Protocol Dependence and Evidence Divergence in the Effects of Simulated Hypoxic Training on Cardiorespiratory Fitness in Healthy College Students: A Narrative Review

Song Shuailong[1], Cao Leyi[1], Shi Bing[2]
[1]School of Physical Education, Qinghai Normal University, Xining, Qinghai, China
[2]School of Physical Education, Shaanxi Normal University, Xi'an, Shaanxi, China

Abstract: Background: Simulated hypoxic training has been increasingly studied in healthy college students, but evidence remains inconsistent regarding whether hypoxia confers additional benefits over intensity-matched normoxic training. Objective: This narrative review examines the sources of divergence in the literature on simulated hypoxic training and cardiorespiratory fitness in healthy college students. Approach: We juxtapose Chinese college-student intervention studies with international synthesis evidence and propose an analytical framework for interpreting divergent effect estimates. Argument: Divergence is not simply a matter of "insufficient evidence." It arises from differences across five domains: the level of comparison, hypoxic dose, intensity anchoring, simulation modality, and outcome temporal window. Chinese college-student studies report signals in cardiac pumping function, hemorheology, antioxidant defense, and some aerobic performance indicators, but these signals are based on small samples, heterogeneous protocols, and incomplete reporting of dose and safety parameters. International synthesis evidence answers different questions at different levels: network meta-analysis ranks hypoxic paradigms within hypoxic training; meta-regression stratifies athletes and non-athletes; and strict meta-analysis directly compares hypoxia with matched normoxia. These answers can coexist logically, but they cannot be compressed into a single claim that "hypoxic training is effective." Conclusions: Future research should move from overall effect testing toward cumulative conditional-effect propositions: under what hypoxic dose, intensity-anchoring method, simulation modality, outcome dimension, and time window does hypoxia produce reproducible and practically meaningful differences relative to matched normoxic training.

Keywords: Simulated Hypoxic Training; Healthy College Students; Cardiorespiratory Fitness; Maximal Oxygen Uptake; Evidence Synthesis

1. Introduction

Cardiorespiratory fitness in healthy college students is an intersection between the evaluation of physical education outcomes and public health monitoring, and it is also an outcome variable with relatively high measurement accessibility in intervention research. Under normoxic conditions, both moderate-intensity continuous training and high-intensity interval training can improve maximal oxygen uptake (VO_2max), but each has constraints in university settings: moderate-intensity continuous training requires a relatively long single-session time investment to accumulate sufficient cardiovascular stimulus, whereas high-intensity interval training imposes higher demands on exercise tolerance and joint and muscle loading. The rationale for introducing simulated hypoxic environments into research is quite direct: at a given absolute power output, hypoxic inhalation increases the relative stress on the cardiorespiratory system, thereby potentially inducing adaptations in aerobic fitness at a lower mechanical load.

This reasoning has a substantial foundation in the athlete altitude-training literature, but when transferred to healthy college students, one premise needs to be re-examined—whether hypoxia provides an adaptive signal additional to intensity-matched normoxic training, or merely makes the same exercise feel more effortful.

English-language synthesis evidence over the past five years has exposed the importance of this premise, and the synthesis evidence itself presents an apparent opposition. Yu et al. (2023) conducted a network meta-analysis including 42

Copyright @ STEMM Institute Press

studies and 1,246 healthy adults, ranking intermittent hypoxic training (IHT) as the hypoxic paradigm with the highest probability (66%) of improving VO_2max in non-athlete populations, and reported a significant association between saturation hours and VO_2max improvement (coefficient 0.004, P = 0.038)[1]. Kohlbrenner et al. (2026), in a meta-regression published in Sports Medicine, further differentiated types of hypoxic strategies: among non-athletes, living-low–training-low (LTH, MD = 1.70, 95% CI 0.85–2.54) and passive hypoxic exposure (PHC, MD = 2.26, 95% CI 1.00–3.52) showed significant effects on VO_2max; the significant effect of living-high–training-low (LHTL) was observed only in athletes (MD = 2.17, 95% CI 0.77–3.56), and PHC even showed a non-significant negative trend in athletes (MD = −1.07)[2]. Dorelli et al. (2025), in a meta-analysis published in the Scandinavian Journal of Medicine & Science in Sports, adopted the strictest comparative logic: including only 35 parallel-design studies and 524 participants, they directly compared aerobic IHT with normoxic training and concluded that IHT had no significant advantage in VO_2max (p = 0.333), peak power output (p = 0.159), or time to exhaustion (p = 0.410); subgroup analyses found no moderating effect of training level (p = 0.690) or exercise modality (p = 0.900), and Egger's test indicated the presence of publication bias (p = 0.004)[3]. An umbrella review including 22 systematic reviews (14 meta-analyses and 3 network meta-analyses) and 487 primary studies adopted a broader position, suggesting that intermittent hypoxic protocols generally improve aerobic and anaerobic performance, but that the magnitude of effects is highly dependent on protocol type, training status, and hypoxic dose, and that these gains are often not accompanied by consistent hematological changes[4].

The tension among these studies is the core issue this review addresses. The tension does not mean that any one of them is wrong. Their divergence precisely delineates several variables in the current evidence system that have not yet been adequately controlled. For ease of subsequent analysis, this review categorizes these variables into five domains and presents an analytical framework for the sources of evidence divergence (Figure 1).

Chinese studies using college students as samples provide a concrete empirical level for these methodological issues. After Liu Xudong implemented 4-week IHT with a decreasing oxygen concentration in ordinary male college students, left ventricular ejection time shortened from 288.6 ± 8.4 ms to 269.3 ± 7.6 ms, stroke volume increased from 52.3 ± 3.8 mL to 55.4 ± 7.2 mL, and cardiac output increased from 3.53 ± 0.31 L/min to 3.82 ± 0.38 L/min; all three reached significant levels, while VO_2max showed an increasing trend but did not reach significance; under the same protocol, red blood cell count and hematocrit increased significantly, erythrocyte deformability improved significantly, and whole blood viscosity and plasma viscosity showed no significant changes[5]. Lu Baoyun conducted 10-week hypoxic treadmill training in non-physical-education female college students and found that VO_2max, vital capacity, and step-test index in the hypoxic group were all very significantly better than those in the normoxic group (P < 0.01) [6]. Hou Zheng et al. used an ETM high-altitude respiratory mask for hypoxic endurance training, and reported that vital capacity, forced expiratory volume in the first second, and VO_2max in experimental group II were significantly improved in both within-group and between-group tests[7]. Tu Yanhao et al. used an artificial oxygen-restricting mouthpiece in 30 plains college students during 35 days of equal-load training; aerobic exercise capacity in the hypoxic group improved significantly, and the improvement in 800-m performance was greater than that in the ordinary group[8]. The inconsistency in the direction of results is precisely the best entry point for understanding the structure of hypoxic training effects.

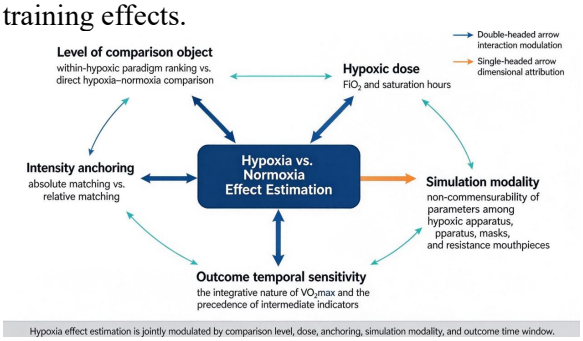


Figure 1. Structural Sources of Divergence in Effect Estimates of Simulated Hypoxic Training: An Analytical Framework

This narrative review focuses on healthy college students and on cardiorespiratory outcomes relevant to simulated hypoxic training. It does not aim to produce a pooled effect estimate; rather, it selects evidence that illustrates the structural sources of divergence in the literature.

 Copyright @ STEMM Institute Press

Chinese-language studies were searched in CNKI and Wanfang Data using combinations of "hypoxic training," "intermittent hypoxia," "simulated hypoxia," "college students," "cardiopulmonary function," and "maximal oxygen uptake." International evidence was searched in PubMed, Web of Science, Scopus, and SPORTDiscus using combinations of "hypoxic training," "intermittent hypoxic training," "normobaric hypoxia," "VO2max," "cardiorespiratory fitness," "healthy adults," and "college students." The search was last updated on September 30, 2026. Reference lists of included reviews were hand-searched.

Studies were considered eligible if they (1) involved healthy college students or healthy young adults without cardiovascular, metabolic, or chronic respiratory disease; (2) used a simulated hypoxic intervention; (3) reported cardiorespiratory, hematological, vascular, autonomic, or aerobic performance outcomes; and (4) were published in Chinese or English. Studies in obese or overweight populations were treated as boundary evidence rather than core evidence for healthy college students. Conference abstracts were included only when they provided otherwise unavailable protocol-level data and were explicitly labeled as such.

Because the included literature was heterogeneous in design, population, and outcome reporting, no pooled effect estimate was calculated. Instead, studies were mapped according to five domains: level of comparison, hypoxic dose, intensity anchoring, simulation modality, and outcome temporal window. Risk of bias was appraised narratively using design-appropriate criteria: the Cochrane risk-of-bias tool for randomized trials, ROBINS-I for non-randomized studies, and AMSTAR-2 or ROBIS for systematic reviews and meta-analyses. Evidence was not graded using GRADE because the review question was explanatory rather than effect-estimating. The key characteristics of the included studies are summarized narratively in the relevant sections.

2. Evidence Landscape in Chinese College Student Samples: Heterogeneity as an Analytical Object

The Chinese college-student literature is characterized less by a single homogeneous intervention than by a family of loosely related protocols. At least four categories of simulated hypoxic modalities have been used: hypoxic apparatus with respiratory masks, artificial oxygen-restricting mouthpieces, ETM high-altitude respiratory masks, and hypoxic sleep. Intervention periods range from 4 to 10 weeks, and outcome reporting varies from studies using VO_2max as the sole primary outcome to studies reporting cardiac pumping function, hemorheology, antioxidant indicators, or submaximal performance simultaneously. This heterogeneity should not be treated simply as an insufficient-evidence gap. Rather, it is an analytical object in its own right: differences in modality, duration, and outcome selection may themselves explain why studies point in different directions.

2.1 Four-Week Intermittent Hypoxia: Antecedent Improvement in Cardiac Pumping Function and Non-Significant VO_2max

Two conference papers by Liu Xudong reported the most complete short-cycle IHT protocol and multidimensional cardiac function data in the Chinese literature[5]. The protocol used a domestic ZL-I hypoxic apparatus, with intermittent inhalation of hypoxic gas through a respiratory mask, alternating 5 minutes of hypoxia and 5 minutes of normoxia, 6 cycles per session for a total of 60 minutes, 5 days per week for 4 consecutive weeks; oxygen concentration decreased progressively, with 14% in week 1, 12% in week 2, 11% in week 3, and 10% in week 4. After training, left ventricular ejection time shortened significantly, stroke volume and cardiac output increased significantly, and although VO_2max increased somewhat, it did not differ significantly from pre-training[5]. Another analysis of the same protocol showed that hemoglobin content increased but not significantly; red blood cell count and hematocrit were significantly higher than before training; erythrocyte deformability increased significantly; and whole blood viscosity, plasma viscosity, and erythrocyte aggregation showed no significant changes[5].

The core information of this result pattern is the separation of adaptation dimensions, rather than the ineffectiveness of hypoxia. Before overall aerobic fitness indexed by VO_2max had undergone detectable changes, central cardiac pumping function and peripheral hemorheology had already shown directionally consistent adjustments. From a mechanistic perspective, this separation is highly explicable. Acute

Copyright @ STEMM Institute Press

normobaric hypoxia (FiO_2 = 14.4%) reduced stroke volume and cardiac output at maximal exercise in healthy men; stroke volume reached a plateau earlier under hypoxic conditions, indicating that during the acute phase of a single hypoxic exercise session, cardiac pumping function is actually suppressed rather than enhanced[9]. The significant improvement in resting cardiac pumping function after 4-week IHT cannot be interpreted as a direct stimulating effect of hypoxia on the heart; it is more likely that repeated hypoxic–normoxic alternating exposure induced a regulatory adaptation during the reoxygenation phases: during each normoxic interval, the recovery of hemodynamic load and possible hyperemic responses accumulated over weeks into improved resting cardiac pumping function. The non-significant change in VO_2max suggests that this central-level adjustment was not yet sufficient to drive an overall increase in peak aerobic capacity, which may require synchronous changes in peripheral skeletal muscle oxidative capacity. The study sample consisted of 10 male college students, and statistical power was insufficient to detect small-to-moderate effects on VO_2max; however, the finding that cardiac pumping function changed before VO_2max suggests that the choice of outcome indicator itself shapes the answer to whether hypoxia is effective.

2.2 Long Duration and Different Simulation Modalities: Boundary Conditions of Positive Signals

Lu Baoyun's study constitutes another extreme in intervention duration and training volume. Thirty non-physical-education female college students were randomly divided into a hypoxic group and a normoxic group, both performing 10 weeks of treadmill training, 4 times per week, 30 minutes per session. In the hypoxic group, vital capacity, vital capacity/body weight index, VO_2max, step-test index, and 50-m and 800-m running performance all increased very significantly ($P < 0.01$), and all indicators in the hypoxic group were superior to those in the normoxic group; BMI and resting heart rate decreased significantly, while lean body mass and standing long jump performance increased significantly[6]. This result forms a direct contrast with the negative VO_2max conclusion of the 4-week study, but two qualifying conditions need attention. First, both groups received training, and the effect of hypoxia was evaluated on top of adaptations already produced by normoxic training, rather than compared with a non-training control group. Second, the study did not report effect sizes, randomization details, or dropouts; with 15 participants per group, the robustness of between-group significance is difficult to fully assess without such information. The difference in VO_2max conclusions between 4 and 10 weeks should therefore not be simply attributed to "the longer the period, the more effective." A more reasonable explanation involves a nonlinear relationship between total training volume and the additional effect of hypoxia: when the cumulative volume of normoxic training itself increases with longer periods, the marginal contribution of hypoxia may be compressed rather than amplified; the non-significant VO_2max in the short period may reflect that adaptations have not yet been integrated enough to change peak performance, whereas significance in the long period may partly arise from between-group separation after normoxic training effects accumulated in both groups.

Hou Zheng et al.'s study using an ETM high-altitude respiratory mask reported significant within-group and between-group improvements in vital capacity, forced expiratory volume in the first second, and VO_2max [7]. The dead-space effect and respiratory resistance of the ETM mask make it difficult to separate the two variables of hypoxia and respiratory muscle loading; the positive results may partly reflect respiratory muscle training effects rather than a pure hypoxic effect.

Tu Yanhao et al. used an artificial oxygen-restricting mouthpiece in 30 plains college students during 35 days of equal-load training; aerobic exercise capacity in the hypoxic group improved significantly, the improvement in 800-m performance was greater than that in the ordinary group, and the conclusion was that "simulated hypoxic environment training using an artificial oxygen-restricting mouthpiece in the plains is feasible"[8]. The FiO_2 of the resistance mouthpiece is difficult to quantify accurately, and increased respiratory work itself may introduce physiological perturbations beyond hypoxia. The common problem of these two studies is that when studies present results only under the label hypoxic group, readers cannot determine whether between-group differences arise from hypoxia itself, changes in respiratory muscle loading, or the interaction between the

 Copyright @ STEMM Institute Press

two.

An 8-week hypoxic training study in obese college students (oxygen concentration 14.7%, intensity 65%–75% VO_2max, 60 minutes per day, 5 days per week) reported that hypoxic intervention was superior to normoxic training in reducing body weight, body fat, and 3000-m performance, and that VO_2max in the experimental group reached 54.83 mL/(kg·min)[10]. This result indicates that under the hypoxic label, population characteristics (differences between obesity and normal weight) and training volume combinations have a substantive impact on the direction of conclusions, and also suggests that the distribution of baseline aerobic fitness is an important source of effect heterogeneity. Liu et al. (2025), in a systematic review and meta-analysis of obese or overweight populations, further showed that exercise interventions in hypoxic environments differ from those in normoxic environments in cardiovascular function and maximal power output, and that the magnitude of this difference is likewise modulated by protocol parameters and participant baseline characteristics[11].

2.3 Incommensurability of Simulation Modalities: from a Broad Label to a Comparable Intervention

The foregoing has shown that Chinese college student studies have used at least four categories of hypoxic simulation modalities: hypoxic apparatus with respiratory masks, artificial oxygen-restricting mouthpieces, ETM high-altitude respiratory masks, and hypoxic sleep. These four modalities are not equivalent in the nature of hypoxic exposure. A hypoxic apparatus can relatively precisely control FiO_2 and deliver stimuli in intermittent cycles; a resistance mouthpiece indirectly reduces inspired oxygen partial pressure by increasing respiratory resistance, and its FiO_2 is difficult to quantify precisely; the dead-space effect and respiratory resistance of the ETM mask likewise make it difficult to separate hypoxia and respiratory muscle load; hypoxic sleep centers on residence rather than training, and hypoxic and exercise stimuli are temporally separated.

This incommensurability has substantive consequences at the level of synthesis evidence. Pérez-Regalado et al. [12] directly compared two major administration modes of normobaric hypoxia in a systematic review and meta-analysis: compared with normoxia, normobaric hypoxia overall showed a moderate improvement in VO_2max (SMD = 0.74, $p = 0.06$), with the effect almost entirely driven by chamber/tent-type devices (SMD = 1.30, $p < 0.01$), whereas the effect of mask-type devices was non-significant; in sedentary populations and with a higher total number of training sessions, the advantage of chambers/tents was particularly evident; severe hypoxia did not produce conclusive VO_2max improvement[12]. The methodological implication of this finding is direct: if simulation modality itself systematically modulates the magnitude of effects, then pooling data from mask-type, resistance-mouthpiece-type, and chamber/tent-type studies into a single effect estimate of hypoxic training is methodologically inadequate. Chinese college student literature is particularly weak in standardized reporting of simulation modalities: measured or estimated FiO_2 values, respiratory resistance parameters, dead-space volume, and actual duration of each exposure were not completely reported in most studies[5,8]. This incommensurability means that seemingly consistent positive or directionally inconsistent results in Chinese literature must be discounted for differences in simulation modality before cross-study comparison.

Yun Dachuan and Hu Yongxin implemented simulated living-high–training-low in 13 male college students (8 hours of sleep in a hypoxic environment daily, O_2= 14%, for 4 consecutive weeks); in the living-high–training-low group, erythrocyte SOD and GSH-Px activities were significantly increased both at rest and after exhaustive exercise, MDA showed no change, and heart rate and blood lactate after quantitative load decreased[13]. This protocol is fundamentally different from IHT in the temporal structure of hypoxic exposure, with saturation hours far higher than short-duration IHT, and hypoxic and exercise stimuli temporally separated. When such results are juxtaposed with IHT data, failure to specify the temporal structure of exposure produces a category error: two interventions with different physiological stimulus properties are discussed as the same kind of hypoxic training. Kohlbrenner et al.'s stratified results have already shown that the significant effect of LHTL was not replicated in non-athletes (only one study met inclusion criteria), and that LTH

and PHC were the strategies showing significant effects in non-athletes[2]. Healthy college students are closer to non-athletes in training status; directly analogizing college student IHT data to athlete LHTL literature lacks justification at both the population and protocol levels.

3. Divergence Structure in English-Language Synthesis Evidence: Different Questions, Different Answers

The apparent contradictions among Yu et al., Kohlbrenner et al., and Dorelli et al. largely stem from the fact that they answer different comparative questions. Yu et al.'s network meta-analysis compared the relative ranking among multiple hypoxic paradigms; Kohlbrenner et al.'s meta-regression set the unit of analysis as the effect of hypoxic strategies relative to their respective control groups, with explicit stratification between athletes and non-athletes; Dorelli et al.'s strict meta-analysis directly compared hypoxic IHT with matched normoxic training. The three studies do not give opposing answers to the same question; rather, they answer at different question levels. Yu et al.'s network meta-analysis compared relative rankings among multiple hypoxic paradigms[1]; the conclusion that IHT had the highest probability (66%) in non-athletes holds within the context of ranking within hypoxic paradigms, answering "if one has already decided to use hypoxic training, which paradigm is most likely to improve VO_2max," rather than "whether hypoxic training is better than doing the same thing under normoxia." Kohlbrenner et al.'s meta-regression set the unit of analysis as the effect of hypoxic strategies relative to their respective control groups, with explicit stratification between athletes and non-athletes[2]; the key finding was that LHTL was significant in athletes, LTH and PHC were significant in non-athletes, and PHC was non-significant in athletes. This stratified pattern itself suggests that the effects of hypoxic training are not an attribute that can be transposed across populations. Dorelli et al.'s strict meta-analysis posed the most challenging question: when hypoxic training is juxtaposed with matched normoxic training, is the between-group difference in VO_2max significant? [3] Its negative conclusion and detection of publication bias suggest that the distribution of positive results in the literature may be systematically biased toward small-sample studies, studies with high analytical flexibility, or studies with selective reporting.

The three studies can logically coexist: IHT may be the most effective choice within hypoxic paradigms for non-athletes[1]; LTH and PHC may produce significant effects in non-athletes when hypoxic intervention is compared with respective control conditions[2]; and the incremental gain of hypoxia relative to intensity-matched normoxic training may be small and current evidence insufficient for robust estimation[3,14]. Applying the statement "hypoxic training is effective" without qualification to healthy college students essentially compresses conclusions from these three different levels into one, thereby losing the necessary sensitivity to protocol parameters and comparative logic.

4. Multidimensional Evidence on Adaptive Mechanisms: the Limited Temporal Sensitivity of VO2max

If hypoxic training indeed produces adaptations in healthy college students, and these adaptations do not always manifest as between-group differences in VO_2max, then a natural question is: in which dimensions do adaptations occur, and whether the pull of these dimensions on VO_2max has temporal precedence. Table 1 classifies major outcome indicators by adaptation dimension and temporal sensitivity to facilitate stratified selection during research design.

Table 1. Adaptation Dimensions and Temporal Sensitivity of Outcome Indicators in Hypoxic Training Research

Adaptation dimension	Representative indicators	Direction of change in hypoxic training (weeks-scale)	Temporal sensitivity	Association with VO_2max
Central cardiac pumping function	Stroke volume, cardiac output, left ventricular ejection time	↑ after 4-week IHT	High (detectable at 4 weeks)	Moderate (a component of VO_2max but not a sufficient condition)
Blood oxygen-carrying and	Red blood cell count, erythrocyte deformability,	Mixed (RBC count ↑, hemoglobin not	Moderate (some indicators change at	Moderate (depending on

 Copyright @ STEMM Institute Press

rheology	hemoglobin	significantly ↑)	4 weeks)	whether threshold is reached)
Vascular endothelial function	FMD, baPWV	↑ (hypoxic group superior to normoxic group)	Relatively high (detectable within weeks)	Indirect (affects peripheral perfusion and afterload)
Autonomic nervous regulation	RMSSD, HF, LF/HF	Mixed (positive and negative results coexist)	Relatively high (detectable within weeks)	Indirect (affects recovery and training tolerance)
Antioxidant defense	SOD, GSH-Px, MDA	↑ (SOD, GSH-Px ↑; MDA unchanged)	Moderate (detectable at 4 weeks)	Indirect (affects redox environment)
Peak aerobic capacity	VO_2max	Mixed (non-significant at 4 weeks, significant at 10 weeks)	Low (requires longer period or larger sample)	— (integrative outcome)

Table note: IHT = intermittent hypoxic training; FMD = flow-mediated dilation; baPWV = brachial–ankle pulse wave velocity; RMSSD = root mean square of successive differences in R-R intervals; HF = high-frequency power; LF/HF = low-frequency/high-frequency ratio; SOD = superoxide dismutase; GSH-Px = glutathione peroxidase; MDA = malondialdehyde.

Table 1 reveals a fact with direct implications for research design: VO_2max is the outcome indicator with the lowest temporal sensitivity in the table, whereas cardiac pumping function, vascular endothelial function, and autonomic nervous regulation can show detectable changes at the 4-week scale. Within a 4–8-week semester-based intervention window, using VO_2max as the sole primary outcome may systematically underestimate effects produced by hypoxic training in dimensions that change earlier. The combination in Liu Xudong's study—significant improvement in cardiac pumping function without significant VO_2max—is precisely a direct example of this temporal mismatch in a Chinese college student sample [5].

4.1 Vascular Endothelial Function and Peripheral Vascular Regulation

In healthy women, the hypoxic group under interval training showed greater improvements than the normoxic interval training group in flow-mediated dilation (FMD, $\eta^2 = 0.428$) and baPWV ($\eta^2 = 0.348$), and hemorheological indicators such as erythrocyte aggregation also changed in ways favorable to perfusion[15]. FMD reflects endothelium-dependent vasodilatory capacity, and baPWV reflects arterial stiffness; both belong to the domain of peripheral vascular function and are different adaptation dimensions from central hematology and peak oxygen transport capacity. A 2026 review in the Journal of Physiology systematically summarized acute and subacute physiological responses to hypoxic exercise in autonomic, pulmonary, cardiovascular, and hematological systems, confirming that hypoxic ventilatory response, sympathetic activation, and hypoxic pulmonary vasoconstriction are core acute mechanisms, and that hypoxic exercise has been explored as a therapeutic intervention in limited contexts such as overweight/obesity, right ventricular remodeling, and aging[9].

The significance of these findings is that hypoxic training may already produce measurable adaptations at the peripheral vascular level at exposure doses that have not yet touched the erythropoietic system or changed peak oxygen transport capacity. If this adaptation can occur within weeks while VO_2max improvement requires synchronous accumulation of central and peripheral adaptations, then college student studies using a 4–8-week window and VO_2max as the sole primary outcome may logically systematically underestimate effects of hypoxic training in the vascular dimension. Outcome selection determines what a study can see: if the research question is set as whether hypoxia changes the dimensional distribution of adaptations, FMD and baPWV should be included in the primary outcome set; if the research question is strictly limited to whether hypoxia amplifies peak aerobic capacity, VO_2max is an appropriate outcome, and current synthesis evidence gives a cautious answer to this [3,14].

4.2 Autonomic Nervous Regulation: Directionality Beyond Sympathetic Activation

Hypoxic exposure causes sympathetic activation in the acute phase, but autonomic adaptations induced by intermittent, moderate-level repeated exposure are not equivalent to the effects of

chronic sustained hypoxia. A study of IHT in healthy women showed that RMSSD and HF (reflecting parasympathetic activity) increased significantly after training in the hypoxic group, and the LF/HF ratio decreased significantly, whereas the LF/HF ratio increased in the normoxic interval training group[15]. This directional difference is noteworthy: hypoxic training did not simply increase sympathetic tone; under appropriate intermittent protocols, it may have promoted recovery or enhancement of parasympathetic regulation. A 2024 narrative review, in distinguishing IHT (intermittent hypoxia during exercise) from IHE (intermittent hypoxic exposure at rest), pointed out that IHT has stronger effects on exercise tolerance, with major adaptations occurring in working skeletal muscle; the smaller effects of IHE may include improvement in autonomic regulation processes, endothelial function, and antioxidant capacity; IHE is particularly suitable for sedentary, elderly, or chronically ill populations, whereas IHT is more suitable for young people who already have a training foundation[16]. For healthy college students—a young but usually not systematically trained group—this distinction has direct design implications: if the goal is improvement in autonomic regulation, IHE or a combination of IHT and exercise may be more worth considering than IHT alone.

The hypoxic effects on autonomic indicators are not consistently positive. Park et al.[15] observed increases in RMSSD and HF and a decrease in LF/HF in healthy women undergoing IHT; however, Burtscher et al.'s narrative review[16] noted that different intermittent hypoxic protocols (including hypobaric vs. normobaric hypoxia, IHE vs. IHT) have inconsistent effects on autonomic indicators, and some intermittent hypobaric hypoxia exposure studies found no significant between-group differences in heart rate, blood pressure, cardiac output, stroke volume, total peripheral resistance, cardiovascular variability, or cardiac-vagal baroreflex function after 4 weeks. This divergence may arise from differences in hypoxic protocols (hypobaric vs. normobaric hypoxia, temporal structure and total amount of exposure) and from differences in measurement timing of autonomic indicators. For healthy college student research, this divergence suggests that the value of HRV or post-exercise heart rate recovery as outcome indicators lies not in providing an always-sensitive effect dimension, but in revealing whether hypoxic training produces regulatory adaptations at the autonomic level that can be transferred to resting and recovery states.

4.3 Heterogeneity of Antioxidant Defense and Hematological Adaptation

Chinese studies of simulated living-high–training-low showed that after 4-week intervention, erythrocyte SOD and GSH-Px activities were significantly increased both at rest and after exhaustive exercise, while MDA did not change, suggesting that enhanced antioxidant defense was sufficient to offset potential increases in lipid peroxidation [13]. The conclusion of the umbrella review is consistent with this direction: gains from intermittent hypoxic protocols are often not accompanied by consistent hematological changes [4], meaning that using increases in hemoglobin mass or red blood cell volume as the core mediating mechanism of hypoxic training lacks general support in healthy non-athlete populations. Kohlbrenner et al. also acknowledged that one mechanism by which LHTL improves VO_2max in athletes may be an increase in hemoglobin mass, but this is not a universal finding [2]. The combination in Liu Xudong's study—non-significant hemoglobin but significant changes in red blood cell count and deformability—further indicates that hematological adaptation itself is a multi-indicator, asynchronous process[5].

These results jointly point to a cautious conclusion: within the hypoxic dose range typically used in healthy college students, the risk of oxidative stress does not constitute a reason to deny the feasibility of hypoxic training, nor can enhanced antioxidant defense or hematological adaptation be invoked as definitive mechanisms by which hypoxic training is superior to normoxic training. The relationships between these adaptations and VO_2max are themselves indirect: enhanced antioxidant defense may improve the redox environment during exercise, and improved erythrocyte deformability may favor microcirculatory perfusion; these are not directly equivalent to improvements in oxygen transport or peak oxygen utilization capacity.

5. How Protocol Parameters Modulate Effects: from Hypoxia to Hypoxic Dose

 Copyright @ STEMM Institute Press

5.1 Saturation Hours: a Variable Closer to "dose" than FiO_2

Hypoxic training is not a single intervention. Inspired oxygen concentrations ranging from 16.5% to 10% correspond to mild to severe hypoxic stress, and FiO_2 alone is insufficient to characterize cumulative exposure. Yu et al.'s meta-regression confirmed a significant association between saturation hours and IHT-induced VO_2max improvement (coefficient 0.004, P = 0.038, 95% CI 0.0002–0.0085), and explicitly recommended saturation hours as an operational indicator for future dose–response research on hypoxic training[1]. Kohlbrenner et al.'s multivariable meta-regression likewise identified the severity, duration, and frequency of hypoxic exposure as significant moderators of VO_2max change: in the PHC model, daily hypoxic duration (B = 2.46) and frequency of hypoxic exposure (B = 0.28) were independently associated with VO_2max change[2]. The concept of saturation hours integrates concentration and time into a cumulative exposure amount, making it more comparable than reporting FiO_2= 15% alone or 60 minutes per session alone. For the short-cycle IHT protocols common in healthy college student research, saturation hours usually fall in the range of tens of hours; living-high–training-low protocols, because they include prolonged nightly hypoxic residence, can reach hundreds of saturation hours. The two cannot be equated in the nature or magnitude of physiological stimulation. A more precise approach is to explicitly calculate saturation hours while reporting FiO_2 and single-exposure duration, enabling dialogue among studies at the dose level.

5.2 Intensity Anchoring: Non-Exchangeability of Absolute and Relative Matching

The methodological detail most easily overlooked in explaining hypoxic training effects is how exercise intensity is matched between hypoxic and normoxic groups. If both groups train at the same absolute power or speed, the relative physiological stress in the hypoxic group will necessarily be higher; between-group differences then reflect the effect of higher relative intensity, not the independent contribution of hypoxia. If both groups are anchored to the same relative intensity (%HRmax, %HRR, or %VO_2max), the absolute power in the hypoxic group will be lower; between-group differences then reflect the effect of achieving equal cardiorespiratory stress at a lower mechanical load—this is closer to the rational application logic of hypoxic training in practice, and is also the reasoning underlying Yu et al.'s discussion of the mechanical-load-sparing value of hypoxia: increasing hypoxic stimulus in a hypoxic environment can maintain cardiorespiratory stress without increasing skeletal muscle mechanical load[1]. A meta-analysis on administration modes of normobaric hypoxia distinguished conditions such as hypoxia–absolute load and hypoxia–relative load in subgroup analyses, indicating that this distinction has been recognized at the synthesis level as a potential effect modifier[12]. Dorelli et al.'s strict meta-analysis did not systematically code intensity anchoring as an independent moderator during study inclusion[3]; the pooled effect size may therefore mix studies under two matching logics. In Chinese college student research, Liu Xudong's protocol used intermittent inhalation with a fixed temporal structure and did not report individualized calibration of exercise intensity[5]; Tu Yanhao et al.'s resistance-mouthpiece protocol likewise lacked details on intensity matching[8]. This reporting deficiency prevents readers from determining whether between-group differences arise from the hypoxic condition itself or from different relative intensities actually experienced by the two groups.

5.3 Temporal Sensitivity of Outcomes and Stratification by Training Status

Table 1 has shown that VO_2max is the outcome indicator with the lowest temporal sensitivity, whereas cardiac pumping function, vascular endothelial function, and autonomic nervous regulation can show detectable changes at the 4-week scale. Within a 4–8-week semester-based intervention window, using VO_2max as the sole primary outcome may systematically underestimate effects of hypoxic training in dimensions that change earlier. From a research design perspective, outcome selection itself is a theoretical decision, not a neutral methodological choice. If the research question focuses on whether peak aerobic capacity is additionally improved by hypoxia, VO_2max is an appropriate outcome, and synthesis evidence from Dorelli et al. and the cardiovascular function meta-analysis gives a cautious negative answer to this. If the research question shifts to whether, under a given mechanical load

Copyright @ STEMM Institute Press

constraint, hypoxia changes the cost or dimensional distribution of adaptation, then cardiac pumping function, vascular endothelial function, autonomic nervous regulation, and submaximal efficiency should be included in the primary outcome set.

Stratification by training status also has substantive consequences. Kohlbrenner et al.'s stratified analysis revealed that the direction and magnitude of hypoxic training effects depend on training status, and that non-athletes are not a homogeneous category[2]. Healthy college students are lower than athletes in training status; their baseline aerobic fitness, heterogeneity of training background, and semester-based time constraints also differentiate them from healthy adults generally defined as non-athletes. Within the college student population, baseline VO_2max may differ significantly between physical education majors and non-physical-education majors, and most studies have not used this variable as a stratification factor during randomization. If the additional gain from hypoxia is greater in subgroups with lower aerobic fitness (because there is more room for adaptation) and approaches zero in high-baseline subgroups, then in an unstratified college student sample, the pooled between-group difference will shrink toward zero, producing an overall negative result that masks the heterogeneous distribution of effects across subgroups. Kohlbrenner et al.'s univariable meta-regression indeed found a significant negative association between baseline VO_2max and VO_2max change with PHC ($B = -0.10$, 95% CI -0.19 to -0.01), providing preliminary synthesis-level evidence for the hypothesis that those with lower baseline may benefit more[2]. Liu Xudong's sample consisted of ordinary college students without specialized sports training experience[5]; Lu Baoyun's sample consisted of non-physical-education female college students[6]; Hou Zheng's study did not report the training background of the sample in detail[7]—these differences in training status may partly explain the inconsistency among their conclusions. Future college student studies should report the distribution of baseline VO_2max at recruitment and include it as a stratification factor or covariate in analyses.

6. Accessible Boundaries of Safety Monitoring in University Settings

For healthy college students, safety concerns regarding simulated hypoxic training are fundamentally different from those for clinical populations. Under conditions where FiO_2 is not lower than 10%–12%, exposure time is controlled, and participants are at rest or performing moderate-intensity exercise, the reported rate of severe hypoxic exposure risk is very low. Liu Xudong's 4-week protocol with FiO_2 progressively decreased to 10% reported no adverse events[5]; simulated living-high–training-low studies at 14% oxygen concentration and 8 hours of sleep per day likewise reported no safety problems[13]. The support strength of these signals is limited: the above studies had small samples (mostly 10–30 participants), safety reporting was usually not systematic, and there was no complete account of SpO_2 monitoring methods, termination criteria, or incidence of adverse events.

Monitoring of peripheral oxygen saturation (SpO_2) has practical operational value in university settings. Real-time SpO_2 monitoring has been used to assess the safety of normobaric hypoxic training, and the development of wearable pulse oximeters has made it possible to continuously obtain heart rate and SpO_2 during training[17]. Some studies have continuously monitored SpO_2 and heart rate during hypoxic exercise to identify abnormal oxygen desaturation or excessive cardiovascular stress, thereby ensuring participant safety [17]. Regarding the setting of SpO_2 thresholds, relevant literature on real-time monitoring and hypoxic training management has discussed using the magnitude of SpO_2 decline or a preset lower limit as a reference for adjusting or stopping training[17], but no systematic dose–safety evidence has been established for healthy college students, and its applicability across simulation modalities, FiO_2, exposure temporal structures, and individuals is limited; this review does not treat it as a universal threshold or unified recommendation. For college student research, a more defensible approach is to preset an individualized SpO_2 lower limit confirmed by medical evaluation (e.g., allowing a certain decline from baseline resting SpO_2, or using a preset range as a reference), and simultaneously specify medical screening, termination criteria, emergency plans, and adverse event reporting procedures, implemented after ethics approval; during training, record SpO_2 monitoring frequency, nadir SpO_2 during exercise, exercise intensity at which the lower limit is reached,

 Copyright @ STEMM Institute Press

symptoms, and recovery time, rather than treating a single value as a cross-study universal safety standard. It should be emphasized that even when FiO_2 is not lower than 10%–12%, exposure time is controlled, and participants are at rest or performing moderate-intensity exercise, the absence of reported adverse events does not equal the absence of risk; protocols that decrease FiO_2 to 10% should be regarded as high-exposure protocols requiring strict monitoring and ethical review, rather than risk-free protocols.

Existing Chinese college student studies are generally weak in reporting safety indicators. Most studies did not systematically report SpO_2 monitoring frequency, nadir SpO_2 during exercise, termination criteria, or occurrence of adverse events. For a paper targeting healthy college students, systematic reporting of safety data is not only an ethical requirement but also a component of methodological quality: it enables readers to judge whether the protocol can be reproduced in similar university settings. A narrative review on intermittent hypoxia and exercise tolerance, when recommending IHE for populations with low exercise tolerance and IHT for young people with a training foundation, also implicitly binds safety boundaries to population characteristics: the applicability of IHT to healthy young people depends on exposure dose being controlled within a reproducible and tolerable range, rather than pursuing the lowest possible hypoxic stimulus[16].

7. Discussion: from Overall Effect Testing to Cumulative Conditional-Effect Propositions

By juxtaposing Chinese college student studies with English-language synthesis evidence, the sources of divergence can be summarized into five domains, which together constitute the analytical framework shown in Figure 1. These domains do not point to a single answer to the overall question of whether hypoxic training is effective; rather, they point to a series of cumulative conditional-effect propositions that need to be answered separately.

First, the incremental effect of hypoxia relative to matched normoxic training is not yet robust at the current level of synthesis evidence, and the scope of this judgment needs to be precisely defined. Dorelli et al.'s strict meta-analysis[3], the detection of publication bias ($p = 0.004$), and the meta-analysis of cardiovascular function in populations without cardiovascular disease[14] jointly point to a cautious conclusion: in VO_2max, peak performance, and conventional cardiovascular function parameters, the additional gain of hypoxic training has not been robustly demonstrated. This does not mean that hypoxic training is equivalent to normoxic training across all outcomes and all protocols. A more precise proposition is: under a given number of saturation hours and intensity anchoring method, for college students with baseline aerobic fitness in a specific range, which outcome dimension will show reproducible between-group differences over what period. This proposition format requires studies to simultaneously report dose parameters, intensity matching logic, baseline characteristics, and outcomes at multiple time points, rather than only reporting endpoint comparisons between hypoxic and normoxic groups.

Second, standardization of simulation modalities is a prerequisite for cross-study accumulation. Pérez-Regalado et al.'s meta-analysis[12] has provided direct quantitative evidence: there is a substantive gap between the effect size of chamber/tent-type devices (SMD = 1.30) and the non-significant effect of mask-type devices, and this gap is further amplified in sedentary populations and under conditions of high session volume. In Chinese college student literature, hypoxic apparatus, resistance mouthpieces, and ETM masks differ markedly in FiO_2 controllability, respiratory work load, and dead-space effects, and reporting completeness is far from sufficient to support pooling across simulation modalities. If simulation modality itself systematically modulates effective dose and effect direction, then pooling data from different simulation modalities into a single effect estimate of hypoxic training is methodologically inadequate. Calculation of saturation hours, calibration and reporting of FiO_2, complete description of intermittent structure, and specification of respiratory resistance and dead-space parameters are necessary conditions for transforming hypoxic training from a broad label into a comparable intervention.

Third, outcome selection determines what a study can see; designs using VO_2max as the sole primary outcome narrow the evidence base. Table 1 shows that adaptation signals in cardiac pumping function, vascular endothelial function (FMD, baPWV), autonomic nervous regulation, and hemorheology can appear at the 4–10-week

Copyright @ STEMM Institute Press

scale, and their direction and magnitude may differ from changes in VO_2max[5,15]. Dorelli et al.'s negative conclusion is mainly based on VO_2max and exercise performance indicators; if the unique value of hypoxia lies in changing the dimensional distribution of adaptation rather than amplifying peak capacity, then synthesizing solely on VO_2max is logically insufficient to negate this possibility. For college student populations aiming at health promotion rather than competitive performance, including vascular function, autonomic nervous regulation, and submaximal efficiency in a stratified outcome set may be practically more relevant than a single VO_2max comparison.

Fourth, the internal heterogeneity of the healthy college student population needs to be addressed directly in design. Differences between physical education majors and non-physical-education majors, and between high-baseline and low-baseline VO_2max, may systematically modulate the direction and magnitude of hypoxic training effects. The negative association between baseline VO_2max and PHC effects found by Kohlbrenner et al. ($B = -0.10$)[2], together with the dose effect of saturation hours confirmed by Yu et al. [1], jointly suggest that "for whom" and "how much" are more explanatory variables than "whether to use hypoxia." Pre-reporting the distribution of baseline VO_2max, incorporating training status as a stratification factor or covariate in analyses, and considering the minimum detectable effect for subgroup comparisons in sample size estimation are basic conditions for enabling college student studies to produce cumulative evidence.

Fifth, confusion over the level of comparison object is the main producer of apparent contradictions in the current literature. The foregoing comparison of three synthesis evidence studies has shown that Yu et al., Kohlbrenner et al., and Dorelli et al. answer questions at different levels. Interpreting IHT's highest ranking within hypoxic paradigms as IHT being superior to normoxic training, or interpreting the absence of a significant advantage of hypoxia relative to matched normoxia as hypoxia being ineffective for all outcomes, both compress conclusions from different levels into the same level. A 2024 narrative review, in distinguishing the applicable populations of IHE and IHT, gave a stepwise recommendation: IHE is suitable for populations with low exercise tolerance (sedentary, elderly, chronic disease), followed by exercise training under normoxia, and finally transition to IHT; IHT is more suitable for young people who already have a certain training foundation[16]. The implication of this stepwise logic for healthy college student research is that "healthy college students" itself is not a homogeneous category that can be directly matched to the IHT-applicable population label; protocol selection should be based on baseline characteristics, not on the general labels of young or healthy.

8. Conclusion

The effects of simulated hypoxic training on cardiorespiratory fitness in healthy college students cannot be summarized by the dichotomy of effective or ineffective. Chinese college-student studies have provided preliminary signals in cardiac pumping function, hemorheology, antioxidant defense, and some aerobic performance indicators, but these signals have not been replicated in a sufficient number of independent samples, and their incremental magnitude relative to matched normoxic training remains unclear. International synthesis evidence has answered different questions at different levels: which hypoxic paradigm ranks highest within hypoxic training, which hypoxic strategies show significant effects relative to their respective controls, and whether hypoxia has an additional gain relative to matched normoxia. The cautious conclusion of strict meta-analysis on the last question should be placed on an equal footing with network meta-analysis and meta-regression, rather than selectively cited or avoided.

The next step in this field is not simply to accumulate more overall tests of hypoxia versus normoxia, but to refine research questions into conditional-effect combinations across five domains: level of comparison, hypoxic dose, intensity anchoring, simulation modality, and outcome temporal window. Calculation and reporting of saturation hours, standardized reporting of FiO_2 and simulation modalities, explicit explanation of intensity-matching logic, stratified design of multidimensional outcomes, and systematic presentation of safety data are operational conditions for elevating these combinations from individual studies to a cumulative knowledge system. For researchers working with healthy college students in physical education and training science, this implies a methodological shift: from asking

 Copyright @ STEMM Institute Press

whether hypoxia is better than normoxia to designing studies that can answer under what comparison level, at what hypoxic dose, under which intensity-anchoring logic, with which simulation modality, and over what outcome time window hypoxia relative to matched normoxic training produces reproducible and practically meaningful differences.

References

[1] Yu Q, Kong Z, Zou L, Chapman R, Shi Q, Nie J. Comparative efficacy of various hypoxic training paradigms on maximal oxygen consumption: A systematic review and network meta-analysis. J Exerc Sci Fit. 2023;21(4):366–375.

[2] Kohlbrenner D, Ghaith A, Fournier J, et al. The Effects of Intermittent Hypoxic Training Strategies on Maximal Oxygen Uptake in Healthy Humans: A Meta-analysis with Meta-regression. Sports Med. 2026;56:2767–2782.

[3] Dorelli G, Giuriato G, Zamboni G, Daini M, Cominacini M, Carbonare LGD, Wang E, Crisafulli E, Schena F, Venturelli M. Aerobic Intermittent Hypoxic Training Is Not Beneficial for Maximal Oxygen Uptake and Performance: A Systematic Review and Meta-Analysis. Scand J Med Sci Sports. 2025;35(6):e70088.

[4] Boulares A, Dupuy O, Bragazzi LN, et al. Effects of Intermittent Hypoxia Protocols on Physical Performance in Trained and Untrained Individuals: An Umbrella Review of Systematic Reviews and Meta-Analyses. Sports Med Open. 2025;11(1):145.

[5] Liu X D. Effects of intermittent hypoxic training on cardiac function in male college students [C]//Committee of Exercise Physiology, Chinese Physiological Society. Proceedings of the 2013 Annual Meeting of the Committee of Exercise Physiology, Chinese Physiological Society and the "Exercise and Health" Academic Symposium. 2013:129–130.

[6] Lu B Y. Effects of hypoxic training on health-related physical fitness in female college students [J]. Journal of Lanzhou University of Arts and Science (Natural Sciences), 2022, 36(1):105–110.

[7] Hou Z, Wang J J, Yin J. Effects of ETM hypoxic endurance training under different load durations on physical fitness in college students [J]. Journal of Anhui Normal University (Natural Science), 2019, 42(2):195–199+204.

[8] Tu Y H, Tan Z Q, She S J N W. Effects of simulated hypoxic training on aerobic exercise capacity and physiological function in college students [J]. Sichuan Sports Science, 2016, 35(1):39–43.

[9] Forbes LM, Hearon CM Jr, Cillay J, Lawley JS, Stembridge M, Subudhi AW, Levine BD, Cornwell WK III. Exercise in acute and subacute hypoxic conditions: physiological responses affecting oxygen delivery and their applications to human health. J Physiol. 2026.

[10] Zhao S Q, Shi H J. Effects of hypoxic training on body composition and aerobic endurance in obese college students [J]. Chinese Journal of School Health, 2016, 37(11):1637–1640.

[11] Liu R, Wen Z, Gou P, et al. Effects of exercise interventions in hypoxic environment on cardiovascular function and maximal power output in obese or overweight individuals: a systematic review and meta-analysis. BMC Cardiovasc Disord. 2025;26(1):203.

[12] Pérez-Regalado S, Benavente C, Almeida F, Feriche B. Impact of the administration type for normobaric hypoxia on the maximal aerobic capacity: A systematic review and meta-analysis. Sports Med Health Sci. 2026;8(2):153–162.

[13] Yun D C, Hu Y X. Effects of simulated living-high–training-low on erythrocyte antioxidant capacity and aerobic exercise capacity in male college students [J]. Journal of Physical Education, 2005(4):53–55.

[14] Rodríguez FR, Vizcaíno MV, Gutiérrez RS, et al. Effectiveness of hypoxic versus normoxic exercise on cardiovascular function in people without cardiovascular diseases: A systematic review and meta-analysis. Eur J Clin Invest. 2025;55(7):e70028.

[15] Park HY, Jung WS, Kim SW, Lim K. Effects of Interval Training Under Hypoxia on the Autonomic Nervous System and Arterial and Hemorheological Function in Healthy Women. Int J Womens Health. 2022;14:79–90.

[16] Burtscher J, Glazachev SO, Kopp M, et al. Effects of interval hypoxic exposures and interval hypoxic training on exercise tolerance (narrative review) [J]. Sports

Medicine: Science and Practice, 2024, 14(2):16–23.

[17] Iatsune K, Yamakawa T. Real-Time physiological monitoring for management of normobaric hypoxic training toward wearable system implementation. Journal of Digital Life. 2022;2:2022.2.2.

 Copyright @ STEMM Institute Press