori, we assume the free parameters to be independent and distributed according to

$$R \sim \text{Beta}\left(\nu_1^{(0)}, \nu_0^{(0)}\right), \tag{30}$$

$$\boldsymbol{\Sigma}_i \sim \mathcal{W}^{-1}\left(\nu_i^{(0)}, \boldsymbol{\Psi}_i^{(0)}\right), \tag{31}$$

$$H_{i,m} \sim \Pi_m\left(\nu_i^{(0)}, \chi_{i,m}^{(0)}\right), \tag{32}$$

where Beta denotes the beta distribution, $\mathcal{W}^{-1}$ denotes the inverse Wishart distribution, and Π_m denotes the conjugate prior of the exponential family distribution corresponding to the density in (26). This distribution exists and its density is of the form

$$\pi_H(\eta \,|\, \nu, \chi) \propto \exp\left(\eta^\top \chi - \nu A(\eta)\right). \tag{33}$$

The parameters ν_0 and ν_1, which represent the effective number of observed samples under each hypothesis, are shared across all priors. Allowing for different initial effective sample sizes, $\nu_0^{(0)}$ and $\nu_1^{(0)}$, is straightforward. We do not introduce this modification here, as it would significantly complicate the notation. In practice, however, it can be the case that, for example, one has more prior knowledge about the cheating rate distribution than the feature distribution.

While not guaranteed to accurately reflect the true prior knowledge in general, we consider the conjugate priors in (30)–(32) a useful and robust choice in practice—mainly for two reasons: First, as will become clear later in this section, conjugate priors significantly reduce computational complexity and eliminate the need for Monte Carlo sampling or numerical optimization in the inference step. Second, they are arguably a good fit for the problem at hand. Since the flagging procedure is assumed to run periodically, a natural prior for the parameters at time t is simply the posterior from the previous time instant, $t-1$. In this chain, later priors are more accurate, as they incorporate more information. This effect is captured by the conjugate priors, which, by construction, remain of the same type after each administration, but increase their effective sample size, reflecting the larger set of training data. In the absence of training data, the *least informative prior* is obtained by setting the effective sample size to its minimum feasible value.

Based on the model specified above, we propose the following approximate posterior update. At time instant t, the approximate posterior distributions of R, $\boldsymbol{\Sigma}_i$, and $H_{m,i}$ are given by

$$Q_R^{(t)} = \text{Beta}\left(\nu_1^{(t)}, \nu_0^{(t)}\right), \tag{34}$$

$$Q_{\boldsymbol{\Sigma}_i}^{(t)} = \mathcal{W}^{-1}\left(\nu_i^{(t)}, \boldsymbol{\Psi}_t^{(t)}\right), \tag{35}$$

$$Q_{H_{i,m}}^{(t)} = \Pi_m\left(\nu_i^{(t)}, \chi_{i,m}^{(t)}\right). \tag{36}$$

The distribution parameters on the right-hand side, $\chi_{i,m}^{(t)}$, $\boldsymbol{\Psi}_t^{(t)}$, and $\nu_i^{(t)}$, are obtained from their previous values, $\chi_{i,m}^{(t-1)}$, $\boldsymbol{\Psi}_t^{(t-1)}$, and $\nu_i^{(t-1)}$, by solving the following system of equations:

$$q_{C_{t,n}}(1) = s_{t,n} d_{t,n} + (1 - s_{t,n}) \frac{\nu_1^{(t)} p_X\left(\boldsymbol{x}_{t,n} \,|\, \hat{\boldsymbol{\eta}}_1^{(t)}, \mathcal{P}_\mathcal{E}\left(\boldsymbol{\Psi}_1^{(t)}\right)\right)}{\sum_{j=0}^{1} \nu_j^{(t)} p_X\left(\boldsymbol{x}_{t,n} \,|\, \hat{\boldsymbol{\eta}}_j^{(t)}, \mathcal{P}_\mathcal{E}\left(\boldsymbol{\Psi}_j^{(t)}\right)\right)}, \tag{37}$$

$$\hat{\eta}_{i,m}^{(t)} \in \arg\max_{\eta \in \mathcal{T}_m} \eta^\top \chi_{i,m}^{(t)} - \nu_i^{(t)} A_{i,m}(\eta), \tag{38}$$

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

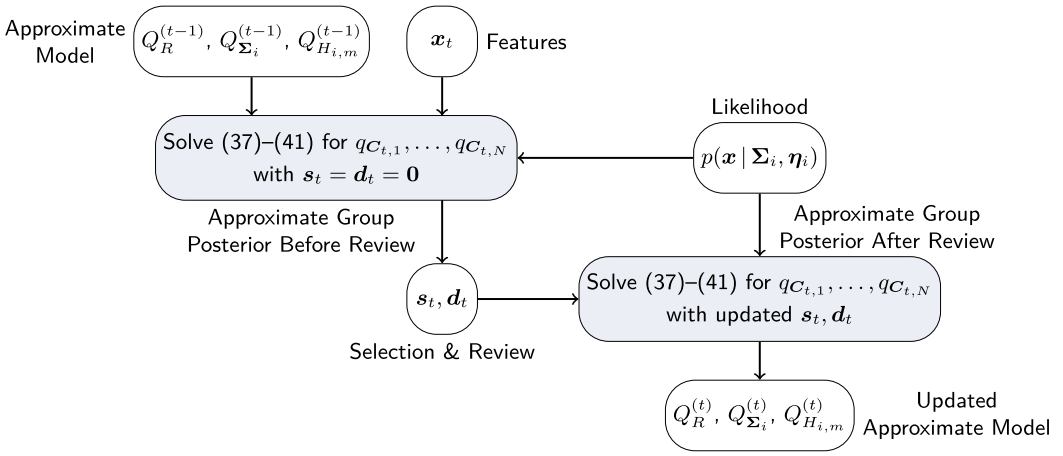


Figure 2. Graphical illustration of the approximate posterior update in (37)–(41).

$$\boldsymbol{\Psi}_i^{(t)} = \boldsymbol{\Psi}_i^{(t-1)} + \sum_{n=1}^{n} q_{C_{t,n}}(i)\, \boldsymbol{z}_{i,t,n} \boldsymbol{z}_{i,t,n}^{\top}, \tag{39}$$

$$\nu_i^{(t)} = \nu_i^{(t-1)} + \sum_{n=1}^{N} q_{C_{t,n}}(i), \tag{40}$$

$$\chi_i^{(t)} = \chi_i^{(t-1)} + \sum_{n=1}^{n} q_{C_{t,n}}(i)\, T(x_{t,n}), \tag{41}$$

where $i \in \{0,1\}$ and $\boldsymbol{z}_{i,t,n} = \Phi^{-1}\left(F_{\boldsymbol{\eta}_i^{(t)}}(\boldsymbol{x}_{t,n})\right)$. The proposed approximate posterior update for a given selection policy is summarized in Algorithm 1, a graphical illustration in analogy to Figure 1 is shown in Figure 2.

Note that the posterior update in Algorithm 1 is run twice—once before the test takers are selected and once after the reviews are completed. The purpose of the first update is to obtain the approximate posterior distributions of the group indicator variables, $q_{C_{t,n}}$. The selection policies discussed in the next section are all based on these distributions. The purpose of the second update is to refine the estimates of $\nu_i^{(t)}, \chi_i^{(t)}$, and $\boldsymbol{\Psi}_i^{(t)}$ by incorporating the outcomes of the reviews. The difference between the first and second runs is reflected in (37), where the probability that the nth test taker of the tth administration is a member of the critical group is estimated from the observed features in case the test taker has not been reviewed ($s_{t,n} = 0$), or is set to the true label, $d_{t,n}$, if the review outcome is known ($s_{t,n} = 1$).

Algorithm 1 Approximate posterior update for given selection policy.

function POSTERIOR UPDATE($\boldsymbol{x}_t$, ξ_t, $\nu_i^{(t-1)}$, $\chi_i^{(t-1)}$, $\boldsymbol{\Psi}_i^{(t-1)}$,)
  Set $\boldsymbol{s}_t \leftarrow 0$ and $\boldsymbol{d}_t \leftarrow 0$
  Solve (37)–(41) for $\nu_i^{(t)}$, $\chi_i^{(t)}$, $\boldsymbol{\Psi}_i^{(t)}$ and $q_{C_{t,n}}$
  Select test takers according to ξ_t and update $\boldsymbol{s}_t$ accordingly
  Review selected test takers and update $\boldsymbol{d}_t$ accordingly
  Solve (37)–(41) for $\nu_i^{(t)}$, $\chi_i^{(t)}$, $\boldsymbol{\Psi}_i^{(t)}$ and $q_{C_{t,n}}$
  return $\nu_i^{(t)}$, $\chi_i^{(t)}$, $\boldsymbol{\Psi}_i^{(t)}$
end function

The idea underlying the update equations in (37)–(41) is to start with the priors in (30)–(32) and approximate the posterior distributions via a mean-field variational approximation (Blei et al., 2017;

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

Šmídl & Quinn, 2006). By construction, this approximation yields posterior distributions that are of the same type as the priors, and whose parameters are determined implicitly by a system of equations. However, this system is still hard to solve since it involves expectations taken with respect to the current posterior distribution. In order to simplify the evaluation of the latter, we use maximum a posteriori probability (MAP) estimates as an approximation of the expected values. More precisely, to evaluate the posterior group membership probabilities in (37), we calculate MAP estimates of the marginal parameters in (38). Once the marginal parameters are fixed, the observations, $\boldsymbol{x}_{t,n}$, can be mapped to the "Gaussian domain," $\boldsymbol{z}_{t,n}$, via (27). The latter can then be used to update the covariance posterior in (39). The effective sample sizes and the parameters of the posterior distributions of the marginal parameters are updated according to (40) and (41). Of course, since all updates are coupled, they cannot simply be performed in sequence, but need to be solved jointly. We will further elaborate on this aspect later in this section.

Finally, note that in the model updates (39)–(41), the posterior group probabilities are used as weights, that is, the observation $x_{t,n}$ contributes to the model of the critical group with weight $q_{C_{n,t}}(1)$ and to the model of the reference group with weight $q_{C_{n,t}}(0) = 1 - q_{C_{n,t}}(1)$. This observation will later be used in the derivation of heuristic selection policies.

Before concluding this section, we briefly discuss some implementation aspects of Algorithm 1:

- There are various ways of solving the equation system in (37)–(41) in practice. A natural approach is to iterate over the updates until all parameters and probabilities have converged. Alternatively, one can use standard numerical solvers. In this case, it is useful to consider only $q_{C_{t,n}}(1)$, $n = 1, \ldots, N$, as free variables since all other variables are fixed once all $q_{C_{t,n}}(1)$ are given. Note that this means that the equation system is of size N in the first update step of Algorithm 1 and of size $N - K$ in the second.
- Depending on the choice of the marginal distributions, the optimization problems in (38) may not have a closed-form solution. However, even if (38) needs to be solved numerically, the underlying distributions are univariate so the number of parameters is typically small. Moreover, for given $\chi_{i,m}$ and $\nu_{i,m}$, the $2M$ problems in (38) can be solved in parallel.
- In (37), the covariance matrix of the feature distribution is projected on the elliptope of correlation matrices. This projection is not strictly necessary. However, in our experiments, it leads to an improved performance. This is likely the case because the true $\boldsymbol{\Sigma}$ has unit diagonal elements if the model in (26) holds exactly. However, in practice, replacing the projection in (37) with the maximum likelihood estimate, $\frac{\boldsymbol{\Psi}_i}{\nu_i + M + 1}$ might improve the performance in some cases.
- In practice, θ_0, θ_1, and r do not necessarily remain constant for all $t = 1, \ldots, T$. While initially the Bayesian model can and will adapt to variations, the more samples are observed, the higher its "inertia" becomes. In order to counteract this effect, past observations can be down-weighted in the posterior update. For example, an exponential decay can be introduced by making the substitutions $\nu_i^{(t-1)} \leftarrow \omega \nu_i^{(t-1)}$, $\chi_i^{(t-1)} \leftarrow \omega \chi_i^{(t-1)}$, and $\boldsymbol{\Psi}_i^{(t-1)} \leftarrow \omega \boldsymbol{\Psi}_i^{(t-1)}$ in (39)–(41), where $0 < \omega < 1$ is a weight that balances adaptivity and steady-state performance—compare (Šmídl & Quinn, 2006, Section 8.6).

4.2. Approximate selection policies

The algorithm proposed in the previous section offers a mechanism to track the posterior distributions of the model parameters for a given selection policy. However, it does not answer the question of how to choose a selection policy in the first place. This question will be addressed in this section.

We propose three selection policies: A detection-greedy policy, an information-greedy policy, and a mixed policy that balances the information-greedy and detection-greedy approaches depending on the estimated accuracy of the current model.

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

4.2.1. Detection-greedy selection policy

Consider the optimal selection policy defined by the optimization problem (9). Adopting a greedy approach and ignoring the expected future rewards, this problem simplifies to

$$\max_{\boldsymbol{s}\in\mathcal{S}_K^N} E_{\boldsymbol{C}_t}\left[\boldsymbol{s}^\top \boldsymbol{C}_t \,|\, \mathcal{F}'_{t-1}\right] = \max_{\boldsymbol{s}\in\mathcal{S}_K^N} \boldsymbol{s}^\top E_{\boldsymbol{C}_t}\left[\boldsymbol{C}_t \,|\, \mathcal{F}'_{t-1}\right] \tag{42}$$

$$= \max_{\boldsymbol{s}\in\mathcal{S}_K^N} \sum_{n=1}^{N} s_{t,n} P\left(C_{t,n} = 1 \,|\, \mathcal{F}'_{t-1}\right). \tag{43}$$

That is, the optimal greedy selection policy is to select the K test takers with the highest posterior probability of being members of the critical group.

Although the detection-greedy policy is much simpler than the optimal one, evaluating $P\left(C_{t,N} = 1 \,|\, \mathcal{F}'_{t-1}\right)$ will still be too complex in most practical cases. Therefore, the heuristic policy suggested here is to select test takers based on the approximate probability $q_{C_{t,n}}(1)$. This leads to the following selection policy:

Selection Policy 1 (Detection-Greedy). At time instant t, the detection-greedy selection policy, $\xi_t^\dagger$, is defined as

$$\xi_t^\dagger\left(\mathcal{S}_t^\dagger\right) = 1, \tag{44}$$

where

$$\mathcal{S}_t^\dagger \coloneqq \arg\max_{\boldsymbol{s}\in\mathcal{S}_K^N} \sum_{n=1}^{N} s_{t,n} q_{C_{t,n}}(1), \tag{45}$$

with $q_{C_{t,n}}(1)$ defined in (38). In words, $\xi_t^\dagger$ flags the K test takers that are most likely to be members of the critical group according to the approximate posterior probability $q_{C_{t,n}}(1)$.

The detection-greedy policy is a natural choice, and, as the numerical examples in Section 6 will demonstrate, works well in many cases. However, by design, it does not take into account the benefits of potential model improvements that could be achieved by reviewing particularly informative cases. The second policy aims to identify the latter.

4.2.2. Information-greedy selection policy

The idea underlying this selection policy is not to review the cases that are most likely to lead to detections, but to review the cases that provide the *most information*. In terms of the optimal selection policy in (9), this roughly corresponds to maximizing only the expected reward term, ρ_t, and ignoring the immediate reward $\boldsymbol{s}^\top \boldsymbol{C}_t$. In this sense, the information-greedy policy complements the detection-greedy policy. However, as was the case for the latter, directly maximizing ρ_t is computationally infeasible. We propose the following heuristic:

Selection Policy 2 (Information-Greedy). At time instant t, the information-greedy selection policy, $\xi_t^\ddagger$, is defined as

$$\xi_t^\ddagger\left(\mathcal{S}_t^\ddagger\right) = 1, \tag{46}$$

where

$$\mathcal{S}_t^\ddagger \coloneqq \arg\max_{\boldsymbol{s}\in\mathcal{S}_K^N} \sum_{n=1}^{N} s_{t,n} H_b\left(q_{C_{t,n}}(1)\right), \tag{47}$$

H_b denotes the binary entropy function, and $q_{C_{t,n}}(1)$ is defined in (37). In words, $\xi_t^\ddagger$ flags the K test takers whose approximate posterior group-membership distribution admits the highest entropy.

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

The rationale for the information-greedy selection policy is that resolving cases with the highest ambiguity has the greatest effect on model updates. More specifically, in (39)–(41), observed samples contribute to the parameter updates of the posterior distributions with weights $q_{C_{t,n}}$. Reviewing cases that are highly likely to belong to a certain group has little effect on these weights. For example, reviewing a case with $q_{C_{t,n}}(1) = 0.99$ will most likely result in a detection, changing the weight from 0.99 to 1.0. This small change has a negligible impact on the model. In contrast, reviewing cases with high entropy, that is, $q_{C_{t,n}}(1) \approx 0.5$, has a strong impact: Without review, the sample contributes equally to both models, reducing separability. With review, it contributes exclusively to the correct model, increasing separability.

On its own, the information-greedy policy is of limited use, as it does not explicitly aim to maximize the number of detections.[2] However, it serves as a useful building block for an adaptive policy that balances model fit and detection rate (DR). Such a policy is proposed next.

4.2.3. Mixed selection policy

The detection-greedy selection policy is a good choice when the learned model is close enough to the true one to produce sufficiently reliable predictions. However, the information-greedy selection policy might be useful to reduce the time and samples needed to reach this point. The idea underlying the proposed mixed policy is to track the model fit and balance the two policies accordingly. However, quantifying model fit is not straightforward. As discussed before, the optimal measure in (9) is too complex to evaluate in practice. Possible proxy measures include, for example, the variance or entropy of the posterior distribution. However, translating the latter into a useful exploration/exploitation trade-off is a non-trivial problem itself.

The heuristic proposed here is to use the detector calibration as a measure of the model accuracy. The underlying idea is the following: If the learned model is a good fit, the resulting (approximate) posterior probabilities will be reasonably accurate. That is, assuming that the posterior probabilities of the K selected test takers being critical group members are $q_1, \ldots, q_K$, the number of critical group members detected in the review process should be close to the sum of these probabilities. More precisely, it should hold that

$$E\left[\sum_{n=1}^{N} D_{t,n}\right] = \sum_{k=1}^{K} q_k \quad \text{and} \quad \operatorname{Var}\left[\sum_{n=1}^{N} D_{t,n}\right] = \sum_{k=1}^{K} q_k(1-q_k). \tag{48}$$

If the number of true positives is significantly lower than one would expect according to (48), the model likely needs improvement. This motivates the following selection policy:

Selection Policy 3 (Mixed). Let $d_{t-1} \coloneqq \sum_{n=1}^{N} D_{t-1,n}$ be the number of detected critical group members in the administration at time instant $t-1$, and let

$$\phi_t = \begin{cases} 1, & \mu_{t-1} - d_{t-1} \le \sigma_{t-1}, \\ \dfrac{\sigma_{t-1}}{\mu_{t-1} - d_{t-1}}, & \text{otherwise}, \end{cases} \tag{49}$$

where μ_t and σ_t^2 denote mean and variance defined in (48). In the administration at time instant t, first select $\lceil \phi_t K \rceil$ test takers according to Selection Policy 1, then select the remaining $K - \lceil \phi_t K \rceil$ test takers according to Selection Policy 2, where $\lceil \cdot \rceil$ denotes the ceiling operation.

In words, ϕ_t is the reciprocal of the number of standard deviations d_{t-1} falls short of its mean. Note that, as long as d_{t-1} remains within one standard deviation, the mixed policy reduces to the detection-greedy policy.

[2] It will be shown later that high detection rates can arise as a side effect of information-greedy flagging.

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

Instead of fixing the shares of test takers selected according to each policy, one can also use a randomized strategy, where test takers are selected sequentially and selection policies 1 and 2 are used with probabilities ϕ and $1-\phi$, respectively.

Finally, note that the distribution of $\sum_{n=1}^{N} D_{t,n}$ is known, namely, it is a Poisson binomial distribution with parameters $q_1, \ldots, q_K$ (Tang & Tang, 2023). Therefore, the probability of observing a certain number of critical group members can be calculated exactly, and can be used to refine Selection Policy 3. However, given the heuristic nature of the proposed selection policies, such refinements are unlikely to be worth the extra computational cost.

5. Performance bounds

Before investigating the performance of the proposed methods numerically, it is useful to establish theoretical bounds. In this section, two such bounds are presented. The first bound is obtained by assuming the model (P_0, P_1, and r) to be known. In this case, the problem in (8) can be solved exactly. The second result provides a simple, distribution-independent bound on the DR of any selection policy in terms of N, K, and r.

Assuming P_0, P_1, and r to be known eliminates the need to learn the model and, in turn, decouples different test administrations. Therefore, the bounds in this section are given for a single administration, and the respective subscripts are omitted.

For the first bound, we establish a connection between the error probabilities of an optimal test for P_0 against P_1 and the objective function in (8).

Lemma 1. *For any P_0, P_1, and r, it holds that*

$$E\left[\sum_{n=1}^{n} D_n\right] \leq \sum_{k=0}^{K-1} E_{\boldsymbol{X}}\left[E\left[C_{(N-k)} \,|\, \boldsymbol{X}\right]\right], \tag{50}$$

where $E\left[C_{(n)} \,|\, \boldsymbol{X}\right]$ denotes the nth-order statistic of $E\left[C_1 \,|\, X_1\right], \ldots, E\left[C_N \,|\, X_N\right]$. It holds that

$$E_{\boldsymbol{X}}\left[E\left[C_{(N-k)} \,|\, \boldsymbol{X}\right]\right] = 1 - \sum_{i=0}^{k} w_{N,k,i}\, V_{N-i,r}(P_0, P_1), \tag{51}$$

where

$$w_{N,k,i} \coloneqq \frac{N-k}{N-i}\binom{N}{k}\binom{k}{i}(-1)^{k-i} \tag{52}$$

and

$$V_{n,r}(P_0, P_1) \coloneqq E_{\Lambda}\left[\left(r\beta(\Lambda) + (1-r)\,(1-\alpha(\Lambda))\right)^n\right]. \tag{53}$$

Here, Λ is a random variable that follows a shifted hyperbolic secant distribution, with PDF

$$p_{\Lambda}(\lambda) = \frac{1}{2}\operatorname{sech}\left(\frac{\pi}{2}(\lambda + \operatorname{logit}(r))\right), \tag{54}$$

and α and β are the error probabilities of a likelihood ratio test for P_0 against P_1:

$$\alpha(\lambda) \coloneqq P_0\left[\log\frac{p_1(X)}{p_0(X)} > \lambda\right], \quad \beta(\lambda) \coloneqq P_1\left[\log\frac{p_1(X)}{p_0(X)} \leq \lambda\right]. \tag{55}$$

The lemma is proven in Appendix B. It is of interest since it provides a way of decomposing the cost function on the left-hand side of (50). The feature distributions, P_0 and P_1, and the share of critical group members, r, only enter (51) via $V_{n,r}$. Assuming that sufficiently many terms $V_{1,r}, V_{2,r} \ldots$ have been calculated, the effect of changes in N and K can be investigated by merely changing the weights

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

$w_{N,k,i}$, which can be calculated efficiently for realistic administration sizes. If the effects of changing P_0 and P_1 are of interest, evaluating the bound in (50) is computationally expensive since all $V_{n,r}$ have to be recalculated or simulated from scratch. In such cases, simply simulating the selection procedure under the distributions of interest is likely to be easier and faster.

The second result in this section is a simple bound on the expected detection *rate* of any selection policy. To derive it, we assume that no detection errors occur or, equivalently, that the selection policy has direct access to $\boldsymbol{C}_t$. In this case, critical group members are missed only if their number in the administration exceeds K. This leads to the following bound:

$$E\left[\frac{\sum_{n=1}^{N} D_{t,n}}{\sum_{n=1}^{N} C_{t,n}}\right] \leq E_Y\left[\min\left\{1, \frac{K}{Y}\right\}\right] = (1-r)^N + \sum_{y=1}^{N} \min\left\{1, \frac{K}{y}\right\} \binom{N}{y} r^y (1-r)^{N-y}, \tag{56}$$

where $Y \sim \text{Bin}(N,r)$, with $\text{Bin}(N,r)$ being the binomial distribution with N trials and success probability r. Note that, in (56), we define the DR to be one if the number of critical group members is zero ("no missed detections").

The bound in (56) is somewhat trivial, but it can be useful to quickly determine lower bounds on K or acceptable ranges for r. For example, assuming that $N = 1{,}000$ and $r = 0.15$, if one seeks to detect at least 90% of critical group members, $K \geq 135$ has to hold irrespective of the feature distributions and the selection policy.

6. Numerical examples

In this section, we present two numerical examples to illustrate the results discussed in the previous sections. The first example uses synthetic data to evaluate the proposed heuristics in a controlled setting. The second uses real-world data, which will be described in more detail below. For context, we compare the proposed procedure to three variants of a reference procedure that combines partial labeling with off-the-shelf classifiers.

In both examples, the marginal feature distributions in (26) are assumed to be beta distributions, that is,

$$h_m(x) = \frac{1}{x(1-x)}, \qquad T(x) = \begin{bmatrix} \log x \\ \log(1-x) \end{bmatrix}, \qquad A_m(\eta) = \log B(\eta_1, \eta_2) \tag{57}$$

for all $m = 1, \ldots, M$, where B denotes the beta function. Every η_m is assumed to have the same prior, namely,

$$\pi(\eta) = \frac{\gamma}{B(\eta_1, \eta_2)} e^{-(\eta_1 + \eta_2)}, \tag{58}$$

where $\gamma \approx 0.4877$. The prior of $\boldsymbol{\Psi}$ is assumed to be $\mathcal{W}^{-1}(1, \boldsymbol{I}_M)$, and the prior of R is assumed to be $\text{Beta}(1,1)$. These choices correspond to initial parameters $\boldsymbol{\nu}^{(0)} = -\chi_m^{(0)} = (1,1)$ and $\boldsymbol{\Psi}^{(0)} = \boldsymbol{I}_M$ and are least informative in the sense discussed in Section 4.1.

To clarify, the proposed method is by no means limited to beta marginals. The main reason for using them here is that in the application that motivated this work, plagiarism detection in a large language test, the feature vectors consist of various text similarity metrics ranging from zero (no similarity) to one (identical texts). These similarity metrics were empirically found to be approximately beta distributed. Beyond this particular application, beta distributions are often a natural choice when the features themselves can be interpreted as probabilities. This is the case, for example, when fusing the output of multiple classifiers.

In both examples, we used the same priors for test takers in the critical and reference groups. That is, no prior knowledge about the feature distributions was assumed. This choice was made for two reasons: First, it corresponds to a worst-case scenario of no prior knowledge and in turn yields conservative performance estimates. Second, our goal was to evaluate how well the proposed method learns from

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

data and leverages review outcomes. This is most apparent when the model is learned from scratch. In practice, of course, prior information can and should be incorporated into the model.

Python code to replicate the experiments in this section is available at Fauss (2025).

6.1. Reference procedure

As discussed in Section 1, the sequential semi-supervised flagging problem addressed in this article is non-standard, and none of the methods found in the literature met all our requirements. Thus, there is no obvious reference procedure to compare the proposed one to.

In order to at least establish a performance baseline, we propose combining off-the-shelf classifiers with *pseudo labeling* (PL), a generic method for learning from partially labeled data. PL was proposed by Lee (2013) and is still widely used today (Kage et al., 2025; Zhai et al., 2019). Its idea is to use a model trained on labeled data to assign labels to unlabeled data. The predicted labels, called pseudo-labels, are then treated as if they were true and used to train the model alongside the original, labeled samples. To reduce noise, pseudo-labeling is usually applied selectively, for example, by only assigning pseudo-labels to samples that were classified with high confidence.

In principle, PL can be combined with any classifier. However, given the complications of the flagging problem, we consider classifiers with the following properties:

1. The classifier outputs *soft labels* (probabilities or scores) that allow samples to be ordered from least to most likely to belong to the critical group.
2. The classifier can be trained via *stochastic gradient descent* (SGD) (or a variant thereof), which naturally lends itself to learning from sequential data.
3. The classifier can operate in a *small-to-medium sample size* regime, minimizing the need for pretraining before use in operation.

Algorithm 2 Generic flagging procedure using SGD and PL.

for $t = 1, \ldots, T$ **do**
    Calculate critical group probability/score for each test taker in administration t
    Select and review K test takers according to chosen policy
    Perform SGD step on labeled data (reviewed test takers)
    Recalculate probabilities/scores using updated classifier
    Pseudo-label unreviewed test takers with sufficiently indicative probabilities/scores
    Perform SGD step on pseudo-labeled data
end for

Note that some advanced methods that are routinely used in more standard settings do not meet the requirements stated above. For example, deep neural networks typically require large datasets to reach their full potential (Sun et al., 2017); many ensemble methods, such as random forests and boosting, are inherently designed for static datasets and not well-suited for sequential learning without substantial modification (Dietterich, 2000; Saffari et al., 2009); and methods such as k-nearest-neighbors or decision-tree classifiers require additional, often sample-hungry calibration procedures to convert hard labels to soft labels (Niculescu-Mizil & Caruana, 2005; Zadrozny & Elkan, 2002).

For the experiments in this section, we used three different classifiers: logistic regression, a linear support vector machine, and a fully-connected neural network with a single hidden layer and rectified linear unit (ReLU) activation function. These classifiers are well-established, highly popular among practitioners, and efficient implementations are readily available in various programming languages. The resulting procedures are referred to as PL-LR, PL-SV, and PL-NN, respectively. The underlying algorithm for combining PL with SGD-based classifiers is summarized in Algorithm 2.

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

All three reference procedures were implemented via the scikit-learn framework (Pedregosa et al., 2011). PL was implemented as follows: During the first 10 administrations, only reviewed samples were used for training. After this warm-up period, the threshold for PL positive samples (critical group members) was set to the lowest predicted probability or score among the confirmed positive samples in the current administration. For example, if three cases were identified as critical group members during the review, and the classifier had predicted these outcomes with probabilities 0.75, 0.8, and 0.9, then all samples whose predicted probability of belonging to the critical group exceeds 0.75 were assigned a positive pseudo-label. Negative pseudo-labels (reference group members) were assigned analogously.

The reference procedures were combined with a detection-greedy flagging policy, that is, the K samples with the highest likelihood of belonging to the critical group were selected for review. We chose this policy because it can be implemented irrespective of whether a classifier outputs probabilities or scores and, as will be shown shortly, the proposed Bayesian procedure typically performed best with a detection-greedy policy.

In addition to the PL and selection rules, all three classifiers have hyper-parameters that need to be set. However, in the sequential settings of the flagging problem, hyper-parameter optimization is non-trivial and requires custom methods (Gama et al., 2014) or substantial modifications of methods for static datasets (Shumway & Stoffer, 2022). Since these complications are well beyond the scope of this article, we decided to set the hyper-parameters by directly optimizing the observed performance. This is clearly not possible in practice, but provides useful insights into the *potential* performance of different methods.

Finally, we focused on tuning two important hyper-parameters: the learning rate used in the SGD steps, which is critical for balancing fast learning with good steady-state performance, and the width of the hidden layer of the neural network. The remaining hyper-parameters were left at their default values in scikit-learn (version 1.6.1). The learning rate was optimized via a grid-search over the interval $[10^{-5}, 10^{1}]$ with 61 logarithmically-spaced grid-points. The width of the hidden network layer was optimized via a grid search over the interval $[M, 10M]$ in steps of $\lfloor \frac{M}{2} \rfloor$, where $\lfloor \bullet \rfloor$ denotes the flooring operation. Results are reported for the parameter values that, averaged over 20 runs, showed the best performance. More details will be discussed in the course of this section.

While clearly not exhaustive, we believe that the three PL procedures detailed above provide a useful baseline for how well the flagging problem can be solved by combining off-the-shelf components.

6.2. Synthetic data

For the first example, we consider $T = 100$ test administrations with $N = 100$ test takers each, approximately 20% of which are members of the critical group ($r = 0.2$). The selection is made based on $M = 10$ features that admit beta-distributed marginals with the following, randomly generated parameters:

$$\begin{bmatrix} \boldsymbol{\eta}_0^\top \\ \boldsymbol{\eta}_1^\top \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}_0^\top \\ \boldsymbol{\beta}_0^\top \\ \boldsymbol{\alpha}_1^\top \\ \boldsymbol{\beta}_1^\top \end{bmatrix} = \begin{bmatrix} 1.73 & 3.16 & 2.46 & 2.58 & 3.08 & 3.02 & 2.10 & 5.27 & 2.78 & 2.24 \\ 1.56 & 2.93 & 2.80 & 3.28 & 3.89 & 5.10 & 1.82 & 1.92 & 2.58 & 3.25 \\ 2.50 & 3.84 & 3.72 & 2.95 & 3.54 & 3.33 & 2.16 & 2.95 & 2.47 & 3.77 \\ 1.75 & 3.49 & 2.07 & 2.43 & 2.75 & 4.02 & 3.36 & 3.71 & 2.80 & 3.08 \end{bmatrix}. \tag{59}$$

For members of the reference group, the features were assumed to be uncorrelated, $\boldsymbol{\Sigma}_0 = \boldsymbol{I}$. For members of the critical group, the correlations were modeled as $\left[\boldsymbol{\Sigma}_1\right]_{ij} = 0.9^{|i-j|}$.

Figure 3 shows the average DR of the selection procedure as a function of the number of administrations. Note that DR here refers to the accumulated fraction of detected members of the critical group, that is,

$$\mathrm{DR}(t) \coloneqq \frac{\sum_{\tau=1}^{t} \sum_{n=1}^{N} D_{\tau,n}}{\sum_{\tau=1}^{t} \sum_{n=1}^{N} C_{\tau,n}}. \tag{60}$$

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

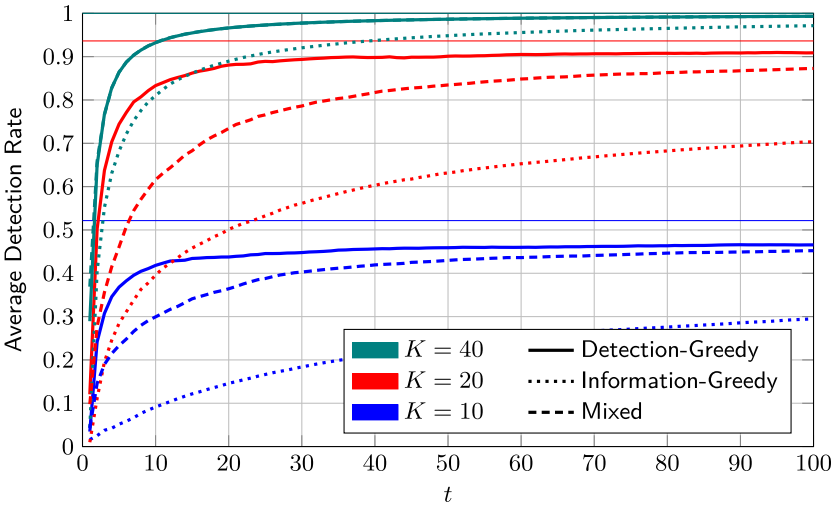


Figure 3. Average accumulated detection rate of the proposed policies against the number of administrations for different review sizes *K*.
Note: Here, $N = 100$, $r = 0.2$, and the parameters of the feature distributions are given in (59). The thin horizontal lines indicate the upper bound on the detection rate in (56).

The results were averaged over 20 runs, and identical feature distributions were used for every run. By inspection of Figure 3, it can be seen that the detection-greedy policy performs best in this example, irrespective of the number of reviewed test takers. The DR of the mixed policy, which balances detection and model improvement, first grows significantly more slowly than that of the detection-greedy policy, but ultimately converges to approximately the same value. This means that the mixed policy first focuses on learning the model, at the expense of a lower DR. However, once the model is sufficiently accurate, it performs just as well or even better than the detection-greedy policy. In contrast, the information-greedy policy shows significantly lower DRs across all experiments. This is expected, as it is not designed to maximize the DR. In fact, it might be somewhat unexpected that the DR does not converge to 50%. In this example, this is the case because the entropy of the posterior group distribution tends to be higher for members of the critical group than for members of the reference group. As a consequence, the information-greedy policy flags more test takers from the former. In general, we observed that the information-greedy policy tends to flag more members of the *smaller* group, which is the critical group in this case.

From Figure 3, one may draw the conclusion that the mixed selection policy is redundant, given that it performs worse than the simpler detection-greedy policy. However, this result does not convey the whole picture. First, for both $K = 10$ and $K = 20$, the mixed policy only performs worse than the detection-greedy policy in the early test administrations, with the largest gap between the two occurring at around $t \approx 10$. As t increases, the gap between the two curves narrows, meaning that the mixed policy performs equally well or even better for virtually all later administrations. Consequently, the mixed policy can be preferable in cases where the first administrations are less critical and can be used primarily for data collection. For example, new items are commonly piloted before being used in high-stakes tests (von Davier, 2010).

A second advantage of the mixed policy is that it leads to a more accurate model. Figure 4 shows plots of the average mean squared error (MSE) of two parameters, R and $\boldsymbol{\Sigma}_1$, over the number of administrations. The MSE was calculated with respect to the posterior distribution at the time and again averaged over 20 runs. For all three policies, the large initial model errors steadily decrease, meaning the procedure does indeed learn the model from the data. However, the information-greedy and the mixed policy learn the parameters faster. Consider, for example, the MSE of R in the top plot of Figure 4. While it takes the detection-greedy policy approximately 75 administrations to reach −40dB,

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

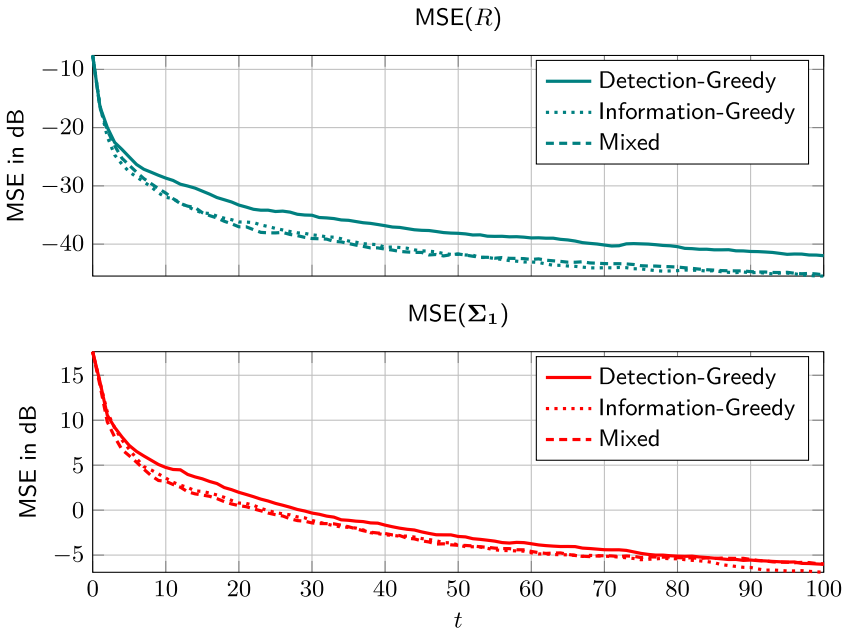


Figure 4. Average MSE against the number of administrations.
Note: Here, $N = 100$, $r = 0.2$, $K = 20$, and the parameters of the feature distributions are given in (59).

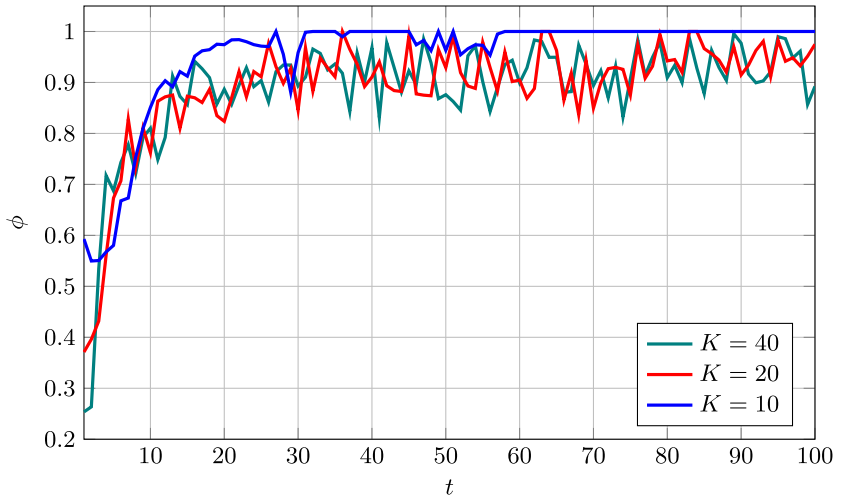


Figure 5. Average values of ϕ_t in (49) against the number of administrations.
Note: Here, $N = 100$, $r = 0.2$, $K = 20$, and the parameters of the feature distributions are given in (59).

the information-greedy and mixed policy only need half as many samples to reach the same error level. For the error in the covariance matrix estimate, shown in the lower plot of Figure 4, this difference is less pronounced, but still noticeable. For example, the information-greedy and mixed policy reach the 0dB level approximately 10 administrations or 1000 samples earlier than the detection-greedy policy.

Finally, some insight into the mixed selection policy can be gained by observing ϕ_t, defined in (49). Its value, averaged over 20 runs, is plotted against the number of administrations in Figure 5. It can be seen that initially a substantial share of samples is selected according to the information-greedy policy, which is in line with the quicker model fit/learning discussed above. However, after approximately

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

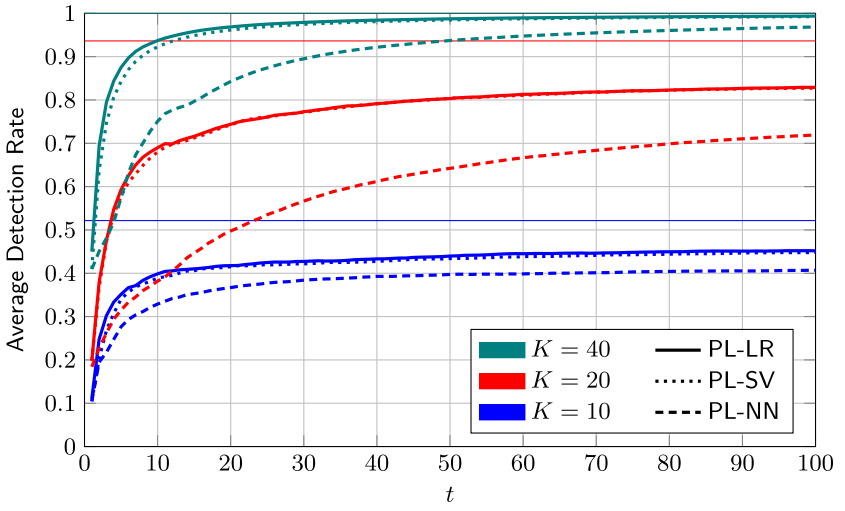


Figure 6. Average accumulated detection rates of reference procedures against the number of administrations for different review sizes *K*.
Note: Here, $N = 100$, $r = 0.2$, and the parameters of the feature distributions are given in (59). The thin horizontal lines indicate the upper bound on the detection rate in (56).

30 administrations, ϕ_t reaches a steady state in which it fluctuates around approximately 0.95, meaning that most reviewed samples are selected according to the detection-greedy policy. Interestingly, the observed ϕ_t curves are almost identical for $K = 20$ and $K = 40$. Only for $K = 10$ the behavior is slightly different, with ϕ_t being strictly one for most administrations. This is likely a consequence of the second-order moment approximation used to calculate ϕ_t, which becomes more accurate as K increases. Nevertheless, the observed quick convergence to a purely detection-greedy policy is beneficial in this case since for $K = 10$ the procedure operates in a regime where the number of reviews, not the model fit, constitutes the main bottleneck.

The results discussed so far illustrate the trade-off between DR and model accuracy. On the one hand, the model needs to be "good enough" to reliably detect critical group members. On the other hand, investing too many samples into model improvements can lead to situations in which an unreasonable number of administrations is required to compensate for the lower initial DR. Based on our experiments, we conjecture that, as a rule of thumb, in cases where a high DR is the main or only goal, the detection-greedy policy should be used. In cases where early administrations can be used for training, or when an accurate model is of concern, for example, to track error rates, feature correlations, or shares of critical group members, the mixed policy can be preferable.

To put the results shown so far into perspective, average DRs of the three reference procedures detailed in Section 6.1 are shown in Figure 6. The corresponding hyper-parameters are given in Table 1. By inspection, PL-SV and PL-LR achieve virtually identical performances for all values of K. PL-NN performs notably worse. We conjecture that this is the case because neural networks have significantly more free parameters and in turn require more training data. In this example, even the smallest neural network, with a hidden layer of width 30, has well over 300 free parameters. For comparison, the proposed Bayesian method has approximately 60 free parameters, and PL-LR and PL-SV both have fewer than 20.

In Figure 7, we compare one of the best-performing baseline methods, PL-SV, to the detection-greedy variant of the proposed method. For large and small values of K, there are no significant performance differences. For $K = 10$, both procedures appear limited mainly by the small number of reviews rather than their respective flagging or learning strategies. At the other end, for $K = 40$, there are enough labeled data points for both to quickly achieve virtually perfect separation. Only for $K = 20$ does the proposed method significantly outperform PL-SV. We conjecture that in this case, in which the share

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

Table 1. Hyper-parameters of reference procedures for different values of K

Procedure	Hyper-parameter	$K = 10$	$K = 20$	$K = 40$
PL-LR	Learning rate	1.5	0.6	0.4
PL-SV	Learning rate	1.2	0.4	1.0
PL-NN	Learning rate	0.1	0.01	0.06
	Hidden layer width	30	70	75

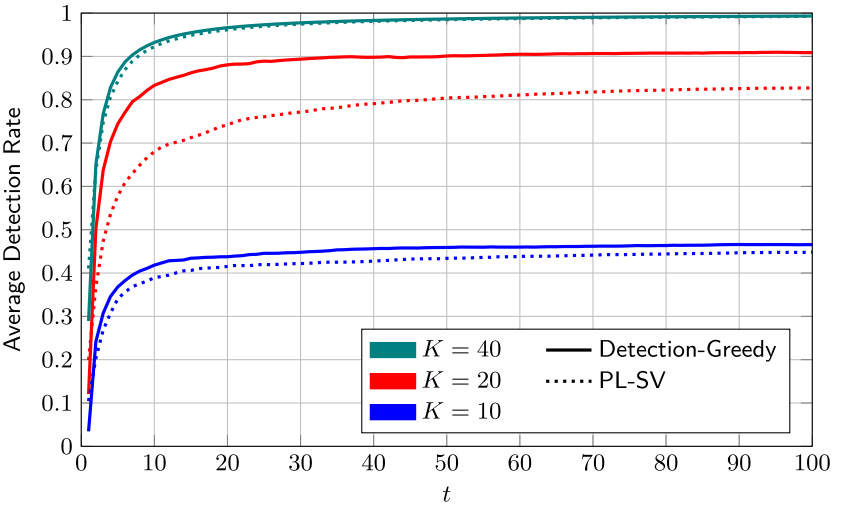


Figure 7. Comparison of average accumulated detection rates of the proposed detection-greedy policy and support-vector-based reference procedure.
Note: Here, $N = 100$, $r = 0.2$, and the parameters of the feature distributions are given in (59).

of reviewed cases equals the rate of critical group members ($\frac{K}{N} = r$), the proposed procedure's more principled approach to data processing and model tracking pays off. Arguably, the case $\frac{K}{N} \approx r$ is also of practical relevance, as one would normally expect to review about as many cases as there are critical group members.

6.3. Real-world data

For the second example, we used real-world data collected from a large language assessment between November 2022 and March 2023. The dataset contains 2,595 data points, each a three-dimensional vector representing different similarity metrics between a submitted response and a potential source text it may have been copied from. Of the 2,595 responses, 745 (≈30%) were labeled as plagiarized (true positives) by human experts, and 1,850 (≈70%) as authentic (false positives), meaning the similarities did not constitute plagiarism under the program's guidelines.[3] The three similarity metrics used in this example all take values in the unit interval, with one corresponding to identical texts. For security reasons, we do not detail how the features were computed from the responses or which items were used for the study. For the purpose of illustrating the proposed flagging procedure, these specifics are of little importance. Interested readers may contact the authors directly.

An additional complication compared to the first example is introduced by the fact that, owing to pre-processing steps that will not be detailed here, the elements of the similarity vectors are left-

[3]Note that the ≈30% positive rate is based on a prefiltered sample and does not reflect operational rates.

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

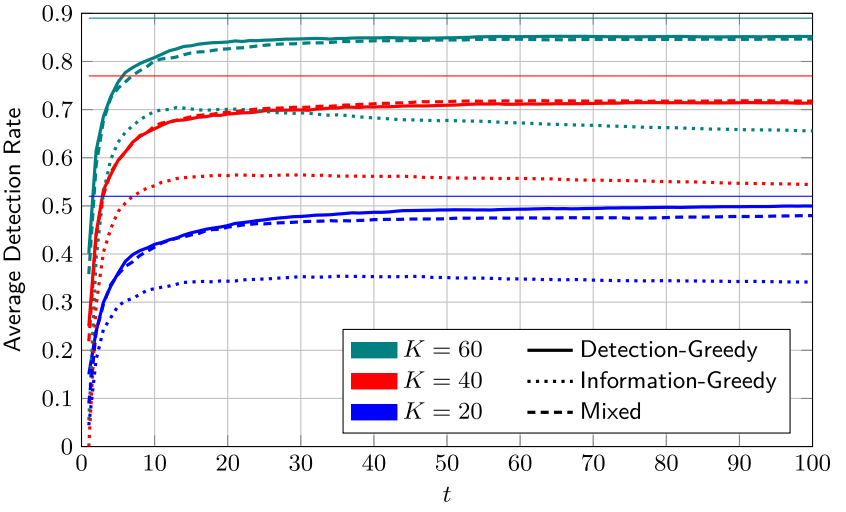


Figure 8. Average detection rate of the proposed policies against the number of administrations for different review sizes *K*.
Note: Here, $N = 100$, $M = 3$, and $r = 0.2$. The thin horizontal lines indicate upper bounds on the detection rate obtained by an "oracle" version of the proposed method that uses a fitted model from the beginning instead of learning it from the data.

truncated at 0.13, 0.7, and 0.3, respectively. In principle, the proposed algorithm can be adapted to use truncated exponential distributions. However, this extension is non-trivial and beyond the scope of this article. Instead, in this example, we disregarded the model mismatch and ran the flagging procedure with the same marginal distributions and parameters as in the first example. This approach is clearly suboptimal. However, we still consider the experiment meaningful since, in practice, model mismatches are inevitable and the flagging procedure should still perform acceptably well under such conditions.

We again used $N = 100$, $r = 0.2$, and $T = 100$. The number of features, $M = 3$, was determined by the data. Samples from the critical and reference groups were generated via bootstrapping. That is, for each administration, we first sampled a vector $\boldsymbol{c}_t$ whose elements are independent Bernoulli random variables with success probability r. For each $c_{n,t}$, the corresponding feature vector $\boldsymbol{x}_t$ was then drawn uniformly and with replacement from the respective data set. In order to investigate the effect of the model mismatch, we fitted two multivariate truncated beta distributions to the true and false positives in our data set. These distributions were defined in analogy to those in Section 3.1, that is, the marginals are truncated beta distributions and the correlations are modeled via a Gaussian copula. This model largely eliminates the model mismatch and, in line with the assumptions underlying Lemma 1, was used to simulate the case in which the distributions P_0 and P_1 are known.

In Figure 8, the average DRs of the three proposed policies are plotted against the number of administrations for different values of K. For comparison, we included bounds obtained by assuming P_0 and P_1 to be known, as explained above. In general, the DR curves behave similar to their counterparts in the first example, which corroborates the findings obtained from synthetic data. Interestingly, in the sample-rich regime ($K = 60$), the performance of the information-greedy policy is on par with that of the detection-greedy and mixed policies. We conjecture that this is the case because samples from the smaller group tend to be more informative. For $K = 60$, the sample size is large enough for the policy to "saturate" the samples from the critical group.

Next, we compare the model fit of the different policies. In lack of a ground truth for the other parameters, we look at the posterior mean of R. As can be seen in Figure 9, for all policies, the estimated rate of critical group members starts high, then quickly drops, and finally converges to a value close to 0.2. This behavior can be explained as follows: Since R is assumed to have been drawn from a uniform prior, the initial mean is 0.5, which is significantly larger than the true value of 0.2. As samples start coming in, the model adjusts to the lower observed rate of critical group members. However, for the first

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

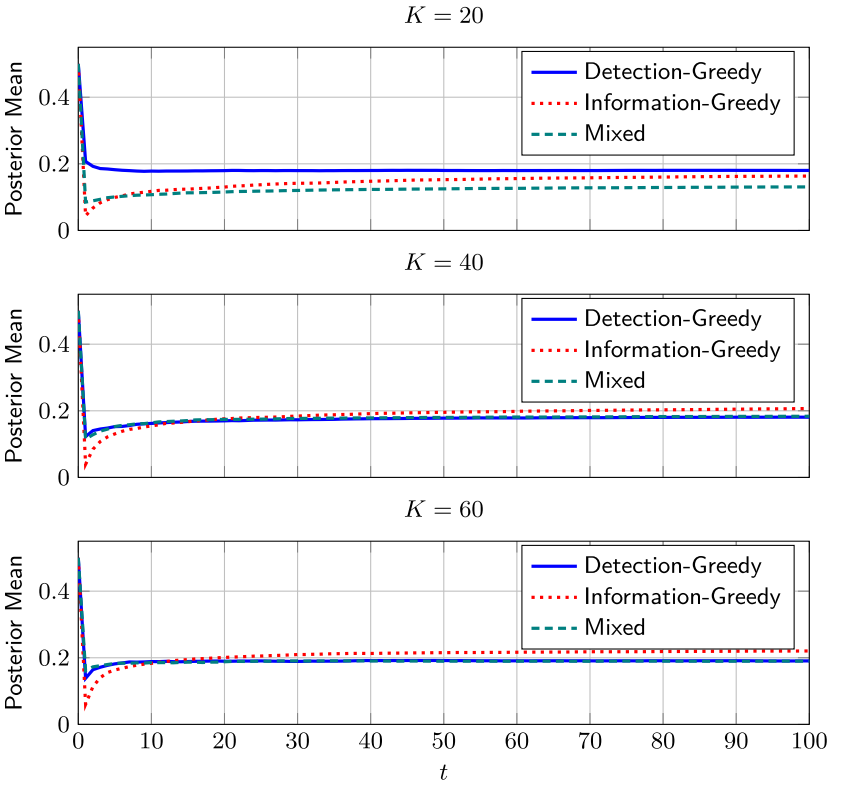


Figure 9. Average posterior mean of R against the number of administrations for different selection policies and values of K, with $N = 100$ and $M = 3$.

administrations, the uncertainty in the model is large, and the detection accuracy is still low. Therefore, only a fraction of the critical group members is detected, and the model goes from overestimating r to underestimating it. This overshoot effect is least pronounced for the detection-greedy policy, which has the highest DR in early administrations (compare Figure 8). We conjecture that this effect also contributes to the higher overall accuracy of the detection-greedy policy when $K = 20$.

The average DRs of the three reference procedures detailed in Section 6.1 are shown in Figure 10, and the corresponding hyper-parameters in Table 2. By inspection, PL-NN performs significantly worse than PL-SV and PL-LR. The latter admit virtually identical performance, with PL-LR performing slightly better for $K = 20$. This finding corroborates that neural networks do not seem to be a good fit for the sequential flagging problem: large networks require too many samples to learn the underlying model, and small networks lack the required expressive power. This problem could possibly be addressed by adaptively growing the network as more data become available, but exploring this approach is well beyond the scope of this section.

In Figure 11, we compare the proposed detection-greedy policy to the best-performing reference method, PL-LR. Interestingly, the latter slightly outperforms the former for larger values of K. We conjecture that this performance gap is mainly caused by the model mismatch discussed above. On the one hand, this result shows that, unsurprisingly, the proposed procedure is not necessarily the best choice under all circumstances. On the other hand, it is reassuring to see that even under moderate model mismatch, the proposed method admits DRs within three percentage points of the best reference method, and accurately estimates the share of critical group members among the test population (compare Figure 9). Also, note that PL-LR uses a close-to-optimal learning rate, which is unlikely to be the case in practice.

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

Table 2. Hyper-parameters of reference procedures for different values of K

Procedure	Hyper-parameter	$K = 10$	$K = 20$	$K = 40$
PL-LR	Learning rate	8.0	1.5	1.5
PL-SV	Learning rate	5.0	2.0	3.0
PL-NN	Learning rate	0.002	$1.25 \cdot 10^{-5}$	0.5
	Hidden layer width	5	3	5

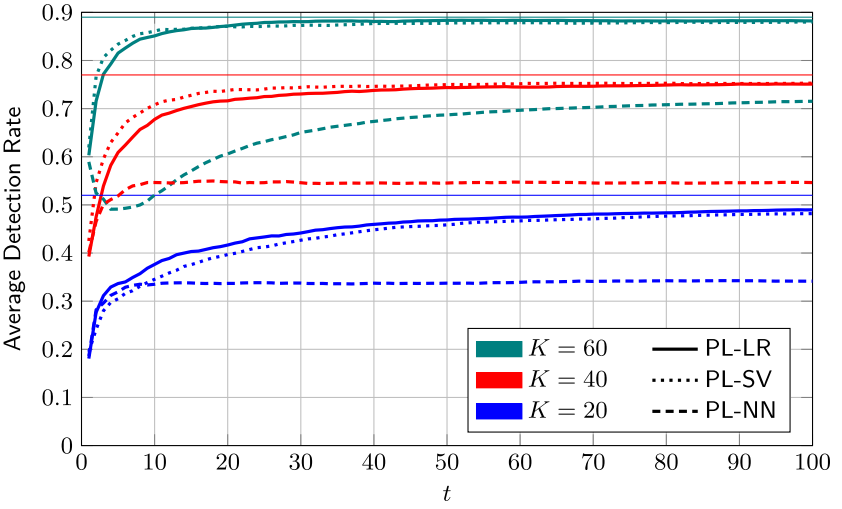


Figure 10. Average detection rate of reference procedures against the number of administrations for different review sizes K.
Note: Here, $N = 100$, $M = 3$, and $r = 0.2$. The thin horizontal lines indicate upper bounds on the detection rate obtained by an "oracle" version of the proposed method that uses a fitted model from the beginning instead of learning it from the data.

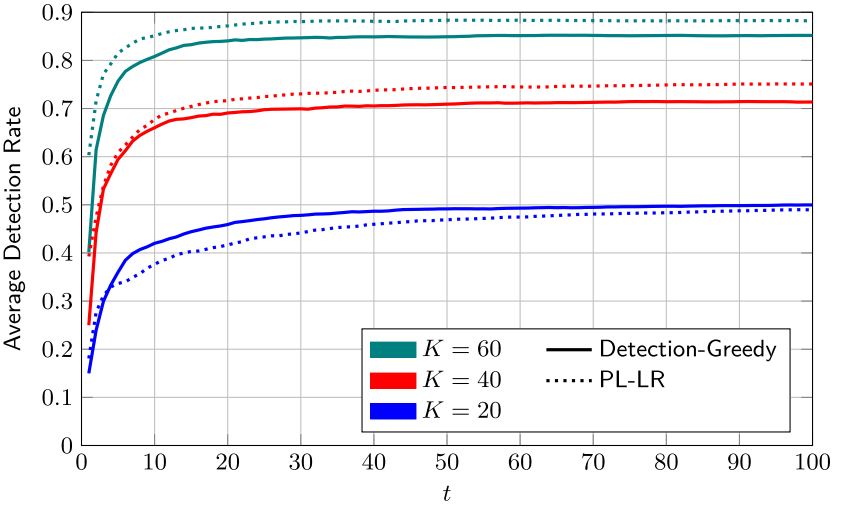


Figure 11. Comparison of average accumulated detection rates of the proposed detection-greedy policy and logistic-regression-based reference procedure.
Note: Here, $N = 100$, $M = 3$, and $r = 0.2$.

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

6.4. Summary

In summary, we conclude from the experiments in this section that the main strengths of the proposed procedure are as follows:

Simplicity: The proposed procedure can be used without modifications for all sample sizes, time horizons, numbers of conducted reviews, feature dimensions, etc. In contrast, many alternative approaches, including the ones presented here, require the user to tune various hyper-parameters. For example, as can be seen from Tables 1 and 2, the optimal learning rates vary widely between different scenarios and can be far from common defaults, such as 0.01 (TensorFlow Contributors, 2025) or 0.001 (PyTorch Contributors, 2025).

Robustness: While not always being the best choice, the proposed procedure performed well in all of our experiments, even under moderate model mismatch. In combination with its simplicity, we believe that this makes it an excellent choice in practice.

Fully Bayesian model: At any point in time, the user has access to the approximate posterior distributions of the quantities of interest and can perform additional inference as desired. While not always necessary, this allows for a much more systematic analysis of both the flagging decisions and the test taker population. Moreover, in particular when using the mixed policy, the procedure self-calibrates so that predicted probabilities can validly be interpreted as such.

As potential weaknesses, we identified the following issues:

Model mismatch: While our experiments showed that the procedure can handle moderate model mismatch, it will break down under severe model mismatch. This means that the user should not apply the procedure blindly, but have some familiarity with the data of interest. For example, a simple sanity check is to verify that the assumed exponential family distributions roughly fit the true marginal distributions of the features.

Compute: In our experiments, the proposed procedure typically took more time to run than the alternatives. This is partly due to its higher complexity and partly because our implementation is not as optimized as those provided by mature machine learning libraries. However, the time required to process a single administration averaged around 2s for $M = 10$ and 0.5s for $M = 3$, making it real-time capable for most practical purposes.

Numerical stability: The proposed procedure requires solving the equation system in (37)–(41). In our experiments, a simple fixed-point iteration converged most of the time. When it did not (fewer than 10 occasions across all experiments), the output of the last iteration was used. The effect of these inaccuracies appears negligible, but, in principle, one should be aware of potential numerical issues.

7. Conclusions and outlook

In this article, we have addressed the problem of automatically flagging test takers who display atypical responses or behaviors for further review by human experts. The primary goal has been to develop a selection policy that efficiently identifies individuals who require additional scrutiny, while also keeping the volume of reviews per test administration manageable. To achieve this, the flagging problem has been formulated as a semi-supervised learning task in a Bayesian framework. The corresponding optimal selection policy has been derived and discussed. Given the computational challenges in implementing this policy and updating the underlying posterior distributions, we have proposed a variational approximation along with three heuristic selection policies, each balancing exploration and exploitation in different ways. Through numerical experiments using both synthetic and real data, the performance of these approximate policies has been evaluated and compared with reference procedures based on off-the-shelf techniques.

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

Beyond the results presented in this article, there are variations and extensions of the proposed procedure that will be considered for future research. Some of these have already been mentioned in the text, but are included here for completeness.

- Assuming the number of reviews to be given and fixed has been convenient for the analysis in this article. In practice, however, other criteria can be more appropriate, such as the detection or false alarm rate. While such variations of the underlying problem formulation will lead to a conceptually similar procedure, we still expect that insights can be gained from explicitly deriving the corresponding selection policies.
- In some applications, it might not be appropriate to assume that the feature distributions and the share of critical group members remain constant between test administrations. For such cases, a variant of the proposed method is needed that tracks the changing parameters over time. Problems of this kind are known as (Bayesian) filtering and have been studied extensively in the literature (Särkkä & Svensson, 2023; Sayed, 2011).
- The flagging problem can be formulated under different assumptions about the available information. For example, the problem simplifies significantly if one replaces $p(c_{t,n} \mid \mathcal{F}'_{t-1})$ with $p(c_{t,n} \mid X_{t,n}, \mathcal{F}_{t-1})$, that is, the decision whether to flag the nth test taker in the tth administration is restricted to depend only on the features extracted from the response under scrutiny. On the other hand, when tests are administered frequently, one may use data collected in multiple administrations, say t, $t+1$, and $t+2$, to flag test takers in administration t, thus complicating the data fusion step.
- In this article, the population of test takers is assumed to be homogeneous, meaning that the feature distribution is identical for all test takers. This is not necessarily the case in practice, where different subgroups of test takers can admit different feature distributions. For example, in language tests, the feature distribution might be affected by a test taker's first language. Similarly, in writing assessments, certain styles or structures can be common among subgroups of test takers that have the same educational background or used the same preparation material. Not taking such differences into account naturally leads to fairness concerns. A common approach to promote fairness is to add a constraint on an appropriate fairness metric. We expect this problem to be a non-trivial extension of the work in this article. For example, it significantly complicates the exploration/exploitation trade-off since, depending on the chosen fairness metric, various groupwise error probabilities need to be tracked and balanced. However, if implemented successfully, such a procedure would enable one to monitor fairness in real time, based on real-world data.

Funding statement. This research received no specific grant funding from any funding agency, commercial, or not-for-profit sectors.

Competing interests. The authors have no competing interests to declare.

References

Adams, R. P., & Ghahramani, Z. (2009). Archipelago: Nonparametric Bayesian semi-supervised learning. In *International conference on machine learning (ICML)* (pp. 1–8). https://doi.org/10.1145/1553374.1553375

Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. *Journal of Network and Computer Applications*, *60*, 19–31. https://doi.org/10.1016/j.jnca.2015.11.016

Bielza, C., & Larrañaga, P. (2020). Non-probabilistic classifiers. In *Data-driven computational neuroscience* (pp. 262–319). Cambridge University Press. https://doi.org/10.1017/9781108642989.011

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, *112*(518), 859–877. https://doi.org/10.1080/01621459.2017.1285773

Booth, B. M., Bosch, N., & D'Mello, S. K. (2023). Engagement detection and its applications in learning: A tutorial and selective review. *Proceedings of the IEEE*, *111*(10), 1398–1422. https://doi.org/10.1109/jproc.2023.3309560

Bruce, R. F. (2001). A Bayesian approach to semi-supervised learning. In J. Tsujii, & K. Su (Eds.), *Natural language processing pacific rim symposium (NLPRS)*, (pp. 57–64). NLPRS.

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

Bulut, O., Gorgun, G., & He, S. (2024). Unsupervised anomaly detection in sequential process data. *Zeitschrift für Psychologie*, *232*(2), 74–94. https://doi.org/10.1027/2151-2604/a000558

Caruana, R., Karampatziakis, N., & Yessenalina, A. (2008). An empirical evaluation of supervised learning in high dimensions. In *International conference on machine learning (ICML)* (pp. 96–103). https://doi.org/10.1145/1390156.1390169

Cizek, G. J., & Wollack, J. A. (2016). *Handbook of quantitative methods for detecting cheating on tests*. Routledge. https://doi.org/10.4324/9781315743097

Cocea, M., & Weibelzahl, S. (2010). Disengagement detection in online learning: Validation studies and perspectives. *IEEE Transactions on Learning Technologies*, *4*(2), 114–124. https://doi.org/10.1109/tlt.2010.14

David, H. A., & Nagaraja, H. N. (2004). *Order statistics*. Wiley. https://doi.org/10.1002/9781118445112.stat00830

Dempster, A. P., Laird, N. M., & Rubin, D. B. (2018). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, *39*(1), 1–22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

Dietterich, T. G. (2000). Ensemble methods in machine learning. In *International workshop on multiple classifier systems (MCS)* (pp. 1–15). https://doi.org/10.1007/3-540-45014-9˙1

Evanini, K., & Wang, X. (2014). Automatic detection of plagiarized spoken responses. In *Workshop on innovative use of NLP for building educational applications (BEA)* (pp. 22–27). https://doi.org/10.3115/v1/w14-1803

Fauss, M. (2025). *Git repository*. https://github.com/mifauss/bayes-flagger

Felder, R. M., & Brent, R. (2009). Active learning: An introduction. *ASQ Higher Education Brief*, *2*(4), 1–5.

Foltýnek, T., Meuschke, N., & Gipp, B. (2019). Academic plagiarism detection: A systematic literature review. *ACM Computing Surveys*, *52*(6). https://doi.org/10.1145/3345317

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys (CSUR)*, *46*(4), 1–37. https://doi.org/10.1145/2523813

Gomaa, W. H., & Fahmy, A. A. (2013). A survey of text similarity approaches. *International Journal of Computer Applications*, *68*, 13–18. https://doi.org/10.5120/11638-7118

He, Q., Meadows, M., & Black, B. (2022). An introduction to statistical techniques used for detecting anomaly in test results. *Research Papers in Education*, *37*(1), 115–133. https://doi.org/10.1080/02671522.2020.1812108

Jaworski, P., Durante, F., Hardle, W. K., & Rychlik, T. (2010). *Copula theory and its applications* (Vol. 198). Springer. https://doi.org/10.1007/978-3-642-12465-5.

Jiang, Y., Hao, J., Fauss, M., & Li, C. (2024). Detecting ChatGPT-generated essays in a large-scale writing assessment: Is there a bias against non-native English speakers? *Computers & Education*, *217*, 105070. https://doi.org/10.1016/j.compedu.2024.105070

Jiffriya, M., Jahan, M. A., & Ragel, R. (2021). Plagiarism detection tools and techniques: A comprehensive survey. *Journal of Science*, *2*(2), 47–64

Kage, P., Rothenberger, J. C., Andreadis, P., & Diochnos, D. I. (2025). *A review of pseudo-labeling for computer vision*. Preprint. https://arxiv.org/abs/2408.07221

Kettler, R. J. (2012). Testing accommodations: Theory and research to inform practice. *International Journal of Disability, Development and Education*, *59*(1), 53–66. https://doi.org/10.1080/1034912x.2012.654952

Kingston, N., & Clark, A. (2014). *Test fraud: Statistical detection and methodology*. Routledge.

Lee, D.-H. (2013). Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In S. Dasgupta, & D. McAllester (Eds.), *ICML workshop on challenges in representation learning* (Vol. 3, p. 896). Journal of Machine Learning Research (JMLR).

Lu, J., Wang, C., & Shi, N. (2023). A mixture response time process model for aberrant behaviors and item nonresponses. *Multivariate Behavioral Research*, *58*(1), 71–89. https://doi.org/10.1080/00273171.2021.1948815

Marianti, S., Fox, J.-P., Avetisyan, M., Veldkamp, B. P., & Tijmstra, J. (2014). Testing for aberrant behavior in response time modeling. *Journal of Educational and Behavioral Statistics*, *39*(6), 426–451. https://doi.org/10.3102/1076998614559412

Mitra, A. (2016). *Fundamentals of quality control and improvement*. Wiley. https://doi.org/10.1002/9781119692379

Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. In *International conference on machine learning (ICML)* (pp. 625–632). https://doi.org/10.1145/1102351.1102430

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Prajapati, A., Bechtel, J., & Ganesan, S. (2012). Condition based maintenance: A survey. *Journal of Quality in Maintenance Engineering*, *18*(4), 384–400. https://doi.org/10.1108/13552511211281552

PyTorch Contributors. (2025). *torch.optim.SGD*. PyTorch. Retrieved May 30, 2025. https://docs.pytorch.org/docs/stable/generated/torch.optim.SGD.html

Rottmann, M., Kahl, K., & Gottschalk, H. (2018). Deep Bayesian active semi-supervised learning. In *International conference on machine learning and applications (ICMLA)* (pp. 158–164). https://doi.org/10.1109/ICMLA.2018.00031

Saffari, A., Leistner, C., Santner, J., Godec, M., & Bischof, H. (2009). On-line random forests. In *International conference on computer vision workshops (ICCV Workshops)* (pp. 1393–1400). https://doi.org/10.1109/iccvw.2009.5457447

Särkkä, S., & Svensson, L. (2023). *Bayesian filtering and smoothing* (Vol. 17). Cambridge University Press. https://doi.org/10.1017/9781108917407

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

Sayed, A. H. (2011). *Adaptive filters*. Wiley. https://doi.org/10.1002/9780470374122
Shumway, R. H., & Stoffer, D. S. (2022). Using cross-validation methods to select time series models. *British Journal of Mathematical and Statistical Psychology*, *75*(1), 1–20. https://doi.org/10.1111/bmsp.12330
Sireci, S. G., Scarpati, S. E., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research*, *75*(4), 457–490. https://doi.org/10.3102/00346543075004457
Šmídl, V., & Quinn, A. (2006). *The variational Bayes method in signal processing*. Springer. https://doi.org/10.1007/3-540-28820-1
Sun, C., Shrivastava, A., Singh, S., & Gupta, A. (2017). Revisiting unreasonable effectiveness of data in deep learning era. In *International conference on computer vision (ICCV)* (pp. 843–852). https://doi.org/10.1109/iccv.2017.97
Tang, W., & Tang, F. (2023). The Poisson binomial distribution—Old & new. *Statistical Science*, *38*(1), 108–119. https://doi.org/10.1214/22-STS852
Taylor, C., Jamieson, J., Eignor, D., & Kirsch, I. (1998). The relationship between computer familiarity and performance on computer-based TOEFL test tasks. *ETS Research Report Series*, *1998*(1), i–30. https://doi.org/10.1002/j.2333-8504.1998.tb01757.x
TensorFlow Contributors. (2025). tf.keras.optimizers.SGD. TensorFlow. Retrieved May 30, 2025. https://www.tensorflow.org/apidocs/python/tf/keras/optimizers/SGD
Van der Linden, W. J., & Guo, F. (2008). Bayesian procedures for identifying aberrant response-time patterns in adaptive testing. *Psychometrika*, *73*(3), 365–384. https://doi.org/10.1007/s11336-007-9046-8
van der Linden, W. J., & Lewis, C. (2015). Bayesian checks on cheating on tests. *Psychometrika*, *80*, 689–706. https://doi.org/10.1007/s11336-014-9409-x
Van Engelen, J. E., & Hoos, H. H. (2020). A survey on semi-supervised learning. *Machine Learning*, *109*(2), 373–440. https://doi.org/10.1007/s10994-019-05855-6
Varshney, P. K. (2012). *Distributed detection and data fusion*. Springer. https://doi.org/10.1007/978-1-4612-1904-0
von Davier, A. (2010). *Statistical models for test equating, scaling, and linking*. Springer. https://doi.org/10.1007/978-0-387-98138-3
Wang, X., Evanini, K., Bruno, J., & Mulholland, M. (2016). Automatic plagiarism detection for spoken responses in an assessment of English language proficiency. In *Spoken language technology workshop (SLT)* (pp. 121–128). https://doi.org/10.1109/SLT.2016.7846254
Wang, X., Evanini, K., Mulholland, M., Qian, Y., & Bruno, J. V. (2019). Application of an automatic plagiarism detection system in a large-scale assessment of English speaking proficiency. In *Workshop on innovative use of NLP for building educational applications (BEA)* (pp. 435–443). https://doi.org/10.18653/v1/W19-4445
Weisstein, E. W. (2024). Pascal's formula. Retrieved October 25, 2024. https://mathworld.wolfram.com/PascalsFormula.html
Wongvorachan, T. (2023). Cheating detection in tests: A systematic review. https://doi.org/10.31234/osf.io/uc5hj
Yan, D., Fauss, M., Hao, J., & Cui, W. (2023). Detection of AI-generated essays in writing assessments. *Psychological Test and Assessment Modeling*, *65*(1), 125–144.
Zadrozny, B., & Elkan, C. (2002). Transforming classifier scores into accurate multiclass probability estimates. In *International conference on knowledge discovery and data mining (KDD)* (pp. 694–699). https://doi.org/10.1145/775047.775151
Zhai, X., Oliver, A., Kolesnikov, A., & Beyer, L. (2019). S4L: Self-supervised semi-supervised learning. In *2019 IEEE/CVF international conference on computer vision (ICCV)* (pp. 1476–1485). https://doi.org/10.1109/iccv.2019.00156
Zhang, Z., Zhang, J., & Lu, J. (2022). Bayesian analysis of aberrant response and response time data. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.841372

A. Proof of Theorem 1

The theorem is proven inductively using techniques from sequential analysis and dynamic programming. Assume that for some $\tau \leq T$, it holds that

$$\max_{\xi_{1:T}} E\left[\sum_{t=1}^{T}\sum_{n=1}^{N} D_{t,n}\right] = \max_{\xi_{1:\tau}} E\left[\sum_{t=1}^{\tau}\sum_{n=1}^{N} D_{t,n} + \rho_\tau(\boldsymbol{X}_{1:\tau}, \boldsymbol{S}_{1:\tau}, \boldsymbol{D}_{1:\tau})\right], \tag{A.1}$$

with ρ_τ as defined in the theorem. We now show that if this is the case, then the optimal selection policy at time τ is of the form (11). By Assumption 3, the selection policy ξ_τ is $\mathcal{F}'_{\tau-1}$-measurable. It hence holds that

$$\max_{\xi_{1:\tau}} E\left[\sum_{t=1}^{\tau}\sum_{n=1}^{N} D_{t,n} + \rho_\tau(\boldsymbol{X}_{1:\tau}, \boldsymbol{S}_{1:\tau}, \boldsymbol{D}_{1:\tau})\right] \tag{A.2}$$

$$= \max_{\xi_{1:\tau-1}} E\left[\max_{\xi_\tau} E\left[\sum_{t=1}^{\tau}\sum_{n=1}^{N} D_{t,n} + \rho_\tau(\boldsymbol{X}_{1:\tau}, \boldsymbol{S}_{1:\tau}, \boldsymbol{D}_{1:\tau}) \,\middle|\, \mathcal{F}'_{\tau-1}\right]\right] \tag{A.3}$$

$$= \max_{\xi_{1:\tau-1}} E\left[\sum_{t=1}^{\tau-1}\sum_{n=1}^{N} D_{t,n} + \max_{\xi_\tau} E\left[\sum_{n=1}^{N} D_{\tau,n} + \rho_\tau(\boldsymbol{X}_{1:\tau}, \boldsymbol{S}_{1:\tau}, \boldsymbol{D}_{1:\tau}) \,\middle|\, \mathcal{F}'_{\tau-1}\right]\right], \tag{A.4}$$

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

where the inner expectation is taken with respect to the distribution of

$$\boldsymbol{D}_\tau, \boldsymbol{S}_\tau, \boldsymbol{C}_\tau \,|\, \mathcal{F}'_{\tau-1}. \tag{A.5}$$

The probability mass function (PMF) of this distribution is given by

$$\begin{aligned} p(\boldsymbol{d}_\tau, \boldsymbol{s}_\tau, \boldsymbol{c}_\tau \,|\, \mathcal{F}'_{\tau-1}) &= p(\boldsymbol{d}_\tau \,|\, \boldsymbol{s}_\tau, \boldsymbol{c}_\tau)\, p(\boldsymbol{s}_\tau, \boldsymbol{c}_\tau \,|\, \mathcal{F}'_{\tau-1}) && \text{(A.6)} \\ &= p(\boldsymbol{d}_\tau \,|\, \boldsymbol{s}_\tau, \boldsymbol{c}_\tau)\, p(\boldsymbol{c}_\tau \,|\, \mathcal{F}'_{\tau-1})\, \xi_\tau(\boldsymbol{s}_\tau), && \text{(A.7)} \end{aligned}$$

where the last equality follows from (4). The inner expectation on the right-hand side of (A.4) can hence be written as

$$\sum_{\boldsymbol{s}_\tau} \sum_{\boldsymbol{c}_\tau} \sum_{\boldsymbol{d}_\tau} \left(\sum_{n=1}^{N} d_{\tau,n} + \rho_\tau(\boldsymbol{X}_{1:\tau}, [\boldsymbol{S}_{1:\tau-1}, \boldsymbol{s}_\tau], [\boldsymbol{D}_{1:\tau-1}, \boldsymbol{d}_\tau]) \right) p(\boldsymbol{d}_\tau \,|\, \boldsymbol{s}_\tau, \boldsymbol{c}_\tau) p(\boldsymbol{c}_\tau \,|\, \mathcal{F}'_{\tau-1}) \xi_\tau(\boldsymbol{s}_\tau). \tag{A.8}$$

Using (6), the inner most sum over $\boldsymbol{d}_\tau$ collapses to a single term

$$\begin{aligned} &\sum_{\boldsymbol{d}_\tau} \left(\sum_{n=1}^{N} d_{\tau,n} + \rho_\tau(\boldsymbol{X}_{1:\tau}, [\boldsymbol{S}_{1:\tau-1}, \boldsymbol{s}_\tau], [\boldsymbol{D}_{1:\tau-1}, \boldsymbol{d}_\tau]) \right) p(\boldsymbol{d}_\tau \,|\, \boldsymbol{s}_\tau, \boldsymbol{c}_\tau) && \text{(A.9)} \\ &\quad = \sum_{n=1}^{N} s_{\tau,n} c_{\tau,n} + \rho_t(\boldsymbol{x}_{1:t}, [\boldsymbol{s}_{1:t-1}, \boldsymbol{s}_\tau], [\boldsymbol{d}_{1:t-1}, \boldsymbol{s}_\tau \odot \boldsymbol{c}_\tau]) && \text{(A.10)} \\ &\quad = \boldsymbol{s}_\tau^\top \boldsymbol{c}_\tau + \rho_t(\boldsymbol{x}_{1:t}, [\boldsymbol{s}_{1:t-1}, \boldsymbol{s}_\tau], [\boldsymbol{d}_{1:t-1}, \boldsymbol{s}_\tau \odot \boldsymbol{c}_\tau]). && \text{(A.11)} \end{aligned}$$

Substituting this result back into (A.8) yields

$$\begin{aligned} &E\left[\sum_{n=1}^{N} D_{\tau,n} + \rho_\tau(\boldsymbol{X}_{1:\tau}, \boldsymbol{S}_{1:\tau}, \boldsymbol{D}_{1:\tau}) \,\middle|\, \mathcal{F}'_{\tau-1} \right] && \text{(A.12)} \\ &\quad = \sum_{\boldsymbol{s}_\tau} \left(\sum_{\boldsymbol{c}_\tau} \boldsymbol{s}_\tau^\top \boldsymbol{c}_\tau + \rho_t(\boldsymbol{x}_{1:t}, [\boldsymbol{s}_{1:t-1}, \boldsymbol{s}], [\boldsymbol{d}_{1:t-1}, \boldsymbol{s}_\tau \odot \boldsymbol{c}_\tau])\, p(\boldsymbol{c}_\tau \,|\, \mathcal{F}'_{\tau-1}) \right) \xi_\tau(\boldsymbol{s}_\tau) && \text{(A.13)} \\ &\quad = \sum_{\boldsymbol{s}_\tau \in \mathcal{S}_K^N} E_{\boldsymbol{C}_\tau}\left[\boldsymbol{s}_\tau^\top \boldsymbol{c}_\tau + \rho_t(\boldsymbol{x}_{1:t}, [\boldsymbol{s}_{1:t-1}, \boldsymbol{s}], [\boldsymbol{d}_{1:t-1}, \boldsymbol{s}_\tau \odot \boldsymbol{C}_\tau]) \,\middle|\, \mathcal{F}'_{\tau-1} \right] \xi_\tau(\boldsymbol{s}_\tau) && \text{(A.14)} \\ &\quad \leq \max_{\boldsymbol{s}_\tau \in \mathcal{S}_K^N} E_{\boldsymbol{C}_\tau}\left[\boldsymbol{s}_\tau^\top \boldsymbol{c}_\tau + \rho_t(\boldsymbol{x}_{1:t}, [\boldsymbol{s}_{1:t-1}, \boldsymbol{s}], [\boldsymbol{d}_{1:t-1}, \boldsymbol{s}_\tau \odot \boldsymbol{C}_\tau]) \,\middle|\, \mathcal{F}'_{\tau-1} \right] && \text{(A.15)} \\ &\quad = r_{\tau-1}(\boldsymbol{X}_{1:\tau}, \boldsymbol{S}_{1:\tau-1}, \boldsymbol{D}_{1:\tau-1}), && \text{(A.16)} \end{aligned}$$

where the last equality holds by definition of r_t, and (A.15) holds since ξ_τ defines a probability distribution over $\mathcal{S}_K^N$. Note that equality in (A.15) holds if and only if ξ_t is of the form (11). To complete the induction step, we substitute (A.16) back into (A.4) and coarsen the σ-algebra from $\mathcal{F}'_{t-1}$ to $\mathcal{F}_{\tau-1}$:

$$\begin{aligned} &\max_{\xi_{1:\tau}} E\left[\sum_{t=1}^{T} \sum_{n=1}^{N} D_{t,n} \right] && \text{(A.17)} \\ &\quad = \max_{\xi_{1:\tau-1}} E\left[\sum_{t=1}^{\tau-1} \sum_{n=1}^{N} D_{t,n} + r_{\tau-1}(\boldsymbol{X}_{1:\tau}, \boldsymbol{S}_{1:\tau-1}, \boldsymbol{D}_{1:\tau-1}) \right] && \text{(A.18)} \\ &\quad = \max_{\xi_{1:\tau-1}} E\left[E\left[\sum_{t=1}^{\tau-1} \sum_{n=1}^{N} D_{t,n} + r_{\tau-1}(\boldsymbol{X}_{1:\tau}, \boldsymbol{S}_{1:\tau-1}, \boldsymbol{D}_{1:\tau-1}) \,\middle|\, \mathcal{F}_{\tau-1} \right] \right] && \text{(A.19)} \\ &\quad = \max_{\xi_{1:\tau-1}} E\left[\sum_{t=1}^{\tau-1} \sum_{n=1}^{N} D_{t,n} + E\left[r_{\tau-1}([\boldsymbol{X}_{1:\tau-1} \boldsymbol{X}_\tau], \boldsymbol{S}_{1:\tau-1}, \boldsymbol{D}_{1:\tau-1}) \,\middle|\, \mathcal{F}_{\tau-1} \right] \right] && \text{(A.20)} \\ &\quad = \max_{\xi_{1:\tau-1}} E\left[\sum_{t=1}^{\tau-1} \sum_{n=1}^{N} D_{t,n} + \rho_{\tau-1}(\boldsymbol{X}_{1:\tau-1}, \boldsymbol{S}_{1:\tau-1}, \boldsymbol{D}_{1:\tau-1}) \right], && \text{(A.21)} \end{aligned}$$

where the last equality holds by definition of ρ_t. Finally, the induction basis is given by $\tau = T$ and $\rho_T = 0$, for which (A.1) is trivially true.

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

B. Proof of Lemma 1

Once N, K, r, P_0, and P_1 are given, the model no longer needs to be inferred from the data and the optimization problem in (8) reduces to a binary detection problem with known distributions. As a consequence, the cost function in (8) simplifies to

$$\max_{\xi_1} E\left[\sum_{n=1}^{n} D_n\right] = E\left[\max_{s\in\mathcal{S}_K^N} \sum_{n=1}^{N} s_n E\left[C_n \,|\, X_n\right]\right] = \sum_{k=0}^{K-1} E_X\left[E\left[C_{(N-k)} \,|\, \boldsymbol{X}\right]\right], \tag{B.1}$$

with $E\left[C_{(N-k)} \,|\, \boldsymbol{X}\right]$ as defined in Lemma 1. The random variables $C_n|\boldsymbol{X}$ are independent and identically distributed and it holds that

$$E\left[C_n \,|\, \boldsymbol{X}\right] = E\left[C_n \,|\, X_n\right] = P(C_n = 1 \,|\, X_n) \tag{B.2}$$

$$= \frac{r p_1(X_n)}{r p_1(X_n) + (1-r) p_0(X)} \tag{B.3}$$

$$= \frac{1}{1 + \exp\left(-\log \frac{p_1(X_n)}{p_0(X_n)} - \operatorname{logit}(r)\right)}. \tag{B.4}$$

Therefore, the CDF of $E\left[C_n \,|\, \boldsymbol{X}\right]$ is given by

$$F(\varepsilon) \coloneqq P\left[E\left[C_n \,|\, \boldsymbol{X}\right] \le \varepsilon\right] \tag{B.5}$$

$$= P\left[\log \frac{p_1(X)}{p_0(X)} \le \operatorname{logit}(\varepsilon) - \operatorname{logit}(r)\right] \tag{B.6}$$

$$= r\beta\left(\operatorname{logit}(\varepsilon) - \operatorname{logit}(r)\right) + (1-r)\left(1 - \alpha\left(\operatorname{logit}(\varepsilon) - \operatorname{logit}(r)\right)\right), \tag{B.7}$$

where $\operatorname{logit}(r) = \log \frac{r}{1-r}$, $P = rP_1 + (1-r)P_0$ and α and β are defined in (55). From (B.7), we can calculate the distributions of the order statistics $E\left[C_{(N-k)} \,|\, \boldsymbol{X}\right]$:

$$F_{(N-k)}(\varepsilon) \coloneqq P\left[E\left[C_{(N-k)} \,|\, \boldsymbol{X}\right] \le \varepsilon\right] \tag{B.8}$$

$$= \sum_{j=0}^{k} \binom{N}{j} \left(F(\varepsilon)\right)^{N-j} \left(1 - F(\varepsilon)\right)^{j} \tag{B.9}$$

$$= \sum_{j=0}^{k} \sum_{i=0}^{j} \binom{N}{j} \binom{j}{i} (-1)^{j-i} \left(F(\varepsilon)\right)^{N-i} \tag{B.10}$$

$$= \sum_{i=0}^{k} \left(F(\varepsilon)\right)^{N-i} \sum_{j=i}^{k} \binom{N}{j} \binom{j}{i} (-1)^{j-i} \tag{B.11}$$

$$= \sum_{i=0}^{k} \left(F(\varepsilon)\right)^{N-i} w_{N,k,i}, \tag{B.12}$$

where (B.9) is shown in David & Nagaraja (2004, Chapter 2), (B.10) follows from the fact that

$$(1-x)^j = \sum_{i=0}^{j} \binom{j}{i} (-1)^{j-i} x^{j-i}, \tag{B.13}$$

(B.11) is obtained by rearranging the lower-triangular sum over j and i (row-first vs. column-first), and (B.12) holds since

$$\sum_{j=i}^{k} \binom{N}{j} \binom{j}{i} (-1)^{j-i} = \frac{N-k}{N-i} \binom{N}{k} \binom{k}{i} (-1)^{k-i} = w_{N,k,i}. \tag{B.14}$$

This identity can be shown by repeated application of Pascal's formula (Weisstein, 2024) and has been verified by computer algebra systems.

Since $E\left[C_n \,|\, \boldsymbol{X}\right]$ is a non-negative random variable, the expected values on the right-hand side of (B.1) can now be written as

$$E_X\left[E\left[C_{(N-k)} \,|\, \boldsymbol{X}\right]\right] = \int_0^1 \left(1 - F_{(N-k)}(\varepsilon)\right) \mathrm{d}\varepsilon \tag{B.15}$$

$$= 1 - \int_0^1 F_{(N-k)}(\varepsilon) \,\mathrm{d}\varepsilon \tag{B.16}$$

$$= 1 - \sum_{j=0}^{k} w_{N,k,i} \int_0^1 \left(F(\varepsilon)\right)^{N-i} \mathrm{d}\varepsilon. \tag{B.17}$$

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press

In order to evaluate the integral in the last equation, we change the integration variable from ε to

$$\lambda = \text{logit}(\varepsilon), \tag{B.18}$$

with the corresponding transformation of the infinitesimal

$$\mathrm{d}\varepsilon = \frac{1}{2}\,\text{sech}\left(\frac{\pi\lambda}{2}\right)\mathrm{d}\lambda. \tag{B.19}$$

It hence holds that

$$\int_0^1 \big(F(\varepsilon)\big)^n\,\mathrm{d}\varepsilon = \int_{-\infty}^{\infty} \big(F(\lambda)\big)^n\,\frac{1}{2}\,\text{sech}\left(\frac{\pi\lambda}{2}\right)\mathrm{d}\lambda \tag{B.20}$$

$$= \int_{-\infty}^{\infty} \big(r\beta(\lambda) + (1-r)(1-\alpha(\lambda))\big)^n\, p_\Lambda(\lambda + \text{logit}(r))\,\mathrm{d}\lambda \tag{B.21}$$

$$= E_\Lambda\Big[\big(r\beta(\Lambda) + (1-r)(1-\alpha(\Lambda))\big)^n\Big], \tag{B.22}$$

where the second equality follows from (B.7), and p_Λ and Λ are as defined in the lemma. This completes the proof.

Cite this article: Fauss, M., Liu, X., Li, C., Choi, I. and Poor, H. V. (2026). Bayesian Selection Policies for Human-in-the-Loop Anomaly Detectors with Applications in Test Security. *Psychometrika*, 279–311. https://doi.org/10.1017/psy.2025.10056

https://doi.org/10.1017/psy.2025.10056 Published online by Cambridge University Press