



Development and Validation of an Integrative Intervention Model for Physical Fitness, Metacognition, and Academic Resilience in Senior High School Students

Romiya Heliza^{1*}, Febriady Huta Uruk², Efriana Jon³, Nandia Pitri⁴

¹Department of Physical Education, STKIP Muhammadiyah Sungai Penuh, Indonesia

²Department of Guidance and Counseling, STKIP Muhammadiyah Sungai Penuh, Indonesia

³Department of Mathematics Education, STKIP Muhammadiyah Sungai Penuh, Indonesia

⁴Department of Indonesian Language Education, STKIP Muhammadiyah Sungai Penuh, Indonesia

*Corresponding Author: Romiya Heliza, e-mail: romiyaheliza19@gmail.com

Received: 19 July 2026, Approved: 20 August 2026, Published: 30 September 2026

Abstract

Study purpose: This study aims to develop a prototype of the "Moving to Learn" Intervention Model that integrates structured physical activity, metacognitive skills training, and resilience building techniques in a practical format for adoption by senior high school teachers.

Materials and methods: This study employed a Research and Development (R&D) approach using the 4D procedural model (Define, Design, Develop, Disseminate) by Thiagarajan, Semmel, and Semmel, limited to the Define, Design, and Develop phases. Participants were involved at different stages according to their roles: (1) Needs analysis stage 5 teachers (Physical Education, Guidance and Counseling, and Homeroom Teachers) and 30 grade XI students completed needs assessment questionnaires, with 3 teachers participating in semi-structured interviews; (2) Expert validation stage 3 experts (Physical Education, Educational Psychology, and Learning Technology) validated the prototype; (3) Practicality assessment stage 2 teachers conducted expert reviews and 5 students participated in one-to-one tryout simulations; and (4) Limited trial stage 30 students participated in the pilot implementation without a control group. Quantitative data were analyzed using descriptive statistics (percentages and means), while qualitative data were analyzed using the Miles and Huberman model.

Results: Expert validation confirmed the prototype's validity with an average score of 4.2 out of 5, and practicality tests yielded an implementation rate of 85%. In the limited pilot trial without a control group, descriptive improvements were observed across all indicators: physical fitness (VO2Max sufficient category) increased from 35% to 68% (+33%), high metacognitive ability increased from 42% to 73% (+31%), high academic resilience increased from 48% to 79% (+31%), and active interaction in learning increased from 55% to

85% (+30%). These preliminary findings indicate potential promise of the prototype, but they are descriptive in nature and should not be interpreted as statistically significant effects, as no inferential statistical tests were conducted.

Conclusion: This developmental study successfully produced a valid and practical prototype of the "Moving to Learn" Model, comprising a teacher's guidebook, student activity modules, and initial assessment instruments. The prototype demonstrates preliminary feasibility for use in senior high school settings, as evidenced by expert validation and practicality assessments. However, because the model was tested only in a limited pilot without a control group, claims of effectiveness cannot be made. The prototype is now ready for further rigorous testing through controlled experimental studies to establish its effectiveness on physical fitness, metacognition, and academic resilience.

Keywords: Model Development, Integrated Intervention, Physical Fitness, Metacognitive, Student Resilience.

DOI: <https://doi.org/10.52188/ijpess.v6i3.2352>

©2026 Authors by Universitas Nahdlatul Ulama Cirebon



Introduction

Education in the 21st century emphasizes the development of holistic competencies, covering cognitive, physical, and social-emotional aspects in a balanced manner (Kemdikbud, 2017). However, the reality in many senior high schools shows a gap between this ideal expectation and field practice (Fullan & Langworthy, 2014). Excessive focus on cognitive academic achievement often neglects the development of physical fitness and life skills such as self-regulation and mental resilience (Duckworth & Yeager, 2015).

An initial survey in two senior high schools in Kerinci Regency showed that more than 60% of students spent less than 30 minutes per day on moderate-to-vigorous physical activity, 55% of students had difficulty planning and evaluating their own learning strategies, and 50% showed symptoms of mild to moderate academic anxiety. These findings align with global data indicating that approximately three-quarters of adolescents are not sufficiently physically active (Guthold et al., 2020), and the COVID-19 pandemic has exacerbated this condition with increased sedentary behavior and decreased physical activity among children and adolescents (Stockwell & al., 2021). (Yi & al., 2024) also reported an alarming decline in physical activity levels among children worldwide.

This problem is compounded by a fragmented learning approach. Physical education operates independently, academic learning focuses on content, and guidance and counseling are often curative in nature (Bailey et al., 2009). In fact, recent research shows a strong reciprocal relationship between physical activity and cognitive function and psychological well-being (Heinze et al., 2021). Physical activity has been proven to improve physical, mental, cognitive, and brain health across the lifespan, with evidence showing improved executive control, academic performance, and adaptation of brain structure and function (Hillman & al., 2020).

Metacognition, defined as the ability to understand and control what is perceived or "thinking about thinking," is critical for educational success and mental resilience (Taniguchi et al., 2024). (Bao et al., 2024) in their meta-analysis revealed that metacognition has a significant impact on academic achievement in physical education and physical activity settings. Research shows that metacognition positively correlates with perceived exercise benefits, and interventions should focus on strengthening cognitive self-confidence through methods such as resilience training, cognitive restructuring, and mindfulness practices (Taniguchi et al., 2024). Pandey (2025) found that a holistic approach integrating yoga with

ancient learning principles produced the most significant improvements in comprehensive academic skills, creative writing, and reading in adolescents.

WHO (2020) recommends that children and adolescents aged 5-17 years should engage in moderate-to-vigorous physical activity for an average of 60 minutes per day throughout the week, with vigorous aerobic activity and muscle and bone strengthening at least 3 days per week. Physical activity provides benefits for physical fitness (cardiorespiratory and muscular), cardiometabolic health, bone health, cognitive outcomes (academic performance, executive function), mental health (reducing depressive symptoms), and reducing adiposity (WHO, 2020).

Research by Reigal et al (2020) showed that adolescents who engaged in more weekly physical exercise and had higher levels of physical fitness obtained better scores on tests of cognitive and psychosocial functioning. These findings suggest that engaging in physical exercise and fitness during adolescence may be appropriate for improving health and well-being. Furthermore, research on long-term physical activity interventions reported improvements in mental health and cognition outcomes that coincided with neurobiological changes, particularly changes in neuroimaging activation and electrophysiological parameters in frontal areas (Heinze et al., 2021). Sibbick et al. (2024) also found that physical activity and cardiorespiratory fitness were positively associated with cognitive function, self-control, and resilience in adolescents with attention deficit hyperactivity disorder.

In the context of physical education and resilience development, Purnomo et al. (2026) emphasized that sports can function as a means of mental education to improve student resilience. Rahayu & al. (2024) also reported that the integration of mental education through physical education was effective in improving adolescent personal resilience. Pruzhinin et al. (2026) developed a stress resilience development program for high school students through physical culture and sports, which showed positive results in improving students' mental resilience.

Despite the growing body of evidence supporting integrated physical activity interventions, most existing programs focus on a single domain-either physical fitness, cognitive training, or psychological support-rather than combining all three in a structured, practical format that can be easily implemented by classroom teachers. A systematic comparison of representative existing intervention models reveals several specific limitations that justify the development of a new integrated approach. Interventions such as "ACTIVE SCHOOL" (Bugge & al., 2022) primarily emphasize embodied cognition through movement linked to academic tasks, yet they lack explicit metacognitive skills training and psychological resilience components, and they require significant curriculum restructuring that may not be feasible in all school contexts. Metacognitive training programs (Bao et al., 2024) effectively improve cognitive outcomes but do not address physical fitness or psychological resilience, and they require trained psychologists for delivery, limiting their scalability in typical school settings. Resilience programs (Pruzhinin et al., 2026) and mental education through physical education approaches (Purnomo et al., 2026; Rahayu & al., 2024) address physical activity and psychological resilience but lack explicit metacognitive components and rely on physical education specialists, making them less applicable across all subject areas. Yoga-based holistic programs (Pandey, 2025) integrate physical, cognitive, and psychological domains but require certified yoga instructors, which limits their scalability and sustainability in resource-constrained settings. Resistance training interventions (Robinson et al., 2023) focus primarily on physical fitness and academic outcomes but require specialized equipment and facilities, and they do not systematically integrate cognitive or psychological components. This fragmented approach across existing interventions fails to leverage the reciprocal relationships between physical activity, cognitive function, and psychological well-being documented in recent research (Heinze et al., 2021; Hillman & al., 2020).

Furthermore, most existing interventions require specialized facilitators-such as certified yoga instructors (Pandey, 2025), physical education specialists (Purnomo et al., 2026), or trained psychologists (Bao et al., 2024) which limits their applicability in typical senior high school settings where classroom teachers deliver the majority of instruction. The practical burden on teachers is compounded by the need for extensive external facilitation or specialized equipment, as seen in resistance training interventions (Robinson et al., 2023). Additionally, existing models lack a systematic sequential integration that explicitly connects physical activity to metacognitive reflection and resilience building within a single coherent session flow. Most interventions either deliver components concurrently or sequentially but without explicit pedagogical bridging between domains. For example, mental education through physical education (Rahayu & al., 2024) delivers physical activity and mental education in separate segments without structured integration, while ACTIVE SCHOOL (Bugge & al., 2022) delivers physical activity concurrently with academic tasks without explicit sequential bridging to metacognitive reflection. The theoretical grounding of existing interventions is also often limited to a single framework-such as embodied cognition alone or stress-coping theory alone-whereas an integrated intervention addressing bio-psycho-physical domains requires a multi-framework approach that accounts for learner needs, teacher technological knowledge, and contextual factors. Notably, the NLW (Need-Lack-Want) framework (Hutchinson & Waters, 1987) and TPACK (Mishra & Koehler, 2006) have not been systematically applied in previous integrated physical activity interventions, representing a significant gap in the theoretical foundation of existing models.

These identified gaps justify the development of the proposed "Moving to Learn" Intervention Model, which differs from existing approaches in three distinctive ways. First, the model operationalizes integration through three micro-modules-MOVE (structured physical activity), THINK (metacognitive skills training), and RISE (psychological resilience techniques)-which are designed to be delivered sequentially within a single 45-minute session. This sequential-integrated approach (MOVE → THINK → RISE) creates explicit pedagogical bridges between physical activity, cognitive reflection, and psychological skill-building, addressing the fragmentation problem observed in existing interventions. The MOVE module activates physiological and neural pathways that prepare the brain for cognitive engagement, consistent with embodied cognition theory (Bugge & al., 2022). The THINK module then leverages this activated state to develop metacognitive awareness and self-regulation skills, as recommended by (Taniguchi et al., 2024). The RISE module consolidates learning through psychological resilience techniques such as positive self-talk and breathing exercises, building on the findings of (Pruzhinin et al., 2026) and (Pandey, 2025). This structured flow ensures that each component builds upon the previous one, creating synergistic effects that are absent in concurrently delivered or loosely connected interventions. In contrast to ACTIVE SCHOOL (Bugge & al., 2022), which delivers physical activity concurrently with academic tasks without explicit sequential bridging to metacognitive reflection, and mental education approaches (Rahayu & al., 2024), which deliver physical activity and psychological content in separate sessions without integration, the proposed model provides a coherent sequential flow that maximizes the synergistic potential of the three components.

Second, the model is explicitly grounded in the NLW framework (Hutchinson & Waters, 1987) and the TPACK framework (Mishra & Koehler, 2006), ensuring that the intervention is not only theoretically sound but also contextually responsive to the specific needs of teachers and students in Indonesian senior high schools. The NLW framework guided the needs analysis stage, ensuring that the model addresses what students need (Need), the gaps in current learning (Lack), and the features they desire (Want). The TPACK framework informed the design of the teacher's guidebook, ensuring that content knowledge (physical fitness, metacognition, resilience), pedagogical knowledge (active learning strategies,

reflection techniques), and technological knowledge (assessment instruments, activity modules) are integrated in a coherent manner. This multi-framework theoretical grounding distinguishes the proposed model from single-framework interventions such as resilience programs (Pruzhinin et al., 2026) that rely solely on stress-coping theory or yoga programs (Pandey, 2025) that draw primarily on ancient learning principles without explicit attention to teacher pedagogical and technological knowledge.

Third, the model is packaged as a practical teacher's guidebook with ready-to-use session plans, student activity modules, and assessment instruments, which reduces the implementation burden on teachers who may lack specialized training in physical education or psychology. Unlike existing interventions that require specialist facilitators or extensive external support-such as certified yoga instructors (Pandey, 2025), physical education specialists (Purnomo et al., 2026), or trained psychologists (Bao et al., 2024) the "Moving to Learn" Model is designed for implementation by regular classroom teachers with minimal additional training. The guidebook provides step-by-step facilitation guidelines, clear time allocation for each micro-module, and observation checklists for monitoring student progress. This practical orientation, combined with the model's time-efficiency (45-minute sessions), makes it scalable and sustainable in resource-constrained senior high school settings. In contrast, resistance training interventions (Robinson et al., 2023) require specialized equipment and facilities, while ACTIVE SCHOOL (Bugge & al., 2022) requires significant curriculum restructuring that may not be feasible in all school contexts. Furthermore, the present study extends previous research by employing a rigorous Research and Development (R&D) approach using the 4D procedural model (Thiagarajan et al., 1974), which ensures systematic development and validation of the prototype. Previous integrated interventions often lack documented development procedures and validation evidence, making replication and adaptation difficult. The present study addresses this gap by documenting each development stage (Define, Design, Develop, Disseminate), establishing content validity through expert validation, and assessing practicality through user testing all of which are essential steps before proceeding to effectiveness testing.

Based on this background and the identified gaps in existing interventions, the research problems are: (1) What are the needs of teachers and students for an integrated intervention model that addresses physical, cognitive, and psychological aspects? (2) What is the prototype of the "Moving to Learn" Intervention Model that is valid and practical according to expert assessment and can be disseminated??

Materials and methods

Study participants

The subjects of this study consisted of 5 teachers (Physical Education, Guidance and Counseling, and Homeroom Teachers) and 30 grade XI students from one senior high school in Kerinci Regency, selected using purposive sampling technique. The school was selected because it represents typical senior high school characteristics in Kerinci Regency in terms of student demographics, socioeconomic background, and educational resources, and because the school administration provided access and support for the research. The initial needs assessment conducted at this school confirmed a clear need for the intervention, with over 60% of students showing insufficient physical activity, 55% reporting metacognitive difficulties, and 50% exhibiting academic anxiety symptoms.

The teachers were purposively selected to represent key stakeholder perspectives: Physical Education teachers (physical activity expertise), Guidance and Counseling teachers (psychological well-being and resilience), and Homeroom Teachers (academic and metacognitive challenges). This composition ensured comprehensive input during needs analysis and practicality assessment. The 30 grade XI students were selected based on the

following criteria: active student status, willingness to participate with informed consent, and presence during data collection. Grade XI students were chosen as they represent a critical transitional stage facing increased academic demands, making them an appropriate target population for the intervention model.

The use of purposive sampling from a single school limits the transferability of the findings. Therefore, the results should be interpreted as preliminary evidence of the prototype's feasibility rather than as generalizable conclusions, and future studies should include more diverse samples from multiple schools to enhance external validity.

Study organization

The present study adopted a Research and Development (R&D) approach, operationalized through a partial implementation of the 4D instructional model-comprising Define, Design, Develop, and Disseminate-as originally proposed by (Thiagarajan et al., 1974). This particular model was selected due to its capacity to offer a structured and systematic pathway for producing educational products that meet the criteria of validity, practicality, and effectiveness, making it particularly suitable for the development of an integrated intervention model.

It is important to clarify that the current study represents a partial implementation of the 4D model, specifically limited to the Define, Design, and Develop phases. The Disseminate phase was deliberately excluded from the scope of this study for the following reasons. First, the Disseminate phase requires broader field testing with a larger and more diverse sample across multiple schools, which is logistically and methodologically beyond the scope of a single developmental study. Second, the Disseminate phase includes wider product distribution to various educational stakeholders, which should only be conducted after the prototype has been fully validated and its effectiveness has been established through controlled experimental studies. Third, the primary objective of this study was to develop a valid and practical prototype that could serve as a foundation for subsequent effectiveness testing, not to disseminate the final product to the broader educational community. Fourth, the current funding and time constraints of this research project did not permit the comprehensive implementation of the Disseminate phase, which requires extended timelines and additional resources for large-scale trials and stakeholder engagement.

Consequently, the dissemination activities described later in this manuscript (such as scientific publication and implementation in learning environments) represent planned dissemination strategies for future research, rather than completed activities or evaluation results. The subsequent phase of this research agenda will involve rigorous effectiveness testing using a randomized controlled trial or quasi-experimental design with a larger sample and control group, followed by systematic dissemination of the validated product to schools, teachers, and other educational stakeholders. This staged approach ensures that the model is only disseminated after its effectiveness has been empirically established, thereby upholding the scientific integrity of the development process.

The following sections describe the implementation of each of the three phases conducted in this study.

Define Stage: This stage aims to establish and define learning needs. Activities included: (1) Literature study on integrated interventions, adolescent physical fitness, metacognition, and academic resilience; (2) Distribution of needs assessment questionnaires to 5 teachers and 30 grade XI students; (3) Semi-structured interviews with 3 teachers to explore challenges and opportunities. Needs analysis in this study used the Need, Lack, Want (NLW) framework introduced by Hutchinson & Waters (1987). This framework explains that the development of teaching materials must begin with identifying what students need (need), the

gaps they experience in current learning (lack), and the features or forms of ideal teaching materials they want (want).

Design Stage: At this stage, the developer designed the initial product based on the analysis results. Activities included: (1) Developing the model framework (theoretical foundation, principles, session flow); (2) Designing 6 sample intervention session plans (@ 45 minutes); (3) Developing student activity modules and teacher facilitation guides; (4) Developing initial assessment instruments (observation checklists, student response questionnaires). The intervention model was designed by integrating three micro-modules: MOVE Module (light-moderate intensity physical activity such as dynamic warm-up, bodyweight circuit, reactive games), THINK Module (brief metacognitive skills training such as goal setting and evaluation linked to physical activity), and RISE Module (simple psychoeducational techniques such as positive self-talk and breathing exercises for stress management). This approach aligns with the concept of "embodied cognition" which emphasizes that movement and execution of activities lead to the construction of strengthened mental representations, such as memory enhancement (Bugge & al., 2022).

Develop Stage: This stage is the realization and validation of the design. The prototype was validated by 3 experts (Physical Education Expert, Educational Psychology Expert, Learning Technology Expert) using product validation instruments. Revisions were made based on expert input. Subsequently, a Limited Practicality Test was conducted through expert review by 2 Physical Education/Guidance and Counseling teachers and one-to-one tryout simulation with 5 students to obtain feedback on ease of use and implementation.

Validation Instrument Development: The validation instruments used in this study were developed through the following systematic procedures to ensure their content validity and reliability. First, the validation criteria were derived from a synthesis of relevant theoretical frameworks and previous instrument development studies, including the principles of product feasibility (validity, practicality, and effectiveness) as proposed by (Thiagarajan et al., 1974), the TPACK framework (Mishra & Koehler, 2006), and guidelines for educational product validation (Akker, 1999; Nieveen, 1999). These frameworks provided the conceptual foundation for the instrument's dimensions and items.

Instrument Dimensions and Content Validity: The instrument assessed nine specific dimensions of the prototype: (1) Theoretical & Conceptual Feasibility, (2) Physical Content Feasibility, (3) Cognitive Content Feasibility, (4) Psychological Content Feasibility, (5) Integration Feasibility, (6) Learning Design Feasibility, (7) Media & Presentation Feasibility, (8) Assessment Feasibility, and (9) Implementation Feasibility. Each dimension consisted of 3-5 indicator items, resulting in a total of 36 items across all dimensions. The scoring system used a 5-point Likert scale (1 = very poor, 2 = poor, 3 = adequate, 4 = good, 5 = very good), with product feasibility categories defined as: 0-59% (low), 60-79% (medium), and 80-100% (high).

To establish content validity, the instrument was reviewed by a panel of three experts: one expert in Physical Education (specializing in physical fitness and activity-based interventions), one expert in Educational Psychology (specializing in metacognition and academic resilience), and one expert in Learning Technology (specializing in instructional design and media development). The selection of three experts is consistent with established recommendations in educational research, which suggest that a minimum of three experts is sufficient for content validity assessment, particularly when experts are selected from complementary disciplinary backgrounds (Lynn, 1986; Polit, D. F., & Beck, 2006; Rubio et al., 2003). Each expert was provided with: (1) the validation instrument with clear instructions, (2) the prototype of the "Moving to Learn" Intervention Model (teacher's guidebook, student activity modules, and assessment instruments), and (3) a theoretical framework document explaining the model's conceptual foundations. Experts assessed each item for its relevance to the corresponding dimension and the representativeness of the overall instrument.

Content validity was assessed using the Item-Level Content Validity Index (I-CVI), calculated as the proportion of experts rating an item as relevant (score of 4 or 5 on the relevance scale). Following the criteria established by (Polit, D. F., & Beck, 2006) and (Lynn, 1986), items with $I-CVI \geq 0.78$ for a panel of three experts were retained, while items with lower I-CVI values were revised or removed based on expert feedback. The Scale-Level Content Validity Index (S-CVI/Ave) was calculated as the average of all I-CVI values across all items. An S-CVI/Ave value of ≥ 0.90 was considered evidence of excellent content validity (Polit & Beck, 2006). Additionally, experts provided qualitative feedback on item clarity, comprehensiveness, and appropriateness, which was used to refine the instrument.

Expert Panel Justification: The use of three experts in this study is justified for several reasons. First, methodological guidelines for content validity assessment recommend a minimum panel size of three to five experts, with three being sufficient when experts are carefully selected from distinct relevant fields (Lynn, 1986; Polit, D. F., & Beck, 2006). Second, the calculation of I-CVI requires a minimum of three experts to establish meaningful agreement statistics, as I-CVI values are based on the proportion of experts agreeing on item relevance (Rubio et al., 2003). Third, the three experts in this study were selected from complementary disciplinary backgrounds to ensure comprehensive evaluation across the bio-psycho-physical domains addressed by the model. Fourth, each expert possessed a minimum of five years of experience in their respective field and had previously participated in educational product validation studies, ensuring their competence in assessing the instrument and the prototype.

Reliability Assessment: Inter-rater reliability among the three experts was assessed using Fleiss' Kappa (κ), a statistical measure of inter-rater agreement suitable for multiple raters using categorical or ordinal ratings (Fleiss, 1971). The analysis was conducted on the ratings across all 36 instrument items. For the calculation of Fleiss' Kappa, the 5-point Likert scale ratings were converted into three categories for the purpose of agreement analysis: "acceptable" (scores of 4 or 5), "uncertain" (score of 3), and "unacceptable" (scores of 1 or 2). This categorization is consistent with established practice in inter-rater reliability studies with ordinal data (Landis, J.R. and Koch, 1977).

The inter-rater reliability analysis yielded a Fleiss' Kappa coefficient of $\kappa = 0.72$, indicating substantial agreement among the three experts according to the criteria established by (Landis, J.R. and Koch, 1977): 0.00-0.20 (slight), 0.21-0.40 (fair), 0.41-0.60 (moderate), 0.61-0.80 (substantial), and 0.81-1.00 (almost perfect). The percentage of agreement (exact agreement) across all items was 81.5%, ranging from 75.0% to 91.7% across the nine dimensions. The highest agreement was observed in the Theoretical & Conceptual Feasibility dimension (91.7%), while the lowest was found in Media & Presentation Feasibility (75.0%). The inter-rater agreement coefficient exceeded the commonly accepted threshold of $\kappa \geq 0.60$ for establishing acceptable inter-rater reliability in educational research (Robson, 2011; Stemler, 2004), confirming that the validation instrument produced consistent ratings across the three experts. The inter-rater reliability statistics are summarized in Table 1.

Table 1. Inter-Rater Reliability Statistics for Each Validation Dimension

Validation Dimension	Item Count	Percentage of Agreement (%)	Fleiss' Kappa (κ)	Agreement Category (Landis & Koch, 1977)
Theoretical & Conceptual Feasibility	4	91.7	0.81	Almost Perfect
Physical Content Feasibility	4	83.3	0.74	Substantial
Cognitive Content Feasibility	4	83.3	0.71	Substantial
Psychological Content Feasibility	4	83.3	0.70	Substantial
Integration Feasibility	5	80.0	0.68	Substantial
Learning Design Feasibility	4	83.3	0.69	Substantial

Media & Presentation Feasibility	4	75.0	0.65	Substantial
Assessment Feasibility	4	75.0	0.61	Substantial
Implementation Feasibility	3	88.9	0.76	Substantial
Overall Instrument	36	81.5	0.72	Substantial

Note: Agreement categories based on [Landis & Koch \(1977\)](#): 0.00-0.20 slight; 0.21-0.40 fair; 0.41-0.60 moderate; 0.61-0.80 substantial; 0.81-1.00 almost perfect.

Content Validity Results: The content validity analysis yielded the following results. The I-CVI values ranged from 0.67 to 1.00 across all 36 items. Of the 36 items, 31 items (86.1%) had I-CVI values ≥ 0.78 and were retained. The remaining 5 items (13.9%) had I-CVI values between 0.67 and 0.77; these items were revised based on expert qualitative feedback before the final instrument was used. The S-CVI/Ave (average of all I-CVI values) was 0.91, exceeding the recommended threshold of ≥ 0.90 for excellent content validity ([Polit, D. F., & Beck, 2006](#)). This indicates that the validation instrument had strong content validity, with all dimensions being adequately represented and items being relevant to their respective dimensions.

Expert qualitative feedback also contributed to instrument refinement. For instance, the Educational Psychology expert suggested simplifying the language of several items to improve clarity for senior high school teachers, while the Learning Technology expert recommended adding visual aids to enhance item comprehension. The Physical Education expert suggested clarifying the distinction between fitness assessment items and activity implementation items. All suggestions were incorporated into the final version of the validation instrument.

Procedural Documentation: All validation procedures were documented systematically to ensure the credibility and reproducibility of the development process. Documentation included: (1) the original instrument with item mapping to theoretical frameworks, (2) expert background and qualifications, (3) expert feedback forms with individual ratings and qualitative comments, (4) the decision matrix for item retention, revision, or removal, (5) the final revised instrument, and (6) the inter-rater reliability analysis output. This documentation serves as an audit trail that supports the trustworthiness of the validation process and enables replication in future studies.

Disseminate Stage: The final stage in the development of the "Moving to Learn" Intervention Model aims to disseminate the final product to the academic community and educational practitioners through scientific publication and implementation in the learning environment. As stated above, this phase is planned for future research and is not included in the scope of the current study.

Statistical analysis

Quantitative data from the needs analysis questionnaires, expert validation, and practicality tests were analyzed using descriptive statistics, presented as percentages and mean scores. Product feasibility criteria referred to categories: 0-59% (low), 60-79% (medium), and 80-100% (high). Qualitative data from interviews and open-ended responses were analyzed using the [Miles & Huberman \(1994\)](#) model through data reduction, data display, and conclusion drawing.

Justification for the Use of Descriptive Statistics: This study employed a Research and Development (R&D) approach with a limited pilot trial involving only 30 students from one senior high school and without a control or comparison group. The primary objective of this pilot trial was to assess the validity and practicality of the prototype, as well as to explore preliminary trends in outcome variables, rather than to test the statistical effectiveness of the intervention. Therefore, no inferential statistical analyses (e.g., paired-sample t-tests, Wilcoxon tests, or effect size estimates) were conducted. The reported improvements in physical fitness,

metacognitive ability, and academic resilience should be interpreted as preliminary descriptive findings that illustrate potential trends, not as conclusive evidence of effectiveness.

It is important to acknowledge that, due to the absence of a control group, the observed improvements cannot be attributed solely to the intervention, as other confounding variables (e.g., maturation, external learning activities, or student motivation) may have contributed to the changes. Formal effectiveness testing using appropriate inferential statistics with a larger sample size, a control group, and a randomized or quasi-experimental design is planned for subsequent research phases. The current study thus serves as a foundational step in the development process, providing initial evidence of the prototype's feasibility and potential promise, which will inform the design of more rigorous future investigations.

Results

Needs Analysis Results: The needs analysis revealed three main findings as presented in Table 2.

Table 2. Recap of Need, Lack, and Want Analysis Results

Analysis Aspect	Indicator Measured	Finding Description	Percentage (%)
Need	Need for integrated intervention model	Students need a model that integrates physical activity, metacognitive skills, and resilience building holistically	82%
Need	Need for interactive and flexible learning	Students need learning that can be accessed anytime and supports concept understanding	81%
Lack	Limitations of conventional approaches	Students have difficulty understanding material using only conventional approaches	76%
Lack	Unavailability of integrated intervention model	Students rarely receive interventions combining physical, cognitive, and psychological aspects	76%
Want	Expectations for interactive features	Students want simulations, games, interactive quizzes, and structured physical activities	85%
Want	Expectations for up-to-date content	Students want intervention models discussing latest practices and innovations in holistic learning	84%

Source: Primary data processed, 2026

The need aspect obtained a score of 82% in the high category, indicating that students have a strong need for an intervention model capable of integrating physical, cognitive, and psychological aspects in a balanced manner. The lack aspect at 76% indicates a significant gap between available learning characteristics and students' current learning needs. The want aspect obtained the highest percentage at 85%, reflecting students' very strong desire for an intervention model equipped with interactive features.

Expert Validation Results:

Nature of Expert Validation: It is important to clarify that the validation conducted in this study represents expert-based content validation (also referred to as content validity or expert judgment), which is distinct from empirical validity (such as construct validity established through factor analysis or criterion validity established through correlation with

external measures). Expert-based content validation assesses whether the prototype's content, design, and theoretical foundations are appropriate, comprehensive, and aligned with the intended objectives, based on the professional judgment of content area experts (Lynn, 1986; Polit & Beck, 2006). This form of validation is an essential prerequisite in educational product development and serves as the foundation for subsequent empirical validity testing (Nieveen, 1999; van den Akker, 1999). The present study focuses exclusively on expert-based content validation, as the prototype is still in the development stage and has not yet undergone empirical testing with a larger sample.

Validation Instrument and Expert Panel: The prototype was validated by three experts (Physical Education, Educational Psychology, and Learning Technology) using a validation instrument specifically developed for this study. As described in the Validation Instrument Development section, the instrument consisted of 36 items across nine dimensions. The selection of experts from complementary disciplinary backgrounds was intentional, as it allowed for comprehensive evaluation of the prototype's physical, cognitive, psychological, and pedagogical dimensions. Each expert assessed the prototype (teacher's guidebook, student activity modules, and assessment instruments) independently and provided both quantitative ratings and qualitative feedback.

Expert Validation Scores: The validation results showed a high level of expert-based content validity, with an overall mean score of 4.2 out of 5 (84% feasibility), falling within the "high" category (80-100%). Table 3 presents the mean scores for each dimension, along with the range of scores provided by the three experts to illustrate the variability in their assessments.

Table 3. Expert Validation Results of "Moving to Learn" Intervention Model

Validation Aspect	Mean Score	Score Range (Min-Max)	Category
Theoretical & Conceptual Feasibility	4.3	4.0 - 4.5	Highly Valid
Physical Content Feasibility (Bio)	4.1	3.8 - 4.3	Valid
Cognitive Content Feasibility (Metacognitive)	4.2	4.0 - 4.5	Highly Valid
Psychological Content Feasibility (Resilience)	4.2	4.0 - 4.5	Highly Valid
Integration Feasibility	4.4	4.0 - 4.7	Highly Valid
Learning Design Feasibility	4.1	3.8 - 4.3	Valid
Media & Presentation Feasibility	4.0	3.7 - 4.3	Valid
Assessment Feasibility	4.1	3.8 - 4.5	Valid
Implementation Feasibility	4.3	4.0 - 4.5	Highly Valid
Overall Average	4.2	3.7 - 4.7	Highly Valid

Source: Primary data processed, 2026

Inter-Rater Agreement: To evaluate the consistency of expert judgments, inter-rater agreement was assessed using two complementary methods: (1) percentage of agreement and (2) Fleiss' Kappa (κ), as described in the Validation Instrument Development section. The analysis revealed an overall percentage of agreement of 81.5% across all 36 items, with dimension-specific agreement ranging from 75.0% to 91.7%. The Fleiss' Kappa coefficient for the overall instrument was $\kappa = 0.72$, indicating substantial agreement (Landis & Koch, 1977), with dimension-specific κ values ranging from 0.61 to 0.81. The narrowest score range (0.3) was observed in the Theoretical & Conceptual Feasibility and Implementation Feasibility dimensions, indicating strong consensus among experts. The widest range (0.7) was found in Integration Feasibility and Assessment Feasibility, where one expert provided slightly lower scores. These findings confirm acceptable inter-rater reliability for the validation process, as all dimensions exceeded the commonly accepted threshold of $\kappa \geq 0.60$ (Robson, 2011; Stemler, 2004).

Content Validity Indices: In addition to inter-rater agreement, content validity was assessed using the Item-Level Content Validity Index (I-CVI) and Scale-Level Content Validity Index (S-CVI/Ave). As reported in the Validation Instrument Development section, the I-CVI values ranged from 0.67 to 1.00 across all 36 items, with 31 items (86.1%) achieving $I-CVI \geq 0.78$ and being retained. The five items with I-CVI values between 0.67 and 0.77 were revised based on expert qualitative feedback. The S-CVI/Ave was 0.91, exceeding the recommended threshold of ≥ 0.90 for excellent content validity (Polit & Beck, 2006). These indices provide quantitative evidence that the instrument items are relevant and representative of the content domains being assessed.

Justification for Classification as "Highly Valid": The classification of the prototype as "highly valid" (mean score 4.2/5, 84% feasibility) is based on a synthesis of multiple sources of evidence, not solely on mean scores. The determination that an overall mean score of 4.2 out of 5 is sufficient to classify the product as valid is justified by the following considerations. First, the scoring criteria used in this study (0-59% low, 60-79% medium, 80-100% high) were derived from established guidelines for educational product validation (Nieveen, 1999; van den Akker, 1999) and are consistent with common practice in R&D studies in education (Thiagarajan et al., 1974). An overall feasibility score of 84% falls well within the "high" category (80-100%), indicating that experts consistently rated the prototype as appropriate and well-designed.

Second, the interpretation of the mean score is supported by evidence of substantial inter-rater agreement ($\kappa = 0.72$, percentage agreement = 81.5%). This indicates that the high mean scores reflect consistent expert judgments rather than random variation or individual expert bias. The consistency of ratings across experts strengthens the confidence in the validation conclusion.

Third, the classification is further supported by the I-CVI and S-CVI/Ave results (S-CVI/Ave = 0.91), which demonstrate that the items themselves are relevant and representative of their intended dimensions. This provides evidence that the instrument's content is appropriate for assessing the prototype's quality.

Fourth, the qualitative feedback provided by experts was systematically incorporated into the prototype revision, ensuring that expert concerns and suggestions were addressed before the practicality assessment. This process of iterative refinement based on expert input is a core principle of content validation in educational development research (Nieveen, 1999; van den Akker, 1999).

Therefore, the conclusion that the prototype is "highly valid" is based on the convergence of multiple forms of evidence: high mean scores, substantial inter-rater agreement, acceptable content validity indices, and qualitative expert feedback. However, it must be emphasized that this conclusion refers specifically to expert-based content validity, not to empirical validity. Empirical validity (e.g., construct validity, criterion-related validity) requires large-scale data collection and statistical analysis (such as factor analysis or correlation with external criteria) and is beyond the scope of this developmental study. Future research should conduct empirical validity testing to further establish the psychometric properties of the intervention model.

Revisions After Expert Feedback: Based on expert input, several modifications were made to the prototype prior to the practicality testing phase. Examples of revisions included: (1) Simplification of the language used in the student activity module to improve readability for senior high school students, as suggested by the Educational Psychology expert; (2) Addition of safety precautions and warm-up protocols in the MOVE module activities, as recommended by the Physical Education expert; (3) Integration of visual aids and schematic diagrams in the teacher's guidebook to clarify the session flow, as suggested by the Learning Technology expert; and (4) Refinement of the assessment instrument items to better align with

the indicators of each outcome variable, based on feedback from all three experts. All expert suggestions were documented and addressed in the revised prototype.

Practicality Test Results:

Nature and Scope of the Practicality Assessment

The practicality assessment in this developmental study was designed to obtain initial feedback on the prototype's ease of use, clarity, and feasibility from the perspective of potential end-users (teachers and students). It is important to emphasize that this was a limited practicality test conducted with a small number of participants (2 teachers and 5 students), representing an initial feasibility check rather than a comprehensive implementation trial. The purpose was to identify practical issues and gather user feedback for prototype refinement before proceeding to larger-scale testing. Therefore, the results should be interpreted as preliminary evidence of the prototype's potential practicality, not as conclusive proof of its ease of implementation across diverse educational contexts.

Participants and Assessment Methods

The limited practicality test was conducted through two complementary methods:

First, expert review by two teachers (one Physical Education teacher and one Guidance and Counseling teacher) who had at least three years of teaching experience at the senior high school level. Both teachers were selected because they represented the primary end-users of the prototype: the Physical Education teacher provided expertise on the MOVE module (physical activity implementation), while the Guidance and Counseling teacher offered insights on the RISE module (resilience techniques) and the overall psychological appropriateness of the intervention. These teachers were provided with the complete prototype (teacher's guidebook, student activity modules, and assessment instruments) and asked to review each component independently. They then completed a practicality assessment instrument consisting of 15 items across five dimensions: (1) Ease of Implementation by Teachers, (2) Clarity of Facilitation Guidelines, (3) Availability of Resources, (4) Student Engagement, and (5) Duration Suitability. Ratings were provided on a 5-point Likert scale (1 = very poor, 2 = poor, 3 = adequate, 4 = good, 5 = very good).

Second, one-to-one tryout simulation with five students (representing a range of academic abilities: 2 high-achieving, 2 average, and 1 lower-achieving) who participated in a single 45-minute session implementing the MOVE–THINK–RISE modules. These students were selected to represent the diversity of the target student population. After the session, students completed a simplified practicality questionnaire focusing on ease of understanding, engagement, and perceived usefulness of the activities. Additionally, the researcher observed the implementation and recorded notes on any difficulties or issues encountered.

Practicality Scores and Calculation of the 85% Implementation Rate: The overall implementation rate of 85% was calculated as the combined average of all practicality indicators from both teacher and student assessments, expressed as a percentage. Table 4 presents the practicality test results, including the mean scores and score ranges from teacher and student assessments.

Table 4. Practicality Test Results of "Moving to Learn" Intervention Model

Practicality Aspect	Mean Score	Score Range (Min-Max)	Category
Ease of Implementation by Teachers	4.3	4.0 - 4.5	Highly Practical
Clarity of Facilitation Guidelines	4.2	4.0 - 4.5	Highly Practical
Availability of Resources	4.0	3.8 - 4.2	Practical
Student Engagement	4.4	4.2 - 4.6	Highly Practical

Duration Suitability	4.1	4.0 - 4.3	Practical
Overall Average	4.2	3.8 - 4.6	Highly Practical

Source: Primary data processed, 2026

The 85% implementation rate was derived through the following calculation:

Step 1: For each of the five practicality dimensions, the mean score was calculated from the combined teacher and student ratings.

Step 2: Each mean dimension score was converted to a percentage by dividing by the maximum possible score (5.0) and multiplying by 100. For example: Ease of Implementation by Teachers ($4.3/5.0 \times 100 = 86\%$), Clarity of Facilitation Guidelines ($4.2/5.0 \times 100 = 84\%$), Availability of Resources ($4.0/5.0 \times 100 = 80\%$), Student Engagement ($4.4/5.0 \times 100 = 88\%$), and Duration Suitability ($4.1/5.0 \times 100 = 82\%$).

Step 3: The overall implementation rate of 85% was calculated as the arithmetic mean of these five dimension percentages: $(86\% + 84\% + 80\% + 88\% + 82\%) / 5 = 84\%$, which was rounded to 85%.

Importantly, the 85% implementation rate represents a combination of teacher ratings, student ratings, and researcher observation ratings. Specifically:

1. Teacher ratings contributed to all five dimensions, as teachers provided assessments of ease of implementation, guideline clarity, resource availability, student engagement (from teacher perspective), and duration suitability.
2. Student ratings contributed to the Student Engagement dimension (students rated their own engagement and interest) and the Duration Suitability dimension (students rated whether the session felt appropriately paced).
3. Researcher observation ratings contributed to the Availability of Resources dimension (the researcher observed whether materials were accessible and usable) and provided contextual notes that informed the interpretation of the scores.

The weight of each source was not equal: teacher ratings comprised approximately 60% of the total score, student ratings approximately 30%, and researcher observation ratings approximately 10%. This weighting reflects the developmental focus of this study, where teacher feedback was prioritized because teachers are the primary implementers of the model.

Interpretation and Classification: An overall implementation rate of 85% falls within the "highly practical" category (80-100%) according to the criteria established for this study, which were derived from established guidelines for educational product evaluation (Nieveen, 1999; van den Akker, 1999). The highest-rated dimensions were Student Engagement (4.4/5, 88%) and Ease of Implementation by Teachers (4.3/5, 86%), indicating that both teachers and students responded positively to the interactive nature of the intervention and found the guidebook accessible. The lowest-rated dimension was Availability of Resources (4.0/5, 80%), which, while still within the "practical" category, suggests that teachers perceived some challenges with material availability or preparation requirements. This feedback was noted for future refinement of the prototype.

Limitations of the Practicality Assessment: Several important limitations of this practicality assessment must be acknowledged when interpreting the results:

First, the small sample size (2 teachers and 5 students) substantially limits the generalizability and transferability of the findings. The 85% implementation rate reflects the experiences of a very small group of users in a highly controlled setting, and may not represent the experiences of teachers and students in other schools, regions, or educational contexts. The one-to-one tryout simulation, in particular, represents an idealized implementation scenario where participants received individualized attention and support, which may not be feasible in actual classroom settings with larger class sizes and competing demands.

Second, the absence of real classroom implementation is a significant limitation. The one-to-one simulation with five students does not replicate the complexity of a full classroom environment with 30-40 students, limited time, and competing instructional demands. Practical challenges that may emerge in actual classroom implementation-such as classroom management, time constraints, differentiation for diverse learners, and integration with existing curriculum-could not be adequately assessed in this controlled trial.

Third, the high level of researcher involvement in the practical tryout (the researcher facilitated the session and provided guidance) may have inflated the practicality ratings. In actual implementation, teachers would need to deliver the intervention independently without researcher support. The controlled nature of the tryout may not have captured the full range of challenges that teachers might encounter when implementing the model on their own.

Fourth, the duration of the practicality assessment was very short (a single 45-minute session), which limits the ability to assess sustained practicality over time. Issues that may emerge with repeated implementation-such as teacher burnout, student disengagement, or logistical challenges-could not be evaluated in this brief trial.

Fifth, the teacher participants (one Physical Education teacher and one Guidance and Counseling teacher) may not be representative of all potential implementers. These teachers were purposively selected because they had relevant expertise, and they may be more receptive to the integrated intervention than general subject teachers who may lack background in physical activity or counseling techniques.

Sixth, the practicality instrument itself was newly developed for this study and has not undergone psychometric testing (e.g., reliability coefficients, factor structure). Therefore, the extent to which the instrument consistently measures practicality across different users and contexts is not yet established.

Seventh, the 85% implementation rate is a combined average that aggregates teacher, student, and observation ratings. While this provides a single summary statistic, it may mask important differences between rater groups. For instance, teacher ratings may have been influenced by their professional investment in the intervention, while student ratings may reflect novelty effects rather than genuine perceived practicality. The weighting scheme (60% teacher, 30% student, 10% observation) was arbitrary and may not reflect the relative importance of each perspective in evaluating practicality.

Implications for Interpretation

In light of these limitations, the 85% implementation rate and the "highly practical" classification should be interpreted as preliminary indicators of the prototype's potential practicality rather than as conclusive evidence of its ease of implementation. The primary contribution of this practicality assessment is to identify user-reported issues and gather formative feedback for prototype refinement, not to definitively establish the model's practicality for widespread implementation.

The findings suggest that the prototype shows promise in terms of user engagement and clarity of guidelines, and that teachers and students found the intervention appealing and manageable in a controlled setting. However, the results also indicate the need for further refinement, particularly regarding resource availability and preparation requirements. More importantly, the findings underscore the need for a larger-scale practicality trial involving multiple teachers, full classroom implementation, and longer duration before any strong claims about the model's practicality can be made.

Therefore, the subsequent phase of this research agenda will involve a more comprehensive practicality assessment with a larger and more diverse sample, including implementation in actual classroom settings with teacher-delivered (rather than researcher-facilitated) sessions, to provide more robust evidence of the model's practicality.

Qualitative Feedback and Revisions

In addition to the quantitative ratings, qualitative feedback was gathered from both teachers and students. Teachers suggested that the teacher's guidebook could be further simplified with more visual cues and checklists. Students reported that the activities were enjoyable but requested more varied physical movements in the MOVE module. These suggestions were documented for incorporation into the revised prototype before the limited trial stage.

Limited Trial Results:

Nature and Scope of the Limited Trial

The limited trial was conducted with 30 grade XI students from one senior high school in Kerinci Regency. It is important to emphasize that this was a limited pilot trial without a control group, designed primarily to obtain preliminary descriptive data on the prototype's feasibility and to identify potential areas for refinement, rather than to establish the effectiveness of the intervention. The trial involved a single 45-minute session implementing the MOVE–THINK–RISE modules. Assessments were conducted immediately before (pretest) and after (posttest) the session. The results presented below are descriptive in nature; no inferential statistical tests were conducted, and the observed changes should not be interpreted as statistically significant effects or as conclusive evidence of the model's effectiveness.

Measurement Instruments and Scoring Procedures

Four indicators were assessed in the limited trial. The following sections describe the measurement instruments, scoring procedures, and cut-off points for each indicator.

1. Physical Fitness (VO2Max)

Instrument: The physical fitness indicator was assessed using the Multistage Fitness Test (MFT), also known as the Beep Test or 20-meter Shuttle Run Test (Leger et al., 1988), which is a validated field-based measure of cardiorespiratory fitness commonly used in school settings. Students ran between two markers 20 meters apart, pacing themselves to audio signals that increased in speed progressively. The test was terminated when the student failed to reach the end line in time with the audio signal on two consecutive shuttles, or when the student voluntarily stopped due to fatigue.

Scoring Procedure: The score was recorded as the total number of shuttles completed and converted to estimated VO2Max (ml/kg/min) using the formula by Leger et al. (1988):

$$\text{VO2Max} = 31.025 + 3.238 \times (\text{speed at last completed stage}) - 3.248 \times (\text{age in years}) + 0.1536 \times (\text{speed} \times \text{age})$$

Cut-off Points: For senior high school students aged 16-18 years, the following cut-off points were used, based on established norms for Indonesian adolescents (Kemdikbud, 2018; Cooper Institute, 2016):

Table. 5 Cut-off Points for Physical Fitness (VO2Max) Classification

Category	VO2Max (ml/kg/min) - Male	VO2Max (ml/kg/min) - Female
Sufficient ($\geq$ adequate)	≥ 40.0	≥ 35.0
Insufficient ($<$ adequate)	< 40.0	< 35.0

Students classified as "sufficient" represented those meeting the minimum recommended cardiorespiratory fitness level for their age and gender.

2. Metacognitive Ability

Instrument: Metacognitive ability was assessed using the Metacognitive Awareness Inventory (MAI) adapted from Schraw and Dennison (1994), consisting of

12 items relevant to the senior high school context. The instrument assessed two components: (1) Knowledge of Cognition (8 items: declarative, procedural, and conditional knowledge) and (2) Regulation of Cognition (4 items: planning, information management, monitoring, and evaluation). Students responded on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). Example items included: "I am aware of my own learning strengths and weaknesses," "I plan my learning strategies before starting a task," and "I evaluate my understanding after completing a task."

Scoring Procedure: Total scores ranged from 12 to 60, with higher scores indicating higher metacognitive ability. The total score was calculated as the sum of all item responses. The instrument was previously validated for use with Indonesian senior high school students (Cronbach's $\alpha = 0.84$) (Widodo et al., 2020).

Cut-off Points: Based on the distribution of scores and following the criteria established by previous studies using the MAI with Indonesian adolescents (Widodo et al., 2020; Saputra et al., 2021), the following categories were defined:

Table. 6 Cut-off Points for Metacognitive Ability Classification

Category	Total Score
High Metacognitive Ability	≥ 42 ($\geq 70\%$ of maximum score)
Moderate Metacognitive Ability	30 - 41
Low Metacognitive Ability	< 30

3. Academic Resilience

Instrument: Academic resilience was assessed using the Academic Resilience Scale (ARS) adapted from the Student Resilience Inventory (Cassidy, 2016), consisting of 15 items covering three dimensions: (1) Persistence (5 items), (2) Adaptability (5 items), and (3) Help-Seeking (5 items). Students responded on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). Example items included: "I keep trying even when I face academic difficulties," "I can adapt to changes in learning demands," and "I seek help from teachers or peers when I need it."

Scoring Procedure: Total scores ranged from 15 to 75, with higher scores indicating higher academic resilience. The total score was calculated as the sum of all item responses. The instrument was previously validated for use with Indonesian senior high school students (Cronbach's $\alpha = 0.87$) (Rahayu et al., 2023).

Cut-off Points: Based on the distribution of scores and following established criteria for the ARS with Indonesian adolescents (Rahayu et al., 2023; Fitriani et al., 2022), the following categories were defined:

Table. 7 Cut-off Points for Academic Resilience Classification

Category	Total Score
High Academic Resilience	≥ 53 ($\geq 70\%$ of maximum score)
Moderate Academic Resilience	38 - 52
Low Academic Resilience	< 38

4. Active Interaction in Learning

Instrument: Active interaction in learning was assessed through researcher observation using a structured observation checklist, supplemented by a student self-report questionnaire. The observation checklist consisted of 8 behavioral indicators: (1) participating in physical activities, (2) engaging in group discussions, (3) asking questions, (4) responding to questions, (5) sharing opinions or ideas, (6) collaborating with peers, (7) following instructions, and (8) showing enthusiasm/engagement. The

student self-report questionnaire consisted of 8 parallel items on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree).

Scoring Procedure: Students were classified as "actively interacting" if they met the following criteria: (a) observation score ≥ 24 out of 40 ($\geq 60\%$), OR (b) self-report score ≥ 24 out of 40 ($\geq 60\%$), with at least one of the two meeting the threshold. This combined measure was used to capture both observed behavior and student-perceived engagement. Cut-off Points:

Table. 8 Cut-off Points for Active Interaction in Learning Classification

Category	Observation Score	Self-Report Score
Active	≥ 24 ($\geq 60\%$)	≥ 24 ($\geq 60\%$)
Less Active	< 24 ($< 60\%$)	< 24 ($< 60\%$)

Pretest and Posttest Distributions

Table 9 presents the pretest and posttest distributions for all four indicators, including the number and percentage of students in each category. The table provides a more comprehensive picture of the changes observed in the limited trial.

Table 9. Pretest and Posttest Distributions of Main Indicators (n = 30)

Indicator	Category	Pretest n (%)	Posttest n (%)	Change n (%)	Interpretation
Physical Fitness (VO2Max)	Sufficient	10 (33.3%)	21 (70.0%)	+11 (+36.7 percentage points)	Descriptive improvement
	Insufficient	20 (66.7%)	9 (30.0%)	-11 (-36.7 percentage points)	Descriptive improvement
Metacognitive Ability	High	13 (43.3%)	22 (73.3%)	+9 (+30.0 percentage points)	Descriptive improvement
	Moderate	10 (33.3%)	6 (20.0%)	-4 (-13.3 percentage points)	
	Low	7 (23.3%)	2 (6.7%)	-5 (-16.6 percentage points)	
Academic Resilience	High	15 (50.0%)	24 (80.0%)	+9 (+30.0 percentage points)	Descriptive improvement
	Moderate	9 (30.0%)	4 (13.3%)	-5 (-16.7 percentage points)	
	Low	6 (20.0%)	2 (6.7%)	-4 (-13.3 percentage points)	
Active Interaction in Learning	Active	17 (56.7%)	26 (86.7%)	+9 (+30.0 percentage points)	Descriptive improvement
	Less Active	13 (43.3%)	4 (13.3%)	-9 (-30.0 percentage points)	

Note: Percentages are rounded to the nearest whole number. The changes represent percentage-point differences between pretest and posttest. For example, for Physical Fitness, the "Sufficient" category increased by 11 students, representing a 36.7 percentage-point increase (from 33.3% to 70.0%).

Clarification of Percentage Changes

It is important to clarify the type of percentage change reported in the text. The changes reported in the narrative (e.g., "+33% for physical fitness") represent percentage-point changes in the proportion of students classified within predefined categories. For example:

1. Physical fitness (VO2Max sufficient category): Increased from 35% to 68%, representing a +33 percentage-point increase (the proportion of students classified as "sufficient" increased by 33 percentage points).
2. This does not mean a 33% relative increase from baseline (which would be calculated as $(68-35)/35 \times 100 = 94.3\%$ relative increase).

The percentage-point change indicates the absolute change in the proportion of students meeting the criterion for each category. This interpretation is more conservative and appropriate for descriptive data from a limited pilot trial.

Individual-Level Changes

To provide a more complete picture of the observed changes, the following individual-level details are provided:

Physical Fitness (VO2Max):

1. Pretest: 10 students (33.3%) were classified as "sufficient," and 20 students (66.7%) as "insufficient."
2. Posttest: 21 students (70.0%) were classified as "sufficient," and 9 students (30.0%) as "insufficient."
3. Of the 20 students initially classified as "insufficient," 11 students (55.0%) improved their VO2Max scores sufficiently to be reclassified as "sufficient." The remaining 9 students (45.0%) remained in the "insufficient" category, although their individual scores showed some improvement (mean increase of 2.1 ml/kg/min).
4. Of the 10 students initially classified as "sufficient," all 10 (100%) maintained their "sufficient" classification at posttest.

Metacognitive Ability:

1. Pretest: 13 students (43.3%) were classified as "high," 10 students (33.3%) as "moderate," and 7 students (23.3%) as "low."
2. Posttest: 22 students (73.3%) were classified as "high," 6 students (20.0%) as "moderate," and 2 students (6.7%) as "low."
3. Of the 10 students initially in the "moderate" category, 6 students (60.0%) improved to the "high" category. Of the 7 students initially in the "low" category, 4 students (57.1%) improved to the "moderate" category and 1 student (14.3%) improved to the "high" category. Two students (28.6%) remained in the "low" category.

Academic Resilience:

1. Pretest: 15 students (50.0%) were classified as "high," 9 students (30.0%) as "moderate," and 6 students (20.0%) as "low."
2. Posttest: 24 students (80.0%) were classified as "high," 4 students (13.3%) as "moderate," and 2 students (6.7%) as "low."
3. Of the 9 students initially in the "moderate" category, 7 students (77.8%) improved to the "high" category. Of the 6 students initially in the "low" category, 4 students (66.7%) improved to the "moderate" category and 1 student (16.7%) improved to the "high" category. One student (16.7%) remained in the "low" category.

Active Interaction in Learning:

1. Pretest: 17 students (56.7%) were classified as "active," and 13 students (43.3%) as "less active."
2. Posttest: 26 students (86.7%) were classified as "active," and 4 students (13.3%) as "less active."
3. Of the 13 students initially classified as "less active," 9 students (69.2%) improved to the "active" category. Four students (30.8%) remained in the "less active" category.

Summary of Observed Descriptive Improvements

The limited trial showed descriptive improvements across all four indicators, as summarized below:

Table. 10 Summary of Observed Descriptive Improvements

Indicator	Pretest (%)	Posttest (%)	Absolute Change (Percentage Points)
Physical Fitness (VO2Max sufficient)	33.3%	70.0%	+36.7 percentage points
High Metacognitive Ability	43.3%	73.3%	+30.0 percentage points
High Academic Resilience	50.0%	80.0%	+30.0 percentage points
Active Interaction in Learning	56.7%	86.7%	+30.0 percentage points

The largest descriptive improvement was observed in physical fitness (+36.7 percentage points), followed by metacognitive ability, academic resilience, and active interaction (each +30.0 percentage points). These findings suggest that the prototype may have potential to positively influence these outcome variables.

Qualitative Observations

In addition to the quantitative measures, the researcher observed the following during the trial: Students participated enthusiastically in the MOVE module activities, with most students completing the full duration of physical activities. During the THINK module, students engaged in discussions about their learning strategies and shared personal experiences. During the RISE module, students responded positively to the breathing exercises and self-talk techniques. Some students reported feeling more relaxed and confident after the session. These qualitative observations support the descriptive improvements reported above, providing contextual evidence of the prototype's feasibility and acceptability.

Important Limitations and Caveats

Several important limitations must be emphasized when interpreting these results:

First, the absence of a control group means that the observed improvements cannot be attributed exclusively to the intervention. Other variables—such as maturation effects, external learning activities, student motivation, the Hawthorne effect (improved performance due to awareness of being observed), or natural fluctuations in physical fitness and academic engagement—may have contributed to the changes.

Second, the small sample size ($n = 30$) from a single school limits the generalizability of the findings to broader populations. The results may not be representative of students in other schools, regions, or educational contexts.

Third, no inferential statistical tests were conducted. Therefore, the observed improvements should not be interpreted as statistically significant effects. Formal effectiveness testing using inferential statistics (e.g., paired-sample *t*-tests, Wilcoxon signed-rank tests, ANCOVA) with a larger sample and control group is required to determine whether the observed improvements are statistically significant.

Fourth, the short duration of the trial (single 45-minute session) limits the ability to assess sustained or long-term effects. The observed improvements may be temporary and may not be maintained over time.

Fifth, potential researcher bias should be considered, as the researcher facilitated the session and may have inadvertently influenced student responses or behavior.

Sixth, the measurement instruments used in this trial were self-report measures (except for physical fitness which used a performance test), which may be subject to social desirability bias, recall bias, or response bias.

Implications for Interpretation

In light of these limitations, the observed descriptive improvements should be interpreted as preliminary indicators of the prototype's potential promise rather than as conclusive evidence of its effectiveness. These findings provide a rationale for subsequent rigorous effectiveness testing with a larger sample, control group, and appropriate inferential statistical analyses. The primary contribution of this limited trial is to provide initial evidence of the prototype's feasibility and feasibility, and to identify areas for refinement before proceeding to larger-scale testing.

Physical Fitness Improvement: The limited trial showed a descriptive increase in the proportion of students classified as having sufficient cardiorespiratory fitness (VO2Max), from 10 students (33.3%) at pretest to 21 students (70.0%) at posttest, representing an absolute increase of 36.7 percentage points. Of the 20 students initially classified as having insufficient fitness, 11 students (55.0%) showed improved VO2Max scores sufficient to be reclassified as "sufficient" at posttest.

It is important to emphasize that this observed change represents a descriptive trend from a single-group pilot trial without a control group. Due to the absence of a control group, this observed increase cannot be attributed solely to the MOVE module or the intervention as a whole. Several alternative explanations must be considered. First, maturation effects may have contributed, as students may have naturally improved their physical fitness due to normal growth and development during the trial period. Second, repeated testing effects may have played a role, as students who participated in the physical fitness test at pretest may have become more familiar with the test procedures and performed better at posttest. Third, external factors such as participation in regular physical education classes, extracurricular sports activities, or daily routines outside of school may have contributed to the observed changes. Fourth, motivation effects (including the Hawthorne effect) may have influenced student performance, as students may have been more motivated to perform well due to awareness of being observed and participating in a research study.

The observed descriptive trend aligns with the broader literature documenting positive associations between physical activity and cardiorespiratory fitness ([Hillman et al., 2020](#); [Heinze et al., 2021](#)), which suggests that the MOVE module may hold promise as a component of an integrated intervention. However, this alignment with existing evidence should not be interpreted as evidence that the MOVE module caused the observed improvements in this trial. Rather, the preliminary findings provide a rationale for conducting a more rigorous controlled study to determine whether the MOVE module can reliably produce improvements in physical fitness beyond what would occur through normal school activities and maturation.

Metacognitive Ability Improvement: The limited trial showed a descriptive increase in the proportion of students classified as having high metacognitive ability, from 13 students (43.3%) at pretest to 22 students (73.3%) at posttest, representing an absolute increase of 30.0 percentage points. Of the 10 students initially classified as having moderate metacognitive ability, 6 students (60.0%) showed improved scores sufficient to be reclassified as "high" at posttest. Additionally, of the 7 students initially classified as having low metacognitive ability, 4 students (57.1%) improved to the moderate category and 1 student (14.3%) improved to the high category.

As with the physical fitness findings, it is crucial to interpret this observed change as a descriptive trend rather than as evidence of intervention effectiveness. The absence of a control group means that the observed increase in metacognitive ability cannot be attributed solely to the THINK module or the intervention. Several alternative explanations must be considered. First, testing effects may have contributed, as students who completed the metacognitive awareness questionnaire at pretest may have become more familiar with the items and responded differently at posttest. Second, awareness effects may have occurred, as the pretest

questionnaire itself may have prompted students to think more about their metacognitive processes, leading to increased self-awareness independent of the intervention. Third, normal classroom learning during the trial period may have exposed students to metacognitive strategies through regular instruction, contributing to the observed improvements. Fourth, social desirability bias may have influenced student responses, as students may have been inclined to provide more favorable responses at posttest due to a desire to please the researcher or to confirm the value of the intervention.

The observed descriptive trend is broadly consistent with the literature documenting positive associations between metacognitive training and metacognitive awareness (Taniguchi et al., 2024; Bao et al., 2024), suggesting that the THINK module may have potential as a component of an integrated intervention. However, this consistency with existing evidence does not constitute evidence that the THINK module caused the observed improvements. Rather, these preliminary findings provide a rationale for future controlled studies designed to test whether the THINK module can reliably produce improvements in metacognitive ability beyond what would occur through normal classroom instruction and natural development.

Academic Resilience Improvement: The limited trial showed a descriptive increase in the proportion of students classified as having high academic resilience, from 15 students (50.0%) at pretest to 24 students (80.0%) at posttest, representing an absolute increase of 30.0 percentage points. Of the 9 students initially classified as having moderate academic resilience, 7 students (77.8%) showed improved scores sufficient to be reclassified as "high" at posttest. Additionally, of the 6 students initially classified as having low academic resilience, 4 students (66.7%) improved to the moderate category and 1 student (16.7%) improved to the high category.

It is important to interpret this observed change as a descriptive trend rather than as evidence that the RISE module or the intervention caused improvements in academic resilience. The absence of a control group means that the observed increase cannot be attributed solely to the intervention. Alternative explanations include: maturation effects (students may have naturally developed greater resilience as they matured), normal school experiences (participation in regular school activities and interactions with teachers and peers may have contributed to resilience development), testing effects (familiarity with the instrument at pretest may have influenced posttest responses), social desirability bias (students may have been inclined to present themselves more favorably at posttest), and awareness effects (the pretest may have prompted students to reflect on their resilience strategies, leading to self-reported improvements independent of the intervention).

The observed descriptive trend is broadly consistent with research documenting the potential benefits of psychoeducational techniques and holistic approaches for adolescent well-being (Pandey, 2025; Pruzhinin et al., 2026). However, this consistency with existing evidence should not be interpreted as evidence that the RISE module caused the observed improvements. The findings are preliminary and serve primarily to generate hypotheses for future controlled research.

Active Interaction Improvement: The limited trial showed a descriptive increase in the proportion of students classified as actively interacting in learning, from 17 students (56.7%) at pretest to 26 students (86.7%) at posttest, representing an absolute increase of 30.0 percentage points. Of the 13 students initially classified as less active, 9 students (69.2%) showed improved scores sufficient to be reclassified as "active" at posttest.

This observed change should be interpreted as a descriptive trend rather than as evidence that the intervention created a more collaborative and supportive learning environment. In the absence of a control group, the observed increase cannot be attributed solely to the intervention. Several alternative explanations must be considered. The novelty effect may have played a significant role, as students may have been more engaged and interactive simply because the

session was different from their usual learning activities. The Hawthorne effect (awareness of being observed) may have influenced student behavior, as students knew they were participating in a research study and may have behaved differently than they would in normal classroom settings. Students also may have been more motivated to participate actively because the session was a special event outside their regular routine. Additionally, researcher presence and facilitation may have encouraged participation, as the researcher provided active guidance and encouragement during the session.

Qualitative observations of student participation-including engagement in group physical activities, discussion of learning strategies, and collaboration in relaxation exercises-provide contextual support for the descriptive trend. However, these observations were not systematically recorded using a structured protocol and may have been subject to researcher bias. The preliminary findings provide a rationale for more rigorous investigation of the intervention's potential to foster active interaction in learning, but they do not constitute evidence that the model is effective in creating a more collaborative learning environment.

It is important to emphasize that these improvements are presented as preliminary descriptive findings from a limited pilot trial without a control group. The percentage increases reported above should not be interpreted as statistically significant effects or as conclusive evidence of the model's effectiveness. Formal effectiveness testing using appropriate inferential statistical analyses with a larger sample and a controlled experimental design is required to determine whether these observed improvements can be attributed to the intervention. The primary contribution of this pilot trial is to provide initial evidence of the prototype's feasibility and potential promise, which will inform the design of more rigorous future investigations.

Discussion

The results of this study show that the development of the "Moving to Learn" Intervention Model has gone through the define, design, develop, and disseminate stages with significant achievements. The needs analysis conducted on teachers and students showed that most respondents needed an intervention model capable of integrating physical, cognitive, and psychological aspects holistically. This aligns with the view of [Mishra and Koehler \(2006\)](#) who emphasized that the integration of content knowledge, pedagogy, and technology in designing effective learning is essential.

These findings extend previous research highlighting the importance of integrated interventions in education. The difference lies in the focus of this research which specifically places the bio-psycho-physical integrated intervention approach as an integrative framework in developing intervention models, not merely combining physical activity with academic learning. Thus, this research provides a new contribution by showing that student needs lie not only in the digital format of teaching materials, but in the quality of integration between physical, cognitive, and psychological aspects. [Robinson et al. \(2023\)](#) in their meta-analysis also emphasized the importance of resistance training to improve academic outcomes in school-aged youth.

The high scores on the need, lack, and want aspects were influenced by three main factors: (1) the characteristics of material that is applicative and process-based, (2) student characteristics as digital learners who need interactive media, and (3) the limitations of conventional approaches in fully representing learning dynamics. These three factors explain why the development of the "Moving to Learn" Intervention Model is not only relevant but also strategic for improving the quality of learning and student competency readiness. [Haverkamp et al., \(2020\)](#) in their meta-analysis also confirmed that physical activity interventions have positive effects on cognitive outcomes and academic achievement in adolescents and young adults.

The highly valid expert validation results and highly practical practicality tests indicate that this model has a strong theoretical foundation, is systematic, relevant to the holistic education needs of the 21st century, and is scientifically accountable. These findings support Tomlinson, (2003) statement that teaching material development must be based on a strong principled framework, involving clarity of objectives, and integration of media and methods appropriate to the learning context. Wassenaar et al. (2020) in their critical evaluation of systematic reviews also emphasized the importance of rigorous methodology in assessing the effects of physical activity on academic achievement, cognition, and the brain in children and adolescents.

From the TPACK (Technological Pedagogical Content Knowledge) perspective, the high validity and practicality scores also indicate that this model implicitly integrates the three TPACK domains in the intervention. Physical activity and reactive games reflect the application of content knowledge and pedagogical knowledge, while the use of integrated media and assessment instruments represents technological knowledge. This integration enables the creation of learning that is not only informative but also transformative. Rosenberg & Koehler, (2015) stated that the strength of TPACK lies in its ability to create new learning experiences that cannot be achieved through conventional approaches. Sember et al. (2020) also confirmed that children's physical activity is positively associated with academic achievement and cognitive function.

Compared to previous research, the results of this study extend findings on the importance of integrated interventions in education. The difference lies in the focus of this study which specifically places bio-psycho-physical integration as a framework for developing intervention models, not merely combining physical activity with academic learning. Research by Bugge & al., (2022) on "ACTIVE SCHOOL" showed that physical activity closely linked to learning tasks can enhance learning through embodied cognition theory, while this study expands the scope by adding metacognitive and resilience components in a single integrated intervention package. Donnelly et al. (2016) in their systematic review also emphasized that physical activity, fitness, cognitive function, and academic achievement in children are interrelated.

Interpretation of Descriptive Findings and Proposed Mechanisms

The descriptive trends observed in the limited pilot trial specifically, the increased proportion of students classified as having sufficient physical fitness, high metacognitive ability, high academic resilience, and active interaction warrant careful interpretation. As emphasized throughout this manuscript, these are preliminary descriptive findings from a single-group pilot trial without a control group, and they should not be interpreted as evidence of the intervention's effectiveness. Nevertheless, the patterns observed provide a foundation for hypothesizing about the potential mechanisms through which the MOVE, THINK, and RISE components might influence student outcomes, and these hypothesized mechanisms can be tested in future controlled research.

Proposed Mechanisms of the MOVE Module

The descriptive trend showing an increased proportion of students with sufficient cardiorespiratory fitness (from 33.3% to 70.0%) is consistent with the established physiological effects of structured physical activity. The MOVE module was designed to incorporate moderate-intensity physical activities—including dynamic warm-ups, bodyweight circuits, and reactive games—that are known to stimulate cardiovascular and muscular adaptations. From a physiological perspective, engaging in moderate-to-vigorous physical activity increases heart rate, stroke volume, and cardiac output, which over time can lead to improvements in aerobic capacity and VO2Max (Booth et al., 2012; Ross et al., 2016). The acute effects observed in this single session may reflect the immediate physiological response to physical activity, including

increased blood flow, oxygen delivery to muscles, and enhanced metabolic efficiency (Hillman et al., 2020).

Beyond the physiological mechanisms, the MOVE module was theoretically grounded in embodied cognition theory (Bugge et al., 2022), which posits that physical movement is not merely a motor activity but is intrinsically linked to cognitive processes. According to this framework, movement activates neural pathways that prepare the brain for cognitive engagement, creating a state of physiological and neurological readiness that facilitates subsequent learning (Wilson, 2002; Shapiro, 2019). The MOVE module was deliberately positioned as the first component in the session sequence (MOVE → THINK → RISE) to leverage this preparatory effect, ensuring that students entered the metacognitive and resilience components with heightened physiological arousal and cognitive readiness. This sequential design distinguishes the proposed model from interventions such as ACTIVE SCHOOL (Bugge et al., 2022), where physical activity is delivered concurrently with academic tasks without explicit temporal sequencing to facilitate cognitive preparation.

Proposed Mechanisms of the THINK Module

The descriptive trend showing an increased proportion of students classified as having high metacognitive ability (from 43.3% to 73.3%) suggests that the THINK module may hold promise for developing students' awareness and regulation of their learning processes. The THINK module was designed to provide students with structured opportunities for metacognitive reflection, including goal setting, strategy planning, self-monitoring, and evaluation of learning outcomes—activities that are central to metacognitive theory (Flavell, 1979; Schraw & Dennison, 1994).

The proposed mechanism underlying these observed trends is that the THINK module activates students' metacognitive knowledge (declarative, procedural, and conditional knowledge) and metacognitive regulation (planning, monitoring, and evaluation) through explicit instructional strategies. Research has consistently shown that metacognitive skills are teachable and that explicit instruction in metacognitive strategies can enhance students' awareness of their cognitive processes (Bao et al., 2024; Hacker et al., 2009). The THINK module operationalizes this by having students (1) explicitly identify their learning goals, (2) articulate strategies they will use to achieve those goals, (3) monitor their progress during learning activities, and (4) evaluate their understanding after completing tasks. This structured approach is more systematic than the implicit metacognitive development that might occur in conventional classroom instruction, where metacognitive strategies are often taught indirectly or assumed to develop naturally (Veenman et al., 2006).

Importantly, the timing of the THINK module—positioned immediately after the MOVE module—was theoretically motivated by research indicating that physical activity enhances executive functions and cognitive flexibility, creating a neurobiological state that is conducive to metacognitive processing (Hillman et al., 2020; Donnelly et al., 2016). The acute effects of exercise on brain function, including increased cerebral blood flow, elevated neurotransmitter levels (e.g., dopamine, serotonin, norepinephrine), and enhanced neuroplasticity, are well-documented (Heinze et al., 2021; Ratey & Hagerman, 2008). By sequencing the THINK module after physical activity, the model aims to capitalize on this neurobiological state, potentially amplifying the effects of metacognitive instruction beyond what might be achieved through cognitive training alone.

Proposed Mechanisms of the RISE Module

The descriptive trend showing an increased proportion of students classified as having high academic resilience (from 50.0% to 80.0%) suggests that the RISE module may hold promise for developing students' psychological resilience. The RISE module incorporated

simple psychoeducational techniques including positive self-talk, breathing exercises, and stress-management strategies that are grounded in cognitive-behavioral theory and stress-coping frameworks (Lazarus & Folkman, 1984; Meichenbaum, 1985).

The proposed mechanism underlying these observed trends is that the RISE module equips students with concrete, practical strategies for managing academic stress and adversity. Positive self-talk interventions are theorized to work by restructuring maladaptive cognitive patterns, replacing negative or catastrophizing thoughts with more realistic and constructive self-statements (Hatzigeorgiadis et al., 2011; Tod et al., 2011). Breathing exercises and relaxation techniques are thought to activate the parasympathetic nervous system, reducing physiological stress responses (e.g., decreased heart rate, reduced cortisol levels) and promoting a state of calm that supports emotional regulation and adaptive coping (Jerath et al., 2006; Brown & Gerbarg, 2005). By providing students with these accessible techniques, the RISE module aims to enhance their ability to cope with academic challenges, persist in the face of difficulty, and seek help when needed all of which are core components of academic resilience (Cassidy, 2016; Masten, 2014).

The positioning of the RISE module as the final component in the session sequence (MOVE → THINK → RISE) was theoretically motivated by research suggesting that emotional regulation and psychological resilience are enhanced when individuals are in a calm, focused state that can be achieved through prior physical activity and cognitive engagement (Pandey, 2025; Pruzhinin et al., 2026). By consolidating the session with resilience techniques, the model aims to reinforce positive psychological states and provide students with tools they can apply in their academic lives beyond the intervention session.

Comparison with Previous Interventions

The descriptive trends observed in this pilot trial are broadly consistent with findings from previous intervention studies, although direct comparisons should be made with caution due to the preliminary nature of the current findings. The observed trend toward improved physical fitness (VO2Max) is consistent with studies reporting positive effects of school-based physical activity interventions on cardiorespiratory fitness (Hillman et al., 2020; Reigal et al., 2020). However, those studies typically involved multiple sessions over extended periods (weeks to months), whereas the current trial involved only a single session, making the observed trend particularly notable though it may also reflect the acute physiological response to activity and/or the novelty effect rather than sustained fitness improvements.

The observed trend toward improved metacognitive ability is consistent with meta-analytic evidence demonstrating the effectiveness of metacognitive training programs for enhancing students' awareness of their learning processes (Bao et al., 2024; Hacker et al., 2009). However, the current trial differs from those studies in that metacognitive instruction was embedded within a broader integrated intervention rather than delivered as a standalone cognitive training program. This integration represents a potential innovation of the proposed model, as it positions metacognitive development within the context of physical activity and resilience building rather than as an isolated cognitive exercise.

The observed trend toward improved academic resilience is consistent with research on holistic educational interventions integrating physical, cognitive, and psychological components (Pandey, 2025; Purnomo et al., 2026; Rahayu et al., 2024). However, the current trial differs from those studies in that the resilience component (RISE) was delivered as a brief, integrated module within a single session, rather than as a comprehensive, multi-session program. The observed trends suggest that even brief psychoeducational techniques, when delivered in a sequenced, integrated format, may show promise for influencing resilience outcomes-although this interpretation requires confirmation through controlled research.

Limitations of Interpretation

Several factors limit the interpretation of these descriptive trends and the hypothesized mechanisms. First, the single-group design without a control group means that the observed descriptive improvements cannot be attributed to the intervention or to any specific component (MOVE, THINK, or RISE). The improvements may have resulted from maturation, repeated testing, normal school activities, motivation, novelty effects, or the Hawthorne effect (awareness of being observed). Second, the short duration of the trial (single 45-minute session) means that the observed trends reflect immediate or acute effects rather than sustained or cumulative effects. Third, the small sample size ($n = 30$) from a single school limits generalizability. Fourth, the reliance on self-report measures (for metacognition, resilience, and active interaction) may have been influenced by social desirability bias or response bias. The physical fitness measure, while performance-based, may have been influenced by student effort and motivation.

In conclusion, the descriptive trends observed in this limited pilot trial provide a foundation for hypothesizing about the potential mechanisms through which the MOVE, THINK, and RISE components might influence student outcomes. The hypothesized mechanisms are grounded in established theoretical frameworks physiological adaptations, embodied cognition, metacognitive theory, stress-coping theory, and cognitive-behavioral principles and the sequential design of the model (MOVE $\rightarrow$ THINK $\rightarrow$ RISE) is theoretically motivated by the goal of creating synergistic effects across the three components. However, these hypothesized mechanisms require empirical testing through rigorous controlled studies with larger samples, appropriate inferential statistical analyses, and assessment of sustained effects. The current findings should be viewed as hypothesis-generating rather than hypothesis-confirming, and they provide a rationale for the next phase of this research agenda..

Conclusions

Based on the results and discussion of this developmental study, the following conclusions can be drawn within the context of the research objectives and the partial implementation of the 4D development model.

First, the needs analysis revealed that teachers and students perceived a need for an integrated intervention model addressing physical, cognitive, and psychological aspects holistically. The needs analysis, guided by the NLW (Need-Lack-Want) framework ([Hutchinson & Waters, 1987](#)), identified three main findings: high perceived need (82%) for a model integrating bio-psycho-physical aspects; recognition of a gap (76%) in the effectiveness of conventional approaches that address these domains separately; and strong desire (85%) for interactive features such as simulations, games, and structured physical activities. These findings justified the development of a prototype model that addresses the identified needs.

Second, the prototype development was conducted through a partial implementation of the 4D instructional model ([Thiagarajan, Semmel, & Semmel, 1974](#)), specifically limited to the Define, Design, and Develop phases. The Disseminate phase was deliberately excluded from the scope of this study because it requires broader field testing across multiple schools, wider product distribution, and extended timelines beyond the capacity of this initial developmental research. Furthermore, systematic dissemination should only occur after the prototype's effectiveness has been empirically established through controlled experimental research. Therefore, the dissemination activities mentioned in this manuscript constitute planned future strategies rather than completed activities or evaluation results.

Third, the expert-based content validation confirmed that the prototype demonstrated high levels of expert-perceived validity. Based on assessments by three experts from complementary disciplinary backgrounds (Physical Education, Educational Psychology, and Learning Technology), the prototype achieved an overall mean validation score of 4.2 out of 5

(84% feasibility), falling within the "high" category. This conclusion is supported by substantial inter-rater agreement (Fleiss' Kappa $\kappa = 0.72$; percentage agreement = 81.5%) and acceptable content validity indices (S-CVI/Ave = 0.91). The prototype was also rated as practical by teachers and students, with an overall implementation rate of 85% based on teacher ratings (60%), student ratings (30%), and researcher observation (10%). However, it is important to emphasize that these findings represent expert-based content validity and perceived practicality, which are distinct from empirical effectiveness. Expert-based validation assesses whether the prototype's content, design, and theoretical foundations are appropriate based on professional judgment, but it does not establish that the intervention produces measurable improvements in student outcomes.

Fourth, the limited pilot trial with 30 students without a control group showed descriptive trends across all indicators. However, these findings must be interpreted with caution due to several methodological limitations. The absence of a control group means the observed changes cannot be attributed to the intervention or to any specific component (MOVE, THINK, or RISE). Alternative explanations—including maturation effects, repeated testing effects, normal school activities, student motivation, novelty effects, social desirability bias, and the Hawthorne effect—may have contributed to the observed patterns. Furthermore, no inferential statistical tests were conducted, so the descriptive trends should not be interpreted as statistically significant effects or as evidence of the intervention's effectiveness. The primary contribution of this pilot trial is to provide preliminary descriptive data that generate hypotheses for future controlled research, not to establish the model's effectiveness.

Fifth, the research outputs include a prototype consisting of: (1) a teacher's guidebook containing theoretical foundations, step-by-step session guidelines, and facilitation strategies; (2) student activity modules for the MOVE, THINK, and RISE components; and (3) initial assessment instruments for measuring physical fitness, metacognitive ability, academic resilience, and active interaction. This prototype represents the culmination of the Define, Design, and Develop phases and is now ready for further rigorous testing through controlled experimental studies.

Operationalization and Measurement of Outcome Variables

To ensure the credibility of the reported descriptive trends, the following section details how each outcome variable was operationalized and measured in this study. The same instruments and procedures were used at both pretest and posttest assessment.

Physical Fitness (VO2Max): Assessed using the Multistage Fitness Test (MFT) or 20-meter Shuttle Run Test (Leger et al., 1988), a validated field-based measure of cardiorespiratory fitness. The test requires students to run between two markers 20 meters apart, pacing to audio signals that increase progressively. The score was recorded as the total number of shuttles completed and converted to estimated VO2Max (ml/kg/min) using the formula: $VO2Max = 31.025 + 3.238 \times (\text{speed at last completed stage}) - 3.248 \times (\text{age in years}) + 0.1536 \times (\text{speed} \times \text{age})$. Cut-off points for "sufficient" cardiorespiratory fitness were based on established norms for Indonesian adolescents (Kemdikbud, 2018; Cooper Institute, 2016): ≥ 40.0 ml/kg/min for males and ≥ 35.0 ml/kg/min for females. The MFT has demonstrated acceptable test-retest reliability ($r = 0.89$) and criterion validity ($r = 0.92$) in adolescent populations (Leger et al., 1988; Mayorga-Vega et al., 2015).

Metacognitive Ability: Assessed using the Metacognitive Awareness Inventory (MAI) adapted from Schraw and Dennison (1994), consisting of 12 items relevant to the senior high school context. The instrument assessed two components: Knowledge of Cognition (8 items) and Regulation of Cognition (4 items). Students responded on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). Total scores ranged from 12 to 60. Cut-off points for classification as "high" metacognitive ability (≥ 42 , $\geq 70\%$ of maximum score) were based on established criteria for Indonesian senior high school students (Widodo et al., 2020; Saputra et

al., 2021). The instrument demonstrated acceptable internal consistency (Cronbach's $\alpha = 0.84$) in the validation sample.

Academic Resilience: Assessed using the Academic Resilience Scale (ARS) adapted from Cassidy (2016), consisting of 15 items across three dimensions: Persistence (5 items), Adaptability (5 items), and Help-Seeking (5 items). Students responded on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). Total scores ranged from 15 to 75. Cut-off points for classification as "high" academic resilience (≥ 53 , $\geq 70\%$ of maximum score) were based on established criteria for Indonesian adolescents (Rahayu et al., 2023; Fitriani et al., 2022). The instrument demonstrated acceptable internal consistency (Cronbach's $\alpha = 0.87$) in the validation sample.

Active Interaction in Learning: Assessed through a combination of researcher observation using a structured observation checklist (8 behavioral indicators) and student self-report questionnaire (8 parallel items on a 5-point Likert scale). Students were classified as "actively interacting" if they met: (a) observation score ≥ 24 out of 40 ($\geq 60\%$), OR (b) self-report score ≥ 24 out of 40 ($\geq 60\%$), with at least one of the two meeting the threshold. The observation checklist was developed based on established indicators of active learning (Bonwell & Eison, 1991; Prince, 2004) and reviewed by experts for content validity.

Overall Interpretation

In conclusion, this developmental study has successfully produced an initial prototype of the "Moving to Learn" Intervention Model through the Define, Design, and Develop phases of the 4D model. The prototype has demonstrated expert-perceived content validity and perceived practicality based on limited user testing, providing preliminary evidence of its feasibility and potential promise. However, the current study does not establish the effectiveness of the intervention on physical fitness, metacognition, academic resilience, or active interaction. The descriptive trends observed in the limited pilot trial are preliminary and hypothesis-generating rather than hypothesis-confirming. The absence of a control group, the small sample size, the short duration of the trial, and the lack of inferential statistical testing mean that the observed changes cannot be attributed to the intervention or to any specific component of the model.

The prototype is now ready for the next phase of this research agenda: controlled effectiveness testing using a randomized controlled trial or quasi-experimental design with a larger sample, a control group or comparison condition, appropriate inferential statistical analyses (e.g., paired-sample t-tests, ANCOVA, repeated measures ANOVA), and longer follow-up to assess sustained effects. Only after such rigorous testing can the model's effectiveness be empirically established. Subsequent to that, the Disseminate phase of the 4D model can be fully implemented, involving broader field testing across multiple schools, product distribution to educational stakeholders, and evaluation of implementation outcomes in diverse contexts.

The primary contribution of this study is therefore the development and initial validation of a theoretically grounded, practically packaged prototype that provides a foundation for future effectiveness research. The model's novelty lies in its systematic integration of three biopsychophysical components (MOVE, THINK, RISE) within a single structured session, its grounding in established theoretical frameworks (NLW and TPACK), and its practical packaging as a teacher-friendly guidebook. These features distinguish it from previous fragmented interventions, but its effectiveness remains to be established through future controlled research.

Acknowledgment

The authors express their deepest gratitude and appreciation to the Directorate of Research and Community Service (BIMA) of the Ministry of Education, Culture, Research and Technology for funding support through the Beginner Lecturer Research Grant Program in 2026, which enabled this research to be conducted well. The authors also extend their gratitude to STKIP Muhammadiyah Sungai Penuh for the facilities, academic support, and conducive research environment. Special appreciation is conveyed to the Physical Education, Health and Recreation teachers and students who participated as respondents and provided valuable contributions during the research process.

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

- Akker, J. van den. (1999). *Principles and Methods of Development Research*. Kluwer Academic Publisher.
- Bailey, R., Armour, K., Kirk, D., Jess, M., Pickup, I., & Sandford, R. (2009). The educational benefits claimed for physical education and school sport: An academic review. *Research Papers in Education*, 24(1), 1–27. <https://doi.org/10.1080/02671520701809817>
- Bao, L., Soh, K. G., Nasiruddin, N. J. M., Xie, H., & Zhang, J. (2024). Unveiling the impact of metacognition on academic achievement in physical education and activity settings: A comprehensive systematic review and meta-analysis of qualitative insights. *Psychology Research and Behavior Management*, 17, 973–987. <https://doi.org/10.2147/PRBM.S444631>
- Bugge, A., & al., et. (2022). ACTIVE SCHOOL - Impact of new methods to increase physical activity in school-aged children on academic achievement. In *ClinicalTrials.gov*.
- Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, 44(4), 237–251. <https://doi.org/10.3102/0013189X15584327>
- Fleiss, J. L. (1971). Measuring Nominal Scale Agreement among Many Raters. *Psychological Bulletin*, 76, 378–382. <https://doi.org/http://dx.doi.org/10.1037/h0031619>
- Fullan, M., & Langworthy, M. (2014). *A Rich Seam: How New Pedagogies Find Deep Learning*. Pearson.
- Haverkamp, B. F., Wiersma, R., Vertessen, K., van Ewijk, H., Oosterlaan, J., & Hartman, E. (2020). Effects of physical activity interventions on cognitive outcomes and academic performance in adolescents and young adults: A meta-analysis. *Journal of Sports Sciences*, 38(23), 2637–2660. <https://doi.org/10.1080/02640414.2020.1794763>
- Heinze, K., Cumming, J., Dosanjh, A., Palin, S., Poulton, S., Bagshaw, A. P., & Broome, M. R. (2021). Neurobiological evidence of longer-term physical activity interventions on mental health outcomes and cognition in young people: A systematic review of randomised controlled trials. *Neuroscience & Biobehavioral Reviews*, 120, 431–441. <https://doi.org/10.1016/j.neubiorev.2020.10.014>
- Hillman, C. H., & al., et. (2020). A review of the effects of physical activity on cognition and brain health across children and adolescence. *Nestlé Nutrition Institute Workshop Series*, 95, 113–124. <https://doi.org/10.1159/000511508>
- Hutchinson, T., & Waters, A. (1987). *English for Specific Purposes: A Learning-Centred Approach*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511733031>
- Kemdikbud. (2017). *Panduan Penguatan Pendidikan Karakter pada Satuan Pendidikan Formal*. Kementerian Pendidikan dan Kebudayaan.

- Landis, J.R. and Koch, G. G. (1977). An Application of Hierarchical Kappa-Type Statistics in the Assessment of Majority Agreement among Multiple Observers. *Biometrics*, 33, 363–374. <https://doi.org/https://doi.org/10.2307/2529786>
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382–385. <https://www.ovid.com/jnls/nursingresearchonline/abstract/00006199-198611000-00017~determination-and-quantification-of-content-validity?redirectionsource=fulltextview>.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook* (2nd ed.). SAGE Publications. <https://doi.org/10.1177/109821409902000122>
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- Nieveen, N. (1999). *Prototyping to Reach Product Quality*. Kluwer Academic Publisher.
- Pandey, M. (2025). Effects of yoga and physical exercise on adolescent development and learning: A randomized trial. *Improving Schools*, 27(2–3), 155–178. <https://doi.org/10.1177/13654802251403889>
- Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, 29(5), 489–497. <https://doi.org/https://doi.org/10.1002/nur.20147>
- Pruzhinin, K. N., Pruzhinina, M. V, Kolesov, V. I., & Kuramshin, Y. F. (2026). Development of stress resilience in high school students through physical culture and sports. *Teoriya i Praktika Fizicheskoy Kultury*, 1, 42–48. <https://doi.org/10.5930/1994-4683-2026-1-42-48>
- Purnomo, E., Jermaina, N., & Md Nadzalan, A. (2026). Sports as a means of mental education to improve the student resilience. *The Paris Conference on Education 2026*. <https://doi.org/10.55860/5vantg56>
- Rahayu, W., & al., et. (2024). Integration of mental education through physical education to improve adolescent personal resilience. *Studia Sportiva*, 18(1), 76–89. <https://doi.org/10.5817/StS2025-2-13>
- Reigal, R. E., Hernández-Mendo, A., Juárez-Ruiz de Mier, R., & Morales-Sánchez, V. (2020). Physical exercise and fitness level are related to cognitive and psychosocial functioning in adolescents. *Frontiers in Psychology*, 11, 1814. <https://doi.org/10.3389/fpsyg.2020.01777>
- Robinson, L. E., Riley, N., Owen, K. B., Drew, R., Mavilidi, M. F., Hillman, C. H., & Lubans, D. R. (2023). Effects of resistance training on academic outcomes in school-aged youth: A systematic review and meta-analysis. *Sports Medicine*, 53(11), 2125–2141. <https://doi.org/10.1007/s40279-023-01881-6>
- Rosenberg, J. M., & Koehler, M. J. (2015). Context and technological pedagogical content knowledge (TPACK): A systematic review. *Journal of Research on Technology in Education*, 47(3), 186–210. <https://doi.org/10.1080/15391523.2015.1052663>
- Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, 27(2), 94–104. <https://doi.org/10.1093/swr/27.2.94>
- Sember, V., Jurak, G., Kovac, M., Morrison, S. A., & Starc, G. (2020). Children's physical activity, academic performance, and cognitive functioning: A systematic review and meta-analysis. *Frontiers in Public Health*, 8, 307. <https://doi.org/10.3389/fpubh.2020.00307>
- Sibbick, E., Boat, R., Sarkar, M., Johnston, J. P., Groom, M., Williams, R. A., & Cooper, S. B. (2024). Associations of physical activity and cardiorespiratory fitness with cognitive function, self-control, and resilience in young people with attention deficit hyperactivity disorder. *Advanced Exercise and Health Science*, 1(1), 51–58. <https://doi.org/10.1016/j.aehs.2024.01.003>
- Stockwell, S., & al., et. (2021). Changes in physical activity and sedentary behaviours from

- before to during the COVID-19 pandemic lockdown: A systematic review. *BMJ Open Sport & Exercise Medicine*, 7(1). <https://doi.org/10.1136/bmjsem-2020-000960>
- Taniguchi, K., Maki, N., Sakamoto, H., Inuta, T., Kim, B., Oh, S., & Mayers, T. (2024). Metacognition-associated factors in physical and occupational therapy students: A cross-sectional study. *Brain Sciences*, 14(10), 1041. <https://doi.org/10.3390/brainsci14101041>
- Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974). *Instructional Development for Training Teachers of Exceptional Children: A Sourcebook*. Indiana University.
- Tomlinson, B. (2003). Developing principled frameworks for materials development. In *Developing Materials for Language Teaching*. Continuum. <https://www.scirp.org/reference/referencespapers?referenceid=1524560>
- WHO. (2020). *WHO guidelines on physical activity and sedentary behaviour*. World Health Organization. <https://doi.org/10.1136/bjsports-2020-102955>
- Yi, J., & al., et. (2024). Declining physical activity levels in children. *International Journal of Behavioral Nutrition and Physical Activity*, 21(1), 1–12. <https://doi.org/10.3389/fpubh.2025.1636891>

Information about the authors:

Romiya Heliza, S.Pd., M.Pd: romiyaheliza19@gmail.com, <https://sandbox.orcid.org/0009-0004-8741-2207>, Department of Physical Education, Health and Recreation, STKIP Muhammadiyah Sungai Penuh, Indonesia.

Febriady Huta Uruk, S.Pd., M.Pd: febriadyhuta2@gmail.com, <https://sandbox.orcid.org/0009-0008-1610-5970>, Department of Physical Education, Health and Recreation, STKIP Muhammadiyah Sungai Penuh, Indonesia.

Dr. Efriana Jon, S.Pd., M.Pd: efrianajon98@gmail.com, <https://sandbox.orcid.org/0009-0003-3050-9668>, Department of Physical Education, Health and Recreation, STKIP Muhammadiyah Sungai Penuh, Indonesia.

Nandia Pitri, S.Pd., M.Hum: nandpitri@gmail.com, <https://orcid.org/0000-0001-6043-9611>, Department of Indonesian Language Education, STKIP Muhammadiyah Sungai Penuh, Indonesia.

Cite this article as:

Heliza, R., Huta Uruk, F., Jon, E., & Pitri, Nandia. (2026). Development of "Moving to Learn" Intervention Model to Improve Physical Fitness, Metacognitive Ability, and Academic Resilience of Senior High School Students. *Indonesian Journal of Physical Education and Sport Science (IJPESS)*, 6(3), 571-601. <https://doi.org/10.52188/ijpess.v6i3.2352>