

Minimax Optimization of Personalized Training Strategies for Student Physical Fitness under Performance Uncertainty

Zuopeng Liu¹, Nur Shakila Mazalan^{2*}, Wan Ahmad Munsif Wan Pa³

¹Faculty of Education, Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia.

E-mail: p153066@siswa.ukm.edu.my

¹Faculty of Education, Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia.

E-mail: shakila@ukm.edu.my

¹Faculty of Education, Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia.

Email: munsif@ukm.edu.my

*Corresponding Author: Nur Shakila Mazalan, E-mail: shakila@ukm.edu.my

Abstract

Making decisions on a personalized physical-fitness programme is a decision problem in which the expected benefit has to be weighed against the uncertain physical-fitness performance and the limited resources of the intervention. This study proposes a framework for prioritizing training needs of students in a public Chinese university dataset with the minimax criterion. The 2,489 observations consisted of two merged student-profile files which included exercise habits, body type, exercise stability, willpower tier, active days and behavioral pattern. The student profile data was clustered using k-means to obtain four readable student profiles. Observed features were used to build a performance proxy and a cross-sectional uncertainty score which are not actual physiological results. Four decision options were tested: foundation consistency, performance enhancement, robust behavioural support, and baseline (non-personalised). A worst-case performance criterion with uncertainty and strategy cost penalties was maximized. For three less successful or more uncertain profiles, a foundation of consistency was recommended, and baseline for the highly active consistent profile. The minimax allocation yielded larger average performance modelled across the set of tested uncertainty weights than the baseline allocation, and smaller uncertainty modelled across the set of tested uncertainty weights than the baseline allocation. The results highlight the potential of minimax reasoning to structure personalized fitness decisions in the presence of uncertainty, and indicate that there is a need for fitness measurements over time as well as empirically determined intervention effects before causal and clinical claims can be made.

Keywords: minimax optimization; robust decision-making; personalized training; student physical fitness; performance uncertainty; physical activity; China

1. Introduction

The principle of personalized physical training is very simple but its decision structure is not easy. The amount of activity participated in, the body profile, the behavioral routines, and the consistency experienced by students vary from one to another when they enter into a training program. The same prescribed program can then have various levels of benefit, adherence, fatigue and uncertainty. While group-average analyses are important for drawing conclusions about central tendencies, they do not directly answer the practical question that must be addressed by a teacher or training system: "Which training priority should go to the individual or profile who might not fare as well as the group average? This is an important question to ask if there is insufficient data. The decision maker does not need to repeat laboratory fitness tests or have experimental estimates of treatment effects for a variety of activity behavior and profile characteristics. In such circumstances, it is easy to have a strategy that seems very appealing with a high expectation for scoring and then when the conditions are not favorable, but still reasonably likely, it may do poorly.

Physical-education and physical-literacy literature validates the value of acknowledging students' diversity, not considering all "students" interchangeable. Physical literacy is now often defined as a multi-dimensional ability that embodies physical competence, motivation, confidence, knowledge and engagement within the context of activity (Cornish et al., 2020; Rudd et al., 2020). The studies with Chinese populations also reveal the relationship between demographic, interpersonal and behavioral characteristics with physical literacy and physical activity participation (Li et al., 2020; Chen et al., 2024; Yang et al., 2024). Intervention reviews reveal that program design is important, and that the content, target, and implementation of physical-literacy interventions differ significantly

(Carl et al., 2022a, 2022b). When actionable information about a person's profile exists, a rule of one size fits all training does not seem defensible based on these findings.

Personalization should not be confused with unconstrained algorithmic optimization. Sports-science research separates the external training load from the internal reaction to the external load. Within the population, the same task can have different physiological (Impellizzeri et al., 2020) and perceptual demands. The quality of measurement, context, and variability of response are key challenges that remain (Rana & Mittal, 2020; Düking et al., 2021; Li et al., 2023), and wearable technologies and data analytics can assist in monitoring. In addition, sports analytics reveals that it is important to have well-defined objectives, validation options, and decision implications (Watanabe et al., 2021). Data-rich decision support in educational settings thus requires a rule for dealing with uncertainty, rather than ranking of expected performance.

One such rule is the minimax rule, which is a form of reasoning. In robust optimization, the decision maker defines a family of possible failure situations, but only allows failure cases that are admissible, and then picks an action that is optimal for the worst case failure. It is not the case that the worst case will definitely happen. Instead, the decision is shielded from potential degradation, which has yet to be addressed by the data present. Modern distributionally robust optimization generalizes this idea to the case of uncertainty over the data generating distribution, and recent statistical research has shed light on when robust rules can offer defensible out-of-sample protection (Duchi et al., 2021; Van Parys et al., 2021; Blanchet et al., 2022; Rahimian & Mehrotra, 2022; Gao & Kleywegt, 2023). A simple finite-scenario approach is better than a complex ambiguity set when the outcome measurements are not very abundant, as it is for a student-fitness application.

This research constructs such a finite model structure based on a student profile repository based on China public data. There are no available records with direct measures of maximal strength, cardiorespiratory fitness, sprint time, flexibility, or repeated measures of pre-post training results. Instead, they identify exercise habits, body type, exercise stability, willpower level, number of active days accumulated and behavior pattern. Such variables can help in student profiling and a fitness based behavioral readiness model but cannot be used to explain training effects. The analysis is therefore carried out in such a way that observed and model derived quantities are separated at each stage. Profiles are defined by the observed variables. Explicit decision features are then developed as a performance proxy and a measure of uncertainty. Lastly, hypothetical strategy-effect parameters are only considered as scenario parameters to illustrate the minimax decision rule. Their effects on intervention are not reported as being estimated empirically.

The difference is at the core of the paper's contribution. One of the most common mistakes in data-driven training research is to move from behavioral data to Prescriptive Solutions, but without consideration for where the empirical evidence leaves off and the assumptions of the model begin. The trouble can give the impression of accuracy which isn't warranted by the data. In order to be reproducible, science must include the target quantity, the estimand, and the inferential boundary must be visible (Lundberg et al., 2021; Nosek et al., 2022; Kapoor & Narayanan, 2023). Hence, personalization is regarded as a decision analysis problem. It is not asking which strategy is implemented under a given loss function and uncertainty set and that the selected strategy would have been proven to enhance physiological fitness; it is asking which is chosen under a given loss function and stated uncertainty set.

The study fills in four interrelated gaps. The theoretical gap is the relationship between a weak and formal robust decision theory in student-centered physical training. The empirical gap refers to a lack of public Chinese student datasets that have multiple objective fitness tests and intervention assignments. The methodological gap is in the fact that, in many cases, only expected-value ranking is used, even though there is no specific protection against negative performance. The actual problem is that the recommendations at the level of the profile are not visible and can not be audited by the teacher, but remain in a black box score. A profile-based minimax model, algebraically explicit with robust counterpart, whose assumptions are reported with results, fills these gaps.

The paper has three bounded contributions. First, it demonstrates that different activity profiles of students can be represented in a decision model without the creation of non-existent physiological variables. Second, it provides a formal worst-case adjusted performance score interpretation of the minimax problem, and shows that the inner worst case problem is equivalent to a closed-form robust counterpart. Third, it tests if the suggested approach varies with the amount of uncertainty. The training-effect matrix is designed to be scenario-based, so the recommendations that are derived from it should be viewed as an application of minimax theory from a methodological perspective. They do not demonstrate effectiveness of treatment. The framing is intentionally conservative and appropriate to the information content of the data.

2. Literature Review and Theoretical Background

2.1 Personalization, Physical Literacy, and Heterogeneous Training Needs

The modern literature on physical education puts forward three dimensions of competence, motivational, behavioral engagement, and contextual support, all of which are related but not identical. The use of a physical-literacy approach proves beneficial as it does not limit participation to one physiological measure. Rudd et al. (2020) highlighted individual enrichment, and how adaptive interaction with the environment can be approached, with Cornish et al. (2020) demonstrating the importance of the interpretation of health being linked to multiple domains and how this can differ. A second important behavioural aspect of activity participation is enjoyment, which has been investigated using short form instruments (Chen et al., 2021). In addition, cross-validated in the Chinese population, it was found that measurement systems cannot be simply imported without taking into account the context (Li et al., 2020). Here the arguments are used to justify the use of the profiles to make decisions, but they do not imply that the labels describing the behaviour in the repository actually measure physical fitness.

Student-centered pedagogy also aids differentiated training priorities. Santos et al. (2022) linked physical literacy to youth development and student-centred approach, while Houser and Kriellaars (2023) defined physical-literacy-enriched pedagogy as a means "to make P.E. responsive to learner needs". Additionally, teacher perceptions and physical-literacy assessment at the teacher level influence the implementation of these ideas (Stoddart & Humbert, 2021; Yin et al., 2024). Choi et al. (2021) demonstrated that structured sport education could have an impact on perceived physical literacy, motivation and activity in a cluster randomized trial conducted at the university level. Kwan et al (2020) also investigated a physical-literacy intervention at university entry. These studies provide empirical basis for predicting student heterogeneity. They also illustrate the importance of exposure to the direct program and measured outcomes for intervention claims. In the repository here analyzed, this design information is missing, and training categories are not used as treatments, but rather as scenarios.

Chinese evidence provides a rationale for activity behavior and support specific to the country. The study conducted by Chen et al. (2024) revealed that there were significant associations between physical literacy, demographic characteristics, and interpersonal support among Chinese adolescents. Yang et al. (2024) examined the relationship between physical literacy and health among medical students in China by exploring the two pathways of physical activity and subjective wellbeing. Liu et al. (2025) demonstrated that competence support, exercise persistence, and perceived competence can be linked in the context of middle-school sports participation. Similarly, Melby et al. (2023) reported cross-sectional relationships between physical literacy, physical activity, and well-being. None of these studies suggests that the profile categories of the present dataset are validated physical-literacy scales. But they do not accept the general idea that the activity behavior and participation are heterogeneous and context dependent.

Content of a custom training plan is important too. Carl et al. (2022a) concluded that interventions involving physical-literacy can also be effective, however, there were differences in settings and program designs when they conducted their meta-analysis. Their review of the companion interventions indicated that they vary in terms of target definition, content structure, and theory to practice (Carl et al., 2022b). O'Connor and Penney (2021) pointed out the need for knowledge and skills to be important for participation in both formal and informal sport settings. Then Schmittwilken et al. (2024) demonstrated the relevance of engaging in extracurricular activities that are aligned to physical-literacy goals. The results do not support the concept of a single universal training dose. Some are also supportive of current use of broad strategy families, which are basic consistency, performance improvement, behavioral support, and baseline provided families, provided they are clearly identified as scenario choices and NOT measured programs.

2.2 Performance Uncertainty in Training Decisions

There are multiple reasons to consider the training response. Different individuals may respond differently with the same external load, and measurement noise may result in a more stable performance over the short term than is truly the case (Impellizzeri et al., 2020). Although monitoring technologies can fill information gaps, they can't eliminate gaps completely. Düking et al. (2021) presented an overview of heart-rate-variability-based endurance training and demonstrated the potential and practical limitations of adaptive monitoring. Although heart-rate variability (HRV) measurement via smart wearable devices has been explored in previous studies, Li et al. (2023) noted a need to more clearly separate the convenience of measurement from the validity of the measurement. Similarly, for wearable kinematic sensing, Rana and Mittal (2020) concluded that they can be used to monitor body movements. Therefore, uncertainty is not just a statistical annoyance, it is a component of the decision context that decides on a selection of a training strategy.

The use of data analysis in sport has grown both the ability to describe and compare performance patterns. Watanabe et al. (2021) observed the increasing importance of big data and analytics in sport decision support and Weakley et al. (2023) demonstrated that feedback has the capability to affect resistance-training performance and

adaptation. The difference which is relevant to this study is easy to understand. Prediction: What will happen? This is a tough question, and one that is the subject of robust optimization: What to opt for when predictions are not certain and poor outcomes need to be protected against. The latter is the subject of the minimax model developed here.

Because the present uncertainty score is intentionally cross sectional. It is an inverse of the performance proxy, adverse profile indicators, missing-profile penalties, and within-cluster dispersion. It is not a variance estimate from one time to the next (longitudinal variance), and should not be called test-retest instability. This constraint is significant as it would allow uncertainty to be estimated from repeated objective outcomes, ideally on exposure to known training exposures. Until these data on uncertainty are available the uncertainty index should be seen as a conservative decision feature. The minimax model is useful because it lets us write out the uncertainty, which would otherwise be treated as a fact that every prediction is equally good.

2.3 Minimax and Robust Optimization

Classical expected value optimization involves ranking strategies based on the average of modeled outcomes. It's appealing when the probability distribution is known and the loss incurred by a poor tail event is not too large. In personalized training, the relevant distribution may, however, be poorly estimated. An irregular pattern of activity could result in a student achieving a great expected improvement with an intensive approach but also having a large amount of variation in adherence or a stable and poor response. The strength of robust optimization is to use an uncertainty set instead of a single point forecast. Then, the decision is chosen from the least favorable state in that set. There are different formulations of the risk concept: risk-neutral, risk-averse, and distributionally robust; and as is natural, training decisions can be made with varying degrees of confidence in the underlying data. There are different forms of the risk concept: risk-neutral, risk-averse, and distributionally robust; and as one might expect, training decisions can be made with varying levels of confidence in data.

There are multiple methods of creating ambiguity sets in modern distributionally robust optimization. Duchi et al. (2021) related strong optimization to generalized empirical likelihood. Van Parys et al. (2021) gave a pioneering finding, in the form of a decision-theoretic result, in which the distributionally robust optimization turns out to be an optimal data-driven decision rule. Recently Wasserstein methods have gained traction, as they define neighborhoods around an empirical distribution and offer out-of-sample protection (Blanchet et al., 2022; Gao & Kleywegt, 2023). Rahimian and Mehrotra (2022) surveyed the entire field and noted that the ambiguity set should be selected in line with the information available. There is here a crucial principle, namely that a sophisticated Wasserstein ambiguity set would include distributional information that this profile dataset does not include.

Therefore, a finite-scenario uncertainty set is more suitable. The model makes the assumption that there is a limited performance degradation from the WORST possible performance from each cluster-strategy combination. The upper limit is tied to the uncertainty score at the time, with the assumed uncertainty reduction of the scenario. The strategy (or strategies) that lead to the greatest lower bound performance, minus a scaled strategy cost, is selected. It's a minimax problem, since after the selection of a strategy, nature selects the worst possible loss within the admissible interval. The mathematical structure is easy to audit and retains the key robust-decision logic. It also permits the use of the sensitivity parameter λ as a meaningful interpretation: the larger the value, the more the algorithm will consider uncertainty protection; the smaller the value, the closer the algorithm will be to the sense of expected performances.

2.4 Methodological Rigour and Inferential Boundaries

Even with a good optimization model the variables that are fed into it can be confusing. For nonexperimental research, Grosz et al. (2020) warned against the use of causal language, and Tennant et al. (2021) illustrated that explicit causal structures were necessary when using observational data in order to answer causal questions. The same was suggested by Lundberg et al. (2021) who claimed that whenever interpreting statistical evidence, researchers should first define the estimand. These are the key principles that directly apply to this study. The estimand does not represent a causal effect of a training programme. It's a model-driven worst-case utility for a profile with the given set of scenario assumptions.

The second boundary is provided by reproducibility. Algorithmic analyses can result in self-consistent results that are very sensitive to choices that are made in preprocessing, leakage or tuning. While Nosek et al. 2022 noted the difference between Replicability, Robustness and Reproducibility, Hardwicke et al. 2021 demonstrated that open data does not necessarily lead to analytically reproducible results. Data leakage was found to be a significant culprit for exaggerated claims in machine learning by Kapoor and Narayanan (2023). The analysis does not make any supervised predictions and the full cluster-selection diagnostics produced in the notebook are reported. Interestingly, the silhouette score of $K=5$ was slightly higher than that of $K=4$. The 4-cluster solution was not used

in order to obtain the most distinct silhouette coefficient but to allow for the interpretability and stability of profile sizes.

2.5 Research gap and research questions.

There is thus a clear gap in the literature where student-centered training merges with strong decision theory. Physical-education research has been found to support individual needs and differentiated pedagogical responses, although the majority of research studies are based on the assessment of observed interventions, associations or average outcomes. While robust optimization offers tools for adverse-state decision-making, these are less frequently expressed in simple, easy-to-implement profile-level rules for training students. The main research questions in this study are as follows: RQ1: Can the available activity data from China be used to distinguish between the behaviorally distinct student profiles? To address RQ2, how should a finite-scenario minimax algorithm distribute the general training priorities among those profiles when two types of penalties are explicitly applied to deal with performance uncertainty and strategy cost? RQ3: Do the profile-level recommendations change if the uncertainty weight is changed?

The expected contribution is methodological; not causal. If a valid result is obtained, there is no evidence that one strategy is fitter than another in the real world. It would demonstrate that if the profile construction and scenario assumptions are as stated then the minimax decision rule will reveal a clear priority set (usually transparent) and will highlight the assumptions underlying those priorities. This separation is useful within the scope of minimax theory and applications as well as providing a realistic balance between the number of data points and the empirical statements.

3. Data and Methods

3.1 Data Source and Analytical Sample

The analysis was performed on the public GitHub repository [haog95/Stu_analyse](https://github.com/haog95/Stu_analyse) (https://github.com/haog95/Stu_analyse), which includes a few student-profile files. There are two files in this decision problem. The physical fitness profile file consisted of 2,434 records with five fields that included student identifier, exercise habit, body type, exercise stability and willpower. There were 2,489 lifestyle-pattern entries in the lifestyle-pattern file, with four data fields: student identifier, total active days, cluster identifier of the source workflow, and behavioral pattern. Students were joined to the files on student identifier using an outer join. The analytical sample produced in the process yielded 2,489 separate student-level observations. Observations that did not contain the physical-profile fields were not removed, but rather were retained in a special Unknown category. The physical-profile file did not contain any duplicate rows and there were 2,434 records in the physical-profile file. The labels for exercise were extremely skewed with 2,420 students labelled as Athlete, 11 students labelled as Regular Exercise and 3 labelled as Occasional Exercise. The body type was more varied, comprising 1,660 Standard, 455 Overweight, 178 Underweight and 141 Obese records. Each category of exercise stability was recorded on the following numbers of records: 2,187 records were labelled “Moderately Stable”, 234 were labelled “General”, 9 were labelled “Very Stable”, and 4 were labelled “Unstable”. Willpower tiers were Silver (608), Gold (608), Bronze (365), Platinum (365), Diamond (244), and Needs Improvement (244). These are not clinical diagnoses or separately developed psychometric scales.

The lifestyle file uncovered a greater variety of behavioural patterns. High-Variance Irregular accounted for 863 records (34.7%), Late-Night Active for 789 (31.7%), Standard Mixed for 611 (24.5%), Early-Morning Regular for 210 (8.4%), and Insufficient Data for 16 (0.6%). The mean number of total active days was 372.15, the standard deviation was 97.79, the median was 359 and the interquartile range was 327 to 402. The lowest was 1 and the highest 873. Since the repository does not supply an overall calendar origin for this field, and no observation-window metadata for this field, it was assumed that it would be a relative activity-volume feature rather than a rate for a specified period.

3.2 Unit of Analysis, Time Period, and Scope

It was the student profile which served as the unit of analysis. The repository did not provide a verified institutional sampling frame, collection dates or repeated fitness-test waves in analyzed files. The study is thus a cross-sectional study from an analysis point of view. The observation level is a single profiled student per student identifier. There are no institutions named in this manuscript as the files seen do not warrant the assignment of the records to a particular university. Data are presented to support an application context that is based in China, but claims of national representativeness are not supported. This was not a new experiment to sample data from, so no prospective power calculation was used; the sample-size justification was based on available-data justification that is recommended for transparent secondary analyses (Lakens, 2022).

3.3 Variable Roles and Measurement Boundaries

All of the variables are specified in Table 1 and the observed features are differentiated from the constructs derived from the model. There is no measurable dependent variable in the unprocessed data because the outcome of training is not measured. A decision-oriented performance proxy is thus defined as the desired target in the optimization model. The inputs observed are exercise habits, body type, exercise stability, willpower tier, active days and behavioural pattern. The membership of clusters is obtained from the preprocessing and the K-means. The indices are indices that are calculated as performance and uncertainty scores. Parameters for the scenarios are: training strategy, cost of strategy, and performance and uncertainty changes specific to the strategy. No mediators or moderators due to lack of data and/or a lack of design to justify causal pathway modeling.

Table 1. Variable definitions and evidential status

Status	Variable	Type	Operational meaning	Role / boundary
Observed	Exercise_Habits	Categorical profile label	Student-level activity orientation	Independent input; not intervention exposure
Observed	Body_Type	Categorical profile label	Standard / overweight / underweight / obese	Independent input; not measured body composition
Observed	Exercise_Stability	Categorical profile label	Very stable / moderately stable / general / unstable	Independent input
Observed	Willpower_Tier	Categorical profile label	Diamond to needs improvement	Independent input; not validated scale
Observed	total_active_days	Count-like continuous feature	Cumulative active-day quantity in source file	Primary continuous activity feature
Observed	Behavioral_Pattern	Categorical profile label	High-variance, late-night, standard mixed, early-morning, insufficient data	Independent input
Derived	Profile_Cluster	K-means class	Four retained behavioral profiles	Personalization unit
Derived	Performance_Proxy_Score	Dimensionless index	Directional combination of observed features	Decision outcome proxy
Derived	Performance_Uncertainty_Score	Dimensionless index	Inverse performance + adverse indicators + missingness + cluster dispersion	Robustness input
Scenario	Training_Strategy	Decision variable	FC / PE / RBS / Baseline	Action in minimax model
Scenario	dP, dU, Cost	Assumed scenario parameters	Strategy-specific gain, uncertainty reduction, cost	Not empirical effect estimates

Note. Observed fields come directly from the two repository files. Derived and scenario quantities were generated by the notebook. No mediator or moderator was estimated.

3.4 Preprocessing

Only one student identifier was used for merging, and not for any modeling. A standardization was used for Continuous total active days. One-hot encoding was performed for the categorical variables. Physical-profile fields that were not available were coded as Unknown and the 16 missing source Cluster_ID values were coded as -1. A total of 32 encoded features ended up in the final model matrix. This method left the entire merged sample intact and allowed the uncertainty construction to see the missingness. In the modeling stage, no names, addresses, or direct identifiers were used to identify any person.

3.5 Student Profiling

The silhouette coefficient was used to determine the optimal number of clusters (K) for k-means clustering, with K ranging from 2 to 10. The observed scores were 0.199 for K=2, 0.262 for K=3, 0.297 for K=4, 0.323 for K=5, 0.291 for K=6, 0.270 for K=7, 0.258 for K=8, 0.257 for K=9, and 0.274 for K=10. The coefficient was the highest with K=5. K was kept as 4 in the notebook as it resulted in a more compact profile solution, which had distinct activity patterns and enough cluster size. This was not a statistical optimum model-selection decision, but rather a judgment-based choice of model. The complete diagnostic sequence is reported in Table 3, which allows the decision to be auditable.

The four-cluster solution retained was found, with cluster sizes of 783, 841, 572 and 293. Their mean total active days were 356.47, 346.93, 337.92, and 553.28, respectively. Clusters were identified by the dominant behavioral information and the interpretation of the profile in the notebook: Late-Night Active (Cluster 0); High-Variance Irregular (Cluster 1); Standard Mixed/Early-Morning (Cluster 2); Highly Active and Consistent (Cluster 3). These names are a summary of the encoded profile structure. They are not clinically or physiologically latent classes.

3.6 Performance Proxy

An explicit 'directional' scoring rule was used to build the performance proxy from the features observed. Higher standardized active days; Athlete label and Regular exercise; Standard and Underweight body type; Very Stable and Moderately Stable exercise stability; higher willpower levels; and the Early-Morning Regular pattern were given positive contributions. Occasional Exercise, Overweight or Obese body type, Unstable exercise behavior, lower willpower tiers, High-Variance Irregular and Late-Night Active patterns and Insufficient Data received negative contributions. Any unrecognized profile fields were given a minor penalty. This results in an indicator of behavior-clinical performance fitness proxy. It cannot be used as a measure of VO2max, strength, sprint speed, body-composition change or a validated total fitness score.

3.7 Performance Uncertainty

The uncertainty index was also explicitly built. It brought together four components: the inverse min-max scaled performance proxy, adverse profile indicators, penalties for unknown profile fields and normalised within-cluster dispersion of the performance proxy. The higher the number, the less favourable the modeled performance, the more unfavourable the behavioral indicators, the more incomplete the information, or the greater the local dispersion of the profile. This score is not a temporal variance estimate, as data are cross-sectional. Performance uncertainty does not refer to repeat-testing variability within the student, but rather in the decision-theoretic sense, is the uncertainty surrounding the modeled profile.

3.8 Training-Strategy Scenarios

There were four decision options defined in the notebook: Foundational Consistency (FC), Performance Enhancement (PE), Robust Behavioral Support (RBS), and Baseline. The 4 scenario costs were 1, 3, 5 and 0. The matrices of strategy-effects identified cluster specific changes in performance and uncertainty. For instance, FC had performance changes of 0.5, 0.4, 0.6, and 0.1 in Clusters 0-3, and uncertainty reductions of 0.4, 0.3, 0.5, and 0.1. PE highlighted Cluster 3 more clearly and uncertainty reductions provided by RBS were increased for the most irregular profile. The values are assumptions about models. Effect sizes are not estimated from intervention data and should not be used as an interpretation.

3.9 Formal Minimax Model

Define a student profile (c) and a candidate strategy (s). Suppose P_c is the measure of current profile performance and U_c is the measure of uncertainty, and that the performance gain and uncertainty reduction due to the scenario are dP_{cs} and dU_{cs} , and the cost of the strategy is C_S . Allow for cost and uncertainty ($\lambda \geq 0$ and $\gamma \geq 0$). Once strategy s is selected, nature will draw an adverse loss $x_{i,cs}$ from the interval $X_{i,cs} = [0, U_c - dU_{cs}]$. The robust decision is given by Equation (1).

Figure 1 summarizes the complete observed-data-to-minimax decision workflow used in the study.

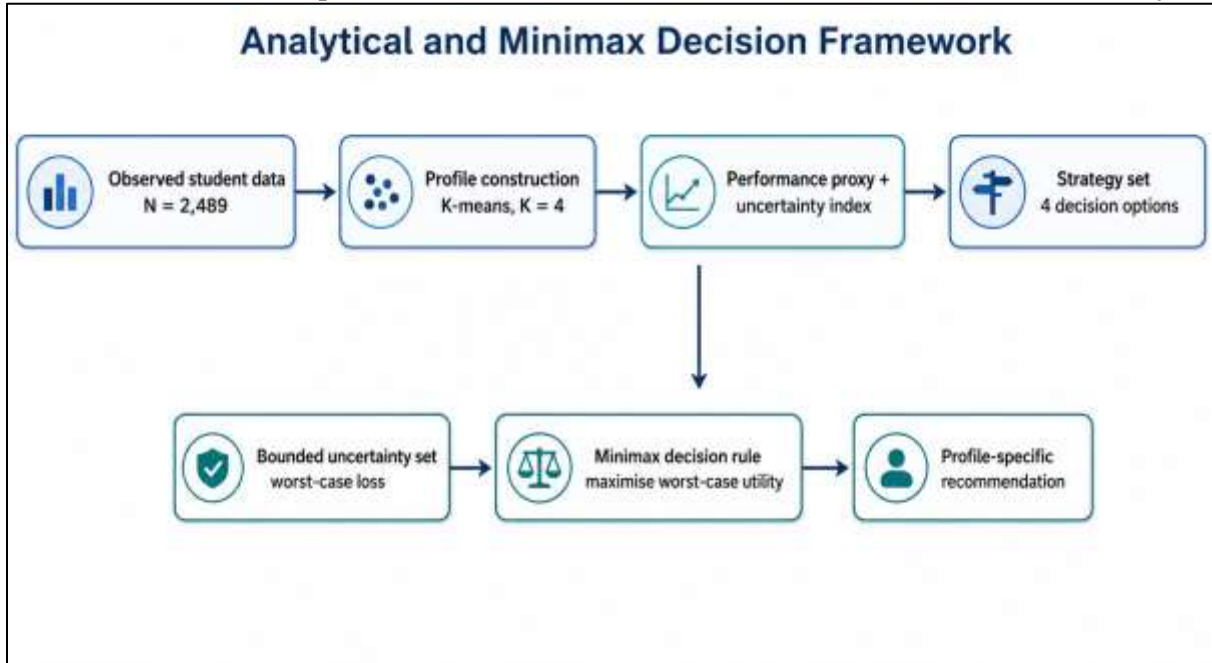


Figure 1. Analytical and minimax decision framework. Source: author-developed from the supplied analysis workflow.

$$s^*_c = \arg \max_s \min_{\{x_{i_{cs}} \in X_{i_{cs}}\}} [P_c + dP_{cs} - \lambda x_{i_{cs}} - \gamma C_s]. \quad (1)$$

For $\lambda \geq 0$, the inner minimum occurs at the largest admissible loss, $x_{i_{cs}} = U_c - dU_{cs}$. The robust counterpart is therefore given by Equation (2).

$$W_{cs} = P_c + dP_{cs} - \lambda (U_c - dU_{cs}) - \gamma C_s. \quad (2)$$

$$s^*_c = \arg \max_s W_{cs}. \quad (3)$$

The notebook used $\lambda = 1.0$ and $\gamma = 0.5$. This closed-form equivalence is the core minimax mechanism examined in the paper.

Proposition 1 (finite-interval robust counterpart). For any profile c and strategy s with $\lambda \geq 0$ and uncertainty set $X_{i_{cs}} = [0, U_c - dU_{cs}]$, the minimax inner problem is equal to $P_c + dP_{cs} - \lambda(U_c - dU_{cs}) - \gamma C_s$. Proof: the objective is affine and non-increasing in $x_{i_{cs}}$ when $\lambda \geq 0$. Therefore, its minimum over the closed interval occurs at the upper endpoint. Substitution yields the stated expression. The proposition is elementary, but it is important because it shows that the reported worst-case-adjusted performance values are not merely heuristic penalties; under the stated interval uncertainty set, they are the exact robust counterpart of the finite minimax problem.

3.10 Comparator Rules and Sensitivity Analysis

The three comparators that were kept in the notebook are shown below. All profiles were given Baseline as a size-1 fits-all. The expected-performance rule opted for strategies based on modeled performance, but with a different worst-case focus. The non-robust profile rule was the same as the robust profile rule, except that it did not introduce a separate adverse-state decision layer. The first sensitivity analysis was performed with a range of λ (sensitivity analysis) while keeping the scenario matrices and cost scale constant at 0.5, 1.0 and 1.5. The primary criterion for robustness was recommendation stability. The scenario matrices are assumed, and not estimated, and thus p-values are not meaningful for the optimization outputs. The checks are the transparency, comparative performance, and sensitivity to uncertainty preference.

3.11 Statistical Reporting and Reproducibility

Descriptive statistics of observed and derived variables. The analysis is not based on normal-theory significance testing, which means that it does not assume normality as a prerequisite that is not required by the decision model and the combination of measurement types (Knief & Forstmeier, 2021). There are no psychometric reliability coefficients reported in this document since source fields are profile labels (categorical instead of multi-item reflective scales). A correlation matrix between the derived performance and uncertainty scores would also be

mechanically confounded with construction since the uncertainty index includes the inverse performance proxy, therefore, this is not shown as independent validity evidence. The silhouette coefficients are used to represent the cluster quality. The model results are deterministic given the input profiles, scenario matrices, lambda and gamma. The entire computation pipeline is packaged in the provided notebook. Authors should store an executable version together with the scenario matrices and information about the version of the software prior to publication.

3.12 Ethics and Privacy

This is a secondary analysis; no new participants were recruited. The public repository is based on identifiers for the students that have been erased due to file linkage. No ethics approval number or consent statement was found for the original data collection in the materials reviewed for this paper. The manuscript thus does not create one. The authors should ensure the provenance, consent and ethics associated with the original repository and the journal's policy regarding secondary analysis of publicly available human-related data before submission.

4. Results

4.1 Dataset Validation and Descriptive Structure

A total of 2,489 student profiles were in the merged analytical sample. A total of 2,489 observations were contributed by the lifestyle file and 2,434 complete profile rows were contributed by the physical-profile file. The outer merge thus generated 55 Unknown observations for the four categories of physical-profile. The sample retained was sufficiently large for exploratory profile segmentation, but was limited by significant imbalance of categories in a number of cases. In particular, 99.42% of the physical-profile file was labeled Athlete and 89.85% was labeled Moderately Stable. The distributions restrict the available independent variation in the delicate effects of exercise-habits. The summary of most observed distributions and two derived decision indices are shown in Table 2.

Table 2. Sample characteristics and descriptive statistics

Characteristic	Value	Interpretation
Analytical sample	2,489	Merged student profiles
Physical-profile records	2,434	Complete in source file
Unknown physical-profile after merge	55	Retained as Unknown
Athlete label	2,420 (99.42%)	Of 2,434 physical-profile rows
Standard body type	1,660 (68.20%)	Of 2,434
Overweight body type	455 (18.69%)	Of 2,434
Underweight body type	178 (7.31%)	Of 2,434
Obese body type	141 (5.79%)	Of 2,434
Moderately stable exercise	2,187 (89.85%)	Of 2,434
High-Variance Irregular pattern	863 (34.7%)	Of 2,489
Late-Night Active pattern	789 (31.7%)	Of 2,489
Standard Mixed pattern	611 (24.5%)	Of 2,489
Early-Morning Regular pattern	210 (8.4%)	Of 2,489
total_active_days	372.15 $\pm$ 97.79	Median 359; IQR 327-402; range 1-873
Performance proxy	2.211 $\pm$ 2.483	Median 2.702; range -7.296 to 10.123
Uncertainty score	2.512 $\pm$ 1.087	Median 2.306; range 0.716 to 5.781

Note. Performance and uncertainty are dimensionless, model-derived scores. Percentages for physical-profile categories use $n=2,434$; lifestyle patterns use $n=2,489$.

The performance proxy had a mean of 2.211 and a standard deviation of 2.483. Scores ranged from -7.296 to 10.123, with a median of 2.702. The uncertainty index had a mean of 2.512 and a standard deviation of 1.087, with a median of 2.306 and a range from 0.716 to 5.781. These values are dimensionless model scores. Their scale has meaning only relative to the scoring rules used in the notebook. They should not be compared with clinical or national physical-fitness standards.

4.2 Student Profiles

The four clusters that have been retained are then located in performance-uncertainty space in Figure 2. The data for cluster 3 was nicely distinct from the others. It had the highest modeled performance (5.542), the lowest uncertainty (1.433) and the highest mean active-day count (553.28). Among the larger groups, Cluster 1 was the least favorable with a performance of 1.333 and uncertainty score of 3.031. The performance of the Cluster 0 was 1.703 with an uncertainty of 2.728. Cluster 2 was in the middle with a performance of 2.489 and uncertainty of 2.006. The profile distribution was 31.5% in Cluster 0, 33.8% in Cluster 1, 23.0% in Cluster 2, and 11.8% in Cluster 3.

Figure 2 visualizes the retained clusters using their mean performance proxy and uncertainty score; bubble size represents cluster size.

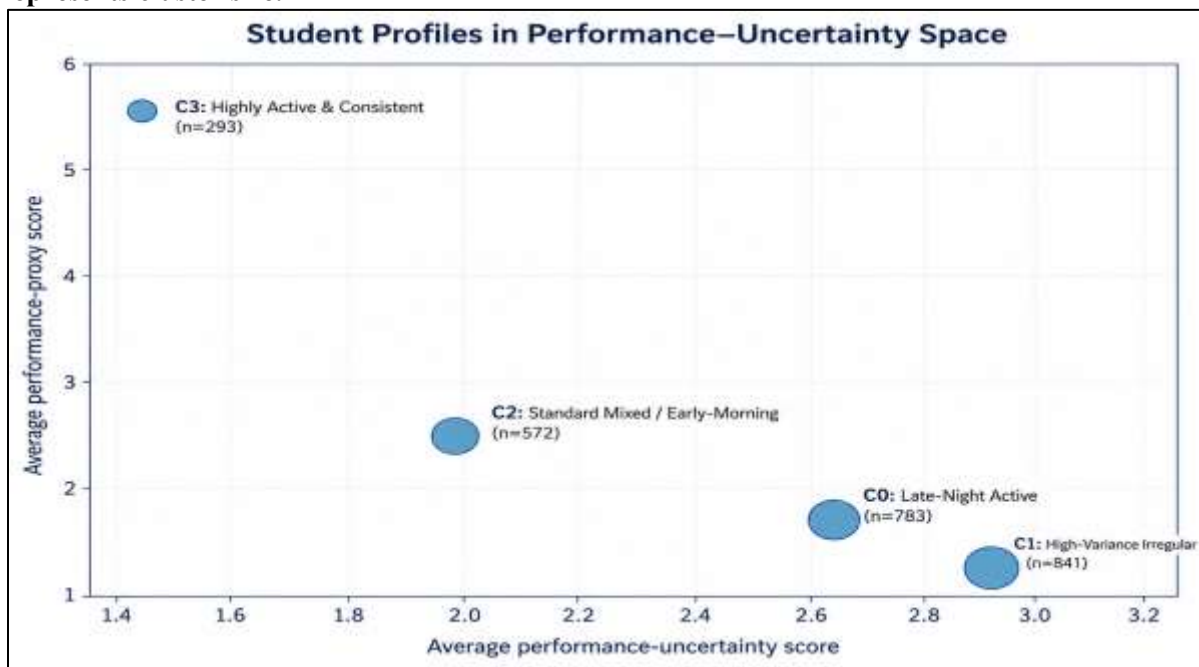


Figure 2. Student profiles in modeled performance-uncertainty space. Bubble size is proportional to cluster size.

The clustering diagnostics will need an expert interpretation. As seen in Table 3, the silhouette coefficient is highest for $K=5$ at 0.323, and $K=4$ at 0.297. The four-cluster solution thus sacrificed geometric separation for a less complicated profile structure. If this tradeoff is disclosed when making a decision-analysis demonstration, then it is acceptable, but it should not be assumed that the labels represent uniquely identified natural groups. A more robust analysis using the raw repeated fitness data would examine whether the minimax recommendation is robust across multiple reasonable cluster solutions for the data and not just uncertainty weights.

Table 3. K-means silhouette diagnostics

Candidate solution	Silhouette coefficient	Decision
K=2	0.199	Not retained
K=3	0.262	Not retained

Candidate solution	Silhouette coefficient	Decision
K=4	0.297	Retained for interpretability
K=5	0.323	Highest silhouette; not retained
K=6	0.291	Not retained
K=7	0.270	Not retained
K=8	0.258	Not retained
K=9	0.257	Not retained
K=10	0.274	Not retained

Note. K=5 produced the highest coefficient. K=4 was retained by the supplied notebook for interpretability and compact profile structure; this is a limitation and is disclosed explicitly.

4.3 Minimax Allocation

The worst case adjusted performance matrix is presented in Table 4. Among Cluster 0, Foundational Consistency had the highest robust value (-0.625), with Baseline (-1.025), Robust Behavioral Support (-2.725), and Performance Enhancement (-2.225) on the lower end. Foundational Consistency topped the pack again for Cluster 1 with -1.498. For Cluster 2, it had the highest value of 1.083. The same was true for Cluster 3, which had a baseline score of 4.109, higher than Foundational Consistency (3.809) and Performance Enhancement (3.509). The model thus suggested FC for Clusters 0-2 and Baseline for Cluster 3.

Figure 3 provides the full cluster-by-strategy worst-case-adjusted performance matrix before Table 4 reports the same values numerically.



Figure 3. Worst-case-adjusted performance by cluster and strategy ($\lambda=1.0$; cost scale $\gamma=0.5$).

Table 4. Main minimax decision results

Cluster	Profile	FC	PE	RBS	Baseline	Selected strategy
Cluster 0	Late-Night Active	-0.625	-2.225	-2.725	-1.025	Foundational Consistency
Cluster 1	High-Variance Irregular	-1.498	-2.998	-3.298	-1.698	Foundational Consistency

Cluster	Profile	FC	PE	RBS	Baseline	Selected strategy
Cluster 2	Standard Mixed / Early-Morning	1.083	-0.817	-1.617	0.483	Foundational Consistency
Cluster 3	Highly Active & Consistent	3.809	3.509	1.809	4.109	Baseline

Note. Values are worst-case-adjusted performance W_{cs} . FC = Foundational Consistency; PE = Performance Enhancement; RBS = Robust Behavioral Support. Strategy effects and costs are scenario assumptions, not estimated treatment effects.

The pattern is self-explanatory in terms of the scenario design. FC was low cost and had moderate improvements in performance and uncertainty for the 3 less favourable profiles. It scored 0.400 higher than Baseline in the high demanding objective (Cluster 0); 0.200 more in the medium demanding objective (Cluster 1); and 0.600 more in the low demanding objective (Cluster 2). The added value of a training scenario did not outweigh the cost penalty in Cluster 3 where the current performance was already high and uncertainty low. Baseline was therefore optimal, by +0.300 relative to FC. This is not a statement that highly active students don't need training. It does not imply that the extra cost of the intervention was not worth the assumed incremental benefit in this scenario matrix and objective.

4.4 Comparison with Alternative Decision Rules

The outcomes of both model types are compared via Figure 4 and Table 5. The average of the modeled performance from the minimax allocation was 2.641, the average of the modeled uncertainty was 2.170, and the average of the modeled scenario cost was 0.882. The one-size-fits-all Baseline had a performance of 2.211 and uncertainty of 2.512 at \$0 cost. The minimax allocation resulted in a 19.47% improvement in the modeled performance index and a 13.62% reduction in the modeled uncertainty index, compared to Baseline. The expected-performance rule yielded a performance of 2.506 and uncertainty of 2.271 and cost 0.544 or 13.35% modeled performance increase and 9.58% modeled uncertainty reduction compared to Baseline.

Figure 4 shows the modeled average performance under the four decision rules. Table 5 adds uncertainty, scenario cost, and the lambda sensitivity results.

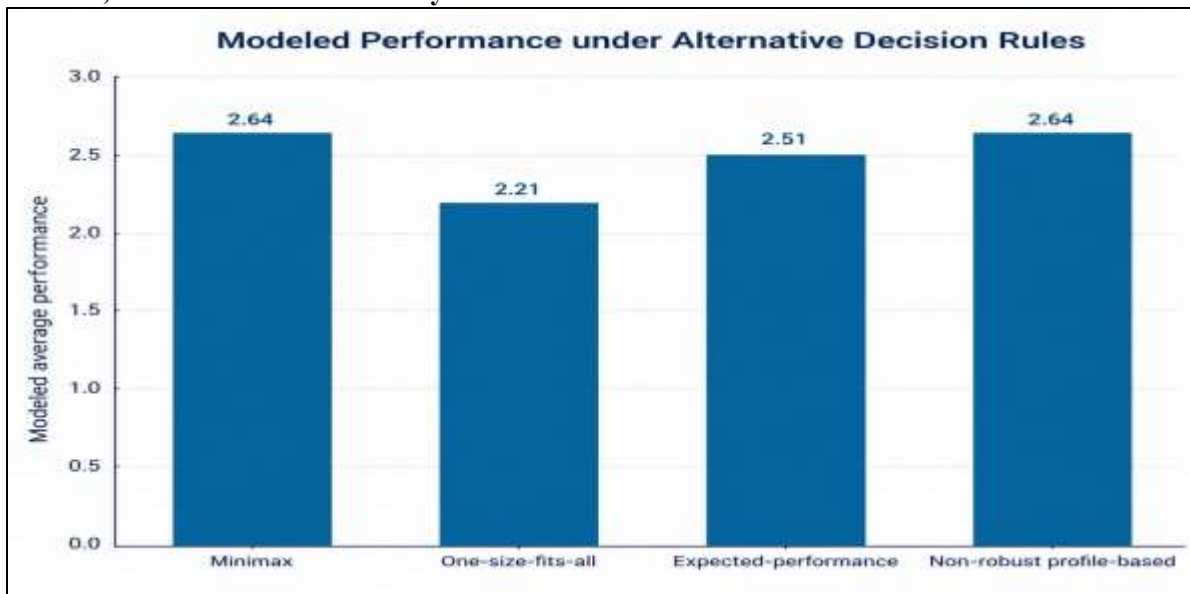


Figure 4. Modeled average performance under alternative decision rules. Values are scenario-model outputs, not observed intervention effects.

Table 5. Comparator and sensitivity results

Decision rule / sensitivity	Avg. performance	Avg. uncertainty	Avg. scenario cost	Profile-level allocation
Minimax	2.641	2.170	0.882	FC / FC / FC / Baseline
One-size-fits-all	2.211	2.512	0.000	Baseline for all clusters
Expected-performance	2.506	2.271	0.544	Notebook comparator
Non-robust profile-based	2.641	2.170	0.882	Same aggregate outcome as minimax
Sensitivity: $\lambda=0.5$	-	-	-	FC / FC / FC / Baseline
Sensitivity: $\lambda=1.0$	-	-	-	FC / FC / FC / Baseline
Sensitivity: $\lambda=1.5$	-	-	-	FC / FC / FC / Baseline

Note. Relative to one-size Baseline, minimax modeled performance was 19.47% higher and modeled uncertainty 13.62% lower. The non-robust profile comparator produced the same aggregate allocation in the supplied notebook.

The non-robust profile-based rule yielded the same aggregate performance, uncertainty and cost as the minimax rule were in the provided notebook. This equality is not a success to be concealed, but rather it is a statement of information. It shows that when looking at the scenario matrices, the profile ranking and the robust ranking coincide. That is, there is not a strong pressure to average profile benefit versus worst case protection in the current scenario. A richer intervention-response matrix might lead to different decisions, and would offer a more discriminating test of minimax value. It is therefore the present result that more clearly illustrates the mechanics and transparency of the robust rule than the superiority over all profile-based alternatives.

4.5 Sensitivity to the Uncertainty Weight

Figure 5 summarizes the lambda sensitivity analysis. At lambda values of 0.5, 1.0, and 1.5, the selected strategies were unchanged: FC for Clusters 0, 1, and 2, and Baseline for Cluster 3. This stability shows that the recommendation is not driven by a narrow choice of the uncertainty weight within the tested range. It does not, however, prove global robustness. The result remains conditional on fixed strategy-effect and cost matrices. Future work should vary those matrices directly, use empirically estimated uncertainty bounds, and test alternative cluster solutions.

Figure 5 visualizes the recommendation stability across the three tested uncertainty weights.

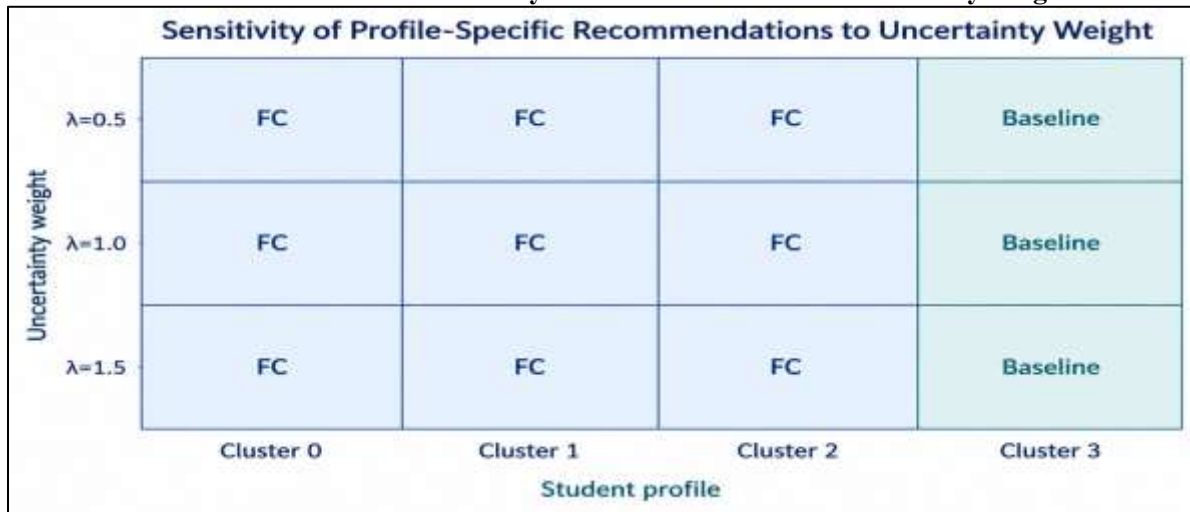


Figure 5. Sensitivity of profile-specific recommendations to the uncertainty weight lambda. FC = Foundational Consistency.

4.6 Main Empirical Conclusion

The observed data facilitates the identification of different behavioural profiles and a repeatable illustration of a minimax allocation rule. Does not make fitness enhancement claims for the selected strategies. The numerical results from Table 5 are model results for the scenarios. Thus, the most defensible conclusion is that when the performance uncertainty representation is at the profile level, and there are explicit assumptions, minimax logic can generate stable and interpretable training priorities. Those priorities are not proven to be effective in the real world in the data.

5. Discussion

5.1 What the Profile Structure Means

The four-cluster solution indicates that the volume of activity, consistency in behavior and modeled performance are not co-occurring in the same manner for all students. Cluster 3 has the highest volume of active days, and the lowest uncertainty model results, whereas Clusters 0 and 1 have more consistent patterns of activity that tend to be late at night. This helps to ensure that training priorities are tailored, rather than standardised. The present profiles are shorter than the full physical-literacy profiles but the same practical concerns remain: students who are similar on one profile may be different on another profile relevant to training needs. The Chinese student population corroborates this. Demographic and interpersonal factors, physical activity, and subjective well-being have been linked to physical literacy (Yang et al., 2024; Chen et al., 2024). These constructs are not directly measured here so are not considered as observed mechanisms. They do, however, emphasize that the amount of activity is not enough to determine training priorities.

5.2 Why Minimax Adds a Distinct Decision Layer

The beauty of minimax reasoning is that it doesn't have to give the same recommendation as the conventional rules. The value that it brings is that it explicitly identifies the thing being protected. The one-size-fits-all rule fails to account for differences in profiles and the expected performance rule focuses on average modeled gain. The minimax rule takes each strategy and considers a worst-case loss under an admissible loss function, and chooses the strategy which performs best over this protected loss. This is important if the loss of performance for a vulnerable profile is deemed to be more expensive than the loss of average gain. This is consistent with the robust-optimization theory, which considers uncertainty as a part of the decision itself (Shapiro, 2021; Rahimian & Mehrotra, 2022). The formal robust counterpart enhances transparency. The worst-case adjusted score is the true value of the inner minimization problem, and is not simply a random penalty. The uncertainty set is still a modelling decision, however. As demonstrated by robust-optimization research, the definition of that set has a significant impact on the usefulness (Duchi et al., 2021; Blanchet et al., 2022; Gao & Kleywegt, 2023). The recommendation was the same for all the lambda values (0.5 to 1.5). This means local decision stability and not global robustness. The test doesn't affect the strategy-effect matrix, it only shifts the decision-maker's preference for uncertainty protection. A multi-variate analysis of performance gain, uncertainty reduction and costs in the future should allow for varying these factors.

5.3 Interpreting the Apparent Gain Over Baseline

The minimax allocation yielded a 19.47% increase in modeled performance and a 13.62% decrease in modeled uncertainty compared with the one size fits all Baseline. The percentages are relative to rules within the model. No improvements in student fitness are noted. The performance measure is an artificial indicator and not a physiological measure. For claims of real training adaptation, there must be observable exposure to the intervention, and there must be measured outcomes (Weakley et al., 2023). The reported gains should thus be interpreted as decision-model gains and not treatment effects. The same overall result was obtained with the minimax and non-robust profile-based rules. This shows one of the big shortcomings of the existing scenario design. In the current parameterisation, the improvement over the universal Baseline is largely due to the effect of personalising the profile. The minimax layer introduces a formal downside protection criterion, and a formalisation of adverse state assumptions, but does not alter the overall allocation compared to the non-robust comparator. Future data should be sufficiently uncertain of the strategies to offer a clear empirical benefit of minimax, such that there is a genuine tradeoff between up and down protection and gain.

5.4 Implications for Student-Centered Physical Training

The three profiles with lower modeled performance or higher uncertainty were chosen as the profiles where the consistency of the foundation is chosen. This can work with strategies focusing on sustainable participation first

and then more demanding performance. Competence, confidence, motivation, and participation are typically components of physical-literacy interventions, and are considered central to the concept and more prominent than the maximal output (Carl et al., 2022a, 2022b). Because the Baseline option was already high activity and low modeled uncertainty, cluster 3 was given the Baseline option. This is not to say that there is no need for training for highly active students. In reality, performance enhancement, recovery management, specialization or injury prevention are all possible needs for such students. Those needs are not measured in the current dataset.

5.5 Methodological Implications for Minimax Applications

The following methodological principles emerge. First is that a solid objective should clearly state the negative variable. A weighted score that looks good might not have a worst-case problem. Second, all the scenario parameters should be labelled as to their evidence status. It is not acceptable to combine observed estimates with expert judgments or hypothetical values. Thirdly, sensitivity analysis should take into consideration both preference parameters and structural assumptions. The points are similar to concerns raised about reproducibility. While robustness is not a synonym for reproducibility, nor is precision in computational outputs a sign of precision in scientific evidence (Nosek et al., 2022; Kapoor & Narayanan, 2023). Once an objective and the inputs to an optimizer are given, an optimizer will always give a clear answer.

5.6 Relation to Contemporary Sports Analytics

The trend of sports analytics is to integrate monitoring, prediction and decision support. They can bolster sport management with big data and wearable sensors can deliver comprehensive data on movement and physiological condition (Watanabe et al., 2021; Rana & Mittal, 2020). But more data does not eliminate the need for decision theory. The more diverse the information, the more explicit the objectives, adverse outcomes, and uncertainty will need to be. Minimax optimization can be useful if there is a need to maximise the performance of the profile under the worst-case scenario instead of just maximising modeled performance. The current structure is intended to be a transition from a behavioral profile to strong decision support. The next phase involves repeated, objective fitness assessments, training exposure, adherence, workloads, fatigue and injury markers, plus time-varying context.

6. Theoretical, Methodological, and Practical Implications

6.1 Theoretical Contribution

The study extends minimax reasoning to the case of a student-fitness decision where the profile-level training decisions are mapped to a bounded adverse loss problem. It proves that a worst-case adjusted score is genuinely minimax if there is an uncertainty set and an inner adverse choice are explicitly mentioned. This distinguishes the formal from the generic view of robust optimization versus penalty scoring and makes it clear what λ means (uncertainty preference) and what γ means (intervention-cost weight).

6.2 Methodological Contribution

The analysis illustrates how to establish a clear decision process for when there is sufficient information to segment but not to estimate treatment effects. The performance measure is referred to as a proxy, the uncertainty measure is considered a type of decision feature (a cross-sectional one), and dP and dU are seen as scenario parameters instead of empirical estimates. Such decisions help to avoid overstatement and will help create workflow that can be enhanced later with more comprehensive longitudinal data.

6.3 Practical Implications

The framework that the authors recommend for physical-education administrators or digital training systems is straightforward: assess students' profile, identify uncertainty, establish reasonable training priorities, outline potential gains and losses, and choose the training that will yield the best possible outcome, even in the face of uncertainty. The scenario matrix should be based on real applications and be estimated using pilot programs, repeated fitness testing or structured expert elicitation.

7. Limitations and Future Research

Several limitations remain. There are no direct physical fitness measures such as aerobic capacity, sprint speed, strength, power, flexibility, or motor competence in the dataset. The performance proxy is thus an operational, rather than measured fitness index. Standardized tests should be included in future data collections, including the

50 m sprint, 800 m or 1000 m endurance run, the standing long jump, sitting and reaching, and suitable strength tests.

The files analyzed also do not have repeats. The uncertainty index is cross-sectional and it is not a measure of intra-individual uncertainty, uncertainty due to training, or test-retest reliability. If repeated trajectories were available, some measure of the uncertainty could be derived from the repeated trajectories, residual variability, or predictive intervals.

The training effects and costs are input values of hypothetical scenarios and not observed effect sizes. To assess performance and uncertainty changes attributable to a strategy, randomized or well-designed quasi-experimental data are required. Another important point is that source variables are highly imbalanced, with the silhouette coefficient being higher for $K=5$ than for $K=4$. Alternative clustering solutions should be tested in the future and a determination made as to the stability of recommendations. Lastly, the repository does not have a proven sampling frame and limited metadata for national representativeness, so the results should be taken with caution and not be extrapolated to the national student population in China. In addition, future testing should be conducted to compare other minimax formulations such as minimax regret, and distributionally robust minimax, and see how the same priorities for profiles persist when there are changing conditions and assumptions and changing uncertainty structures.

8. Conclusion

The aim of this study was to design a clear minimax algorithm for personalized training priorities based on student activity profile data in China. In this study, a clear minimax algorithm for student prioritization of training based on student activity profile data in China was proposed. Four behavioral profiles were identified through the analysis, and model-based performance and uncertainty indices were employed to aid in formulating a minimax decision based on a finite number of scenarios. Foundational Consistency was chosen for the three lower performing and/or less certain profiles and Baseline for the very active and consistent profile. The suggested rate was the same for lambda values of 0.5, 1.0, and 1.5.

Main contribution is methodological. An explicit statement of downside protection is included in the framework, directly derived from minimax logic, when performance is uncertain, but at the same time the distinction between what has been observed and what has been assumed regarding the model is clear. The findings are not evidence of the objective fitness improving strategies selected. This must be backed up by fitness scores over time and by training responses observed. In this context, the study offers a repeatable and traceable use of the minimax theory in the process of personalized training decision making within these bounds.

REFERENCES

1. Blanchet, J., Murthy, K., & Si, N. (2022). Confidence regions in Wasserstein distributionally robust estimation. *Biometrika*, 109(2), 295-315. <https://doi.org/10.1093/biomet/asab026>
2. Carl, J., Barratt, J., Töpfer, C., Cairney, J., & Pfeifer, K. (2022b). How are physical literacy interventions conceptualized? A systematic review on intervention design and content. *Psychology of Sport and Exercise*, 58, 102091. <https://doi.org/10.1016/j.psychsport.2021.102091>
3. Carl, J., Barratt, J., Wanner, P., Töpfer, C., Cairney, J., & Pfeifer, K. (2022a). The effectiveness of physical literacy interventions: A systematic review with meta-analysis. *Sports Medicine*, 52(12), 2965-2999. <https://doi.org/10.1007/s40279-022-01738-4>
4. Chen, C., Weyland, S., Fritsch, J., Woll, A., Niessner, C., Burchartz, A., Schmidt, S. C. E., & Jekauc, D. (2021). A short version of the Physical Activity Enjoyment Scale: Development and psychometric properties. *International Journal of Environmental Research and Public Health*, 18(21), 11035. <https://doi.org/10.3390/ijerph182111035>
5. Chen, W., Zhou, B., Wang, X., & Li, L. (2024). Research on the relationship between physical literacy and demographic variables and interpersonal support for physical exercise among adolescents in China. *PLOS ONE*, 19(10), e0311793. <https://doi.org/10.1371/journal.pone.0311793>
6. Choi, S. M., Sum, K. W. R., Leung, F. L. E., Wallhead, T., Morgan, K., Milton, D., Ha, S. C. A., & Sit, H. P. C. (2021). Effect of sport education on students' perceived physical literacy, motivation, and physical activity levels in university required physical education: A cluster-randomized trial. *Higher Education*, 81(6), 1137-1155. <https://doi.org/10.1007/s10734-020-00603-5>
7. Cornish, K., Fox, G., Fyfe, T., Koopmans, E., Pousette, A., & Pelletier, C. A. (2020). Understanding physical literacy in the context of health: A rapid scoping review. *BMC Public Health*, 20, 1569. <https://doi.org/10.1186/s12889-020-09583-8>

8. Duchi, J. C., Glynn, P. W., & Namkoong, H. (2021). Statistics of robust optimization: A generalized empirical likelihood approach. *Mathematics of Operations Research*, 46(3), 946-969. <https://doi.org/10.1287/moor.2020.1085>
9. Düking, P., Zinner, C., Trabelsi, K., Reed, J. L., Holmberg, H.-C., Kunz, P., & Sperlich, B. (2021). Monitoring and adapting endurance training on the basis of heart rate variability monitored by wearable technologies: A systematic review with meta-analysis. *Journal of Science and Medicine in Sport*, 24(11), 1180-1192. <https://doi.org/10.1016/j.jsams.2021.04.012>
10. Gao, R., & Kleywegt, A. J. (2023). Distributionally robust stochastic optimization with Wasserstein distance. *Mathematics of Operations Research*, 48(2), 603-655. <https://doi.org/10.1287/moor.2022.1275>
11. Grosz, M. P., Rohrer, J. M., & Thoemmes, F. (2020). The taboo against explicit causal inference in nonexperimental psychology. *Perspectives on Psychological Science*, 15(5), 1243-1255. <https://doi.org/10.1177/1745691620921521>
12. Hardwicke, T. E., Bohn, M., MacDonald, K., Hembacher, E., Nuijten, M. B., Peloquin, B. N., deMayo, B. E., Long, B., Yoon, E. J., & Frank, M. C. (2021). Analytic reproducibility in articles receiving open data badges at the journal *Psychological Science*: An observational study. *Royal Society Open Science*, 8, 201494. <https://doi.org/10.1098/rsos.201494>
13. Houser, N., & Kriellaars, D. (2023). 'Where was this when I was in physical education?' Physical literacy enriched pedagogy in a quality physical education context. *Frontiers in Sports and Active Living*, 5, 1185680. <https://doi.org/10.3389/fspor.2023.1185680>
14. Impellizzeri, F. M., Marcora, S. M., & Coutts, A. J. (2020). Internal and external training load: 15 years on. *International Journal of Sports Physiology and Performance*, 15(2), 270-273. <https://doi.org/10.1123/ijsp.2019-0935>
15. Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*, 4(9), 100804. <https://doi.org/10.1016/j.patter.2023.100804>
16. Knief, U., & Forstmeier, W. (2021). Violating the normality assumption may be the lesser of two evils. *Behavior Research Methods*, 53, 2576-2590. <https://doi.org/10.3758/s13428-021-01587-5>
17. Kwan, M. Y. W., Graham, J. D., Healey, C., Paolucci, N., & Brown, D. M. (2020). Stopping the drop: Examining the impact of a pilot physical literacy-based intervention program on physical activity behaviours and fitness during the transition into university. *International Journal of Environmental Research and Public Health*, 17(16), 5832. <https://doi.org/10.3390/ijerph17165832>
18. Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, 8(1), 33267. <https://doi.org/10.1525/collabra.33267>
19. Li, K., Cardoso, C., Moctezuma-Ramirez, A., Elgalad, A., & Perin, E. (2023). Heart rate variability measurement through a smart wearable device: Another breakthrough for personal health monitoring? *International Journal of Environmental Research and Public Health*, 20(24), 7146. <https://doi.org/10.3390/ijerph20247146>
20. Li, M. H., Sum, R. K. W., Tremblay, M., Sit, C. H. P., Ha, A. S. C., & Wong, S. H. S. (2020). Cross-validation of the Canadian Assessment of Physical Literacy Second Edition (CAPL-2): The case of a Chinese population. *Journal of Sports Sciences*, 38(24), 2850-2857. <https://doi.org/10.1080/02640414.2020.1803016>
21. Liu, Y., Yan, J., & Li, J. (2025). The relationship between physical education teachers' competence support and middle school students' participation in sports: A chain mediation model of perceived competence and exercise persistence. *PLOS ONE*, 20(1), e0314338. <https://doi.org/10.1371/journal.pone.0314338>
22. Lundberg, I., Johnson, R., & Stewart, B. M. (2021). What is your estimand? Defining the target quantity connects statistical evidence to theory. *American Sociological Review*, 86(3), 532-565. <https://doi.org/10.1177/00031224211004187>
23. Melby, P. S., Elsborg, P., Bentsen, P., & Nielsen, G. (2023). Cross-sectional associations between adolescents' physical literacy, sport and exercise participation, and wellbeing. *Frontiers in Public Health*, 10, 1054482. <https://doi.org/10.3389/fpubh.2022.1054482>
24. Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, 73, 719-748. <https://doi.org/10.1146/annurev-psych-020821-114157>

25. O'Connor, J., & Penney, D. (2021). Informal sport and curriculum futures: An investigation of the knowledge, skills and understandings for participation and the possibilities for physical education. *European Physical Education Review*, 27(1), 3-26. <https://doi.org/10.1177/1356336X20915937>
26. Rahimian, H., & Mehrotra, S. (2022). Distributionally robust optimization: A review. *Open Journal of Mathematical Optimization*, 3, 4. <https://doi.org/10.5802/ojmo.15>
27. Rana, M., & Mittal, V. (2020). Wearable sensors for real-time kinematics analysis in sports: A review. *IEEE Sensors Journal*, 21(2), 1187-1207. <https://doi.org/10.1109/JSEN.2020.3019016>
28. Rudd, J. R., Pesce, C., Strafford, B. W., & Davids, K. (2020). Physical literacy-A journey of individual enrichment: An ecological dynamics rationale for enhancing performance and physical activity in all. *Frontiers in Psychology*, 11, 1904. <https://doi.org/10.3389/fpsyg.2020.01904>
29. Santos, F., Newman, T. J., Aytur, S., & Farias, C. (2022). Aligning physical literacy with critical positive youth development and student-centered pedagogy: Implications for today's youth. *Frontiers in Sports and Active Living*, 4, 845827. <https://doi.org/10.3389/fspor.2022.845827>
30. Schmittwilken, L., Harding-Kuriger, J., & Carl, J. (2024). Aligning extracurricular school activities with physical literacy: Pilot evaluation through self-study of practice. *Frontiers in Sports and Active Living*, 6, 1415689. <https://doi.org/10.3389/fspor.2024.1415689>
31. Shapiro, A. (2021). Tutorial on risk neutral, distributionally robust and risk averse multistage stochastic programming. *European Journal of Operational Research*, 288(1), 1-13. <https://doi.org/10.1016/j.ejor.2020.05.047>
32. Stoddart, A. L., & Humbert, M. L. (2021). Teachers' perceptions of physical literacy. *Curriculum Journal*, 32(4), 741-757. <https://doi.org/10.1002/curj.107>
33. Tennant, P. W. G., Murray, E. J., Arnold, K. F., Berrie, L., Fox, M. P., Gadd, S. C., Harrison, W. J., Keeble, C., Ranker, L. R., Textor, J., Tomova, G. D., Gilthorpe, M. S., & Ellison, G. T. H. (2021). Use of directed acyclic graphs to guide analyses of causal questions: A review. *International Journal of Epidemiology*, 50(2), 620-632. <https://doi.org/10.1093/ije/dyaa213>
34. Van Parys, B. P. G., Mohajerin Esfahani, P., & Kuhn, D. (2021). From data to decisions: Distributionally robust optimization is optimal. *Management Science*, 67(6), 3387-3402. <https://doi.org/10.1287/mnsc.2020.3678>
35. Watanabe, N. M., Shapiro, S., & Drayer, J. (2021). Big data and analytics in sport management. *Journal of Sport Management*, 35(3), 197-202. <https://doi.org/10.1123/jsm.2021-0067>
36. Weakley, J., Cowley, N., Schoenfeld, B. J., Read, D. B., Timmins, R. G., Garcia-Ramos, A., & McGuckian, T. B. (2023). The effect of feedback on resistance training performance and adaptations: A systematic review and meta-analysis. *Sports Medicine*, 53(9), 1789-1803. <https://doi.org/10.1007/s40279-023-01877-2>
37. Yang, X., Wang, M., Wang, J., Zhang, S., Yang, X., & Zhao, L. (2024). Physical literacy and health of Chinese medical students: The chain mediating role of physical activity and subjective well-being. *Frontiers in Public Health*, 12, 1348743. <https://doi.org/10.3389/fpubh.2024.1348743>
38. Yin, H., Omar Dev, R. D., Soh, K. G., Li, F., & Lian, M. (2024). Assessment and development of physical education teachers' physical literacy: A systematic review. *PLOS ONE*, 19(7), e0307505. <https://doi.org/10.1371/journal.pone.0307505>