International Journal of Human Movement and Sports Sciences 13(6): 1182-1193, 2025 http://www.hrpub.org
DOI: 10.13189/saj.2025.130602

Development of a Sensory Integration-Based Assessment System for Fundamental Motor Skills in Children Aged 5–6 Years

Wei Huang, Watthanapong Khongsuebsor*

Department of Health and Sport Science, Faculty of Education, Mahasarakham University, Thailand

Received July 12, 2025; Revised September 24, 2025; Accepted October 14, 2025

Cite This Paper in the Following Citation Styles

(a): *[1] Wei Huang, Watthanapong Khongsuebsor , "Development of a Sensory Integration-Based Assessment System for Fundamental Motor Skills in Children Aged 5–6 Years," International Journal of Human Movement and Sports Sciences, Vol. 13, No. 6, pp. 1182 - 1193, 2025. DOI: 10.13189/saj.2025.130602.*

(b): *Wei Huang, Watthanapong Khongsuebsor (2025). Development of a Sensory Integration-Based Assessment System for Fundamental Motor Skills in Children Aged 5–6 Years. International Journal of Human Movement and Sports Sciences, 13(6), 1182 - 1193. DOI: 10.13189/saj.2025.130602.*

Abstract This study aimed to develop an evaluation system for assessing gross motor development in children aged 5–6 years based on sensory integration theory. Early identification of motor skill delays is essential for guiding timely interventions in young children. Using the Delphi method across three rounds with 12 experts, 10 key indicators were systematically identified and refined. Pretesting and retesting with 450 children from licensed kindergartens across 14 cities in Guangxi Province, China, demonstrated strong content validity (IOC = 0.80–1.00) and high reliability (Spearman's r = 0.790–0.932, $p < 0.05$). Revisions to the indicators were made based on expert feedback and test results to ensure clarity, relevance, and practical applicability. The resulting tool offers a reliable and valid framework for educators and therapists to detect developmental delays early and design targeted interventions. Beyond its current application, the model highlights potential for extension to younger children (ages 3–4), cross-cultural testing, and adaptation into digital or app-based formats to enhance accessibility and longitudinal tracking. This research contributes to early childhood education and therapy by embedding sensory integration principles into a comprehensive, multidimensional approach to motor skill assessment.

Keywords Fundamental Motor Skills, Sensory Integration, Early Childhood Assessment, Delphi Method, Physical Education

1. Introduction

Fundamental motor skills are critical for the physical, cognitive, psychological, and social development of young children. Early childhood represents a crucial period for establishing a foundation in physical capabilities, which significantly impacts later motor skill acquisition and overall growth. In recent years, China's rapid economic and social development, alongside improved living conditions, has raised concerns about the declining physical health of young children [1]. Evidence suggests a continuous decline in children's physical fitness, with early onset trends manifesting at increasingly younger ages. Despite growing recognition of the importance of sensory integration in motor development, few assessment systems comprehensively incorporate sensory integration dimensions. Thus, there is a pressing need for a scientifically validated, comprehensive assessment tool that integrates both fundamental motor skills and sensory processing abilities to better understand and support young children's development.

Despite the recognized importance of motor development, particularly gross motor skills in holistic child development, traditional assessment tools remain limited, often due to cultural incongruities and insufficient comprehensiveness. Sensory integration theory [2] emphasizes the pivotal roles of the vestibular, proprioceptive, and tactile systems in motor development. However, few existing evaluation models adequately

incorporate these sensory components. Addressing this gap, the present study aims to develop a culturally adapted model for evaluating gross motor development through the lens of sensory integration.

In China, approximately 5% of the 1.4 billion population are children aged 3–6 years [3]. As the future workforce and drivers of national development, the growth and well-being of young children are of strategic importance. Recognizing this, the Chinese government has introduced policies emphasizing early childhood development. Key initiatives include the "Regulations on Kindergarten Work" (1996) [4], the "Guide to Learning and Development for Children Aged 3–6" (2012) [5], and the "Outline for the Construction of a Sports Power" (2019) [6], all of which underscore the significance of structured physical development during early years. As a result, Chinese parents increasingly prioritize their children's physical and motor development.

Children aged 5–6 years are in a transitional phase preparing for primary school, coinciding with sensitive periods of gross motor skill, sensory integration, and neuromuscular control development [7]. However, the lack of formally trained PE teachers in many kindergartens hinders both the guidance of physical activities and the accurate assessment of children's motor skills [8]. Commonly used assessment tools, such as the "National Physical Fitness Measurement Standard Manual (Children's Section)," focus primarily on performance metrics such as the 10-meter shuttle run, standing long jump, and balance beam walking, while neglecting broader developmental dimensions.

Gross motor development is intricately linked to physical, cognitive, and psychological outcomes. It involves the coordination of large muscle groups and is one of the earliest acquired motor skills in childhood. According to Ayres [9], sensory integration in the brain's process of organizing and interpreting sensory input is fundamental to producing effective motor responses. Luo Yijun [10] further emphasized that movement itself constitutes an act of sensory integration, requiring simultaneous engagement across multiple sensory systems. Strengthening vestibular, proprioceptive, and tactile experiences can stimulate neural development and promote overall well-being.

Historically, gross motor assessment tools in China were often adapted from Western instruments without establishing localized norms, leading to cultural mismatches and reduced validity. International tools such as the Test of Gross Motor Development (TGMD), Peabody Developmental Motor Scales, and the Children's Motor Assessment Scale typically feature broad age ranges, extensive indicators, and high testing demands, rendering them impractical for widespread use with Chinese children aged 5–6 years.

To address the challenges posed by the declining physical fitness of young children and the lack of comprehensive assessment tools integrating sensory processing, this study aims to develop an assessment system for fundamental motor skills in children aged 5–6 years based on sensory integration principles. The proposed model is designed to establish localized norms, simplify assessment procedures, enhance accessibility, and promote widespread application. By advancing the evaluation and development of gross motor skills among preschool children, the model seeks to improve comprehensiveness, effectiveness, and scientific validity of motor skill assessments. Furthermore, it enables accurate identification of current developmental statuses and gaps, providing a theoretical foundation for preschools and related institutions to design targeted physical and psychological development programs. Ultimately, the model contributes to the broader goal of enhancing public health through early childhood interventions.

2. Materials and Methods

2.1. Research Design

This research is a developmental study employing a multi-phase approach, integrating expert consensus and empirical validation grounded in sensory integration theory. The three-round Delphi method was conducted to identify, refine, and finalize indicators for evaluating gross motor development in children aged 5–6 years. The Delphi method, a structured communication technique, facilitates consensus among experts through iterative rounds of questionnaires with controlled feedback [11], [12]. It is especially effective in fields with limited empirical data and where expert judgment is essential [13], [14].

Following the Delphi rounds, pretesting and retesting of the assessment items were performed to ensure clarity, feasibility, and reliability of the indicators. Content validity was evaluated using the Item-Objective Congruence (IOC) method, with a minimum threshold of 0.70 [15]. Revisions to the indicators were made iteratively based on expert feedback and preliminary test results to enhance precision and contextual appropriateness.

2.2. Research Participants

A total of 450 children aged 5 to 6 years participated in the study. These children were recruited from licensed preschools in 14 cities across Guangxi Province, China. The inclusion criteria required that participants display typical developmental characteristics and be currently enrolled in government-approved institutions. Written informed consent was obtained from parents or legal guardians. This study was approved by the Mahasarakham University Ethics Committee (Approval number: 555-460/2024).

Additionally, 12 experts with backgrounds in early childhood education, kinesiology, pediatric rehabilitation, and sensory integration were selected for the Delphi panel.

These individuals were selected based on their qualifications, practical experience, and academic contributions in relevant fields, ensuring a comprehensive and credible consensus process.

2.3. Procedures and Instruments

The study was conducted in several stages, beginning with the Delphi method to establish expert consensus.

1. Stage 1: First Round of the Delphi Process

A literature review was conducted to design an initial set of 3 primary indicators and 50 secondary indicators based on established tools such as the Peabody Developmental Motor Scale and the Gross Motor Function Measure. Experts evaluated the importance of each indicator and were invited to provide open-ended feedback and suggest additional items.

2. Stage 2: Second Round of the Delphi Process

Indicators that received low mean scores or showed disagreement among experts were removed. New indicators proposed by the panel were incorporated. A revised questionnaire was sent to the same experts, accompanied by a summary of Round 1 results to support informed reevaluation.

3. Stage 3: Final Round of the Delphi Process

The final survey included the refined list of indicators. Experts reviewed and confirmed the final set, leading to consensus on 10 core indicators that represent key dimensions of sensory integration and gross motor development.

Following the Delphi stages, pretesting and retesting were carried out with the child participants to evaluate item clarity, comprehensibility, and temporal reliability. Content validity of the instrument was assessed using the Item-Objective Congruence (IOC) index (Keeney et al., 2011). Indicators meeting or exceeding the IOC threshold of 0.70 were retained.

The final set of 10 core indicators and their scoring criteria is presented in Appendix A, while the relationships among indicators, observed behaviors, and sensory integration components are summarized in Appendix B.

2.4. Data Analysis

Quantitative analyses were used to evaluate both expert consensus and instrument reliability. Kendall's coefficient of concordance (W) was calculated to assess the degree of agreement among the 12 experts across the three Delphi rounds [11], [13]. This non-parametric statistic ranges from 0 (no agreement) to 1 (perfect agreement), with higher values indicating stronger consensus.

Test-retest reliability was evaluated using both Pearson's and Spearman's correlation coefficients to determine the consistency and stability of the children's performance over time [14]. Pearson's correlation was used to assess linear relationships, while Spearman's rank-order correlation was used for monotonic trends in ordinal data.

Indicators with low mean scores or high variability were eliminated, following best practices in Delphi-based instrument development to ensure precision, validity, and applicability [15].

3. Results

1. Stage 1: Delphi method First round

First-round expert questionnaire: The researchers distributed a preliminary selection criteria framework, based on the literature review and practical experience, to relevant experts in the field in the form of a questionnaire. The experts evaluated and screened the indicators based on their theoretical and practical knowledge. Each expert was asked to rate the relevance and importance of each criterion and provide their professional opinions.

The first round of the Delphi method was implemented in September 2024, and 12 questionnaires were issued. And a total of 12 questionnaires were recovered with a recovery rate of 100%, of which 12 were valid with an effective recovery rate of 100%

As illustrated in Table 1, it summarizes the reliability analysis from the first round of the Delphi process. A panel of 12 experts assessed 50 evaluation items. The Kendall's coefficient of concordance (W) was calculated to be 0.418, indicating a moderate to substantial level of agreement among the experts. Furthermore, the chi-square (χ^2) statistic was 245.959, demonstrating that the level of agreement was statistically significant ($p < 0.05$). These results suggest a satisfactory degree of consensus among the expert panel, supporting the reliability and validity of the data collected during this round for further analysis in subsequent Delphi iterations.

Table 1. Specific data of Column/Row

Kendall W			
Number of experts	Number of evaluation items	Kendall	χ^2
12	50	0.418	245.959

As shown in Table 2, the importance scores from the expert questionnaires were used for data processing. During the initial screening of indicators, items with an average score of less than 3.5 were eliminated. A total of 21 indicators were removed, including C26, C28, L6, L15, C24, B43, L16, C30, B46, B37, L10, C21, C29, B44, C33, C32, C34, L2, C31 (microcomputer), C35, and L8. Consequently, 29 indicators remained for further evaluation. In the second round, experts were asked to reassess the retained indicators in terms of their importance, comprehensiveness, stability, safety, and effectiveness.

Table 2. The first round of expert indicators screening results

Item	AVG	SD	CV	CP
L19	5	0	0	100
C27	4.833	0.577	0.119	91.7
L9	4.667	0.651	0.14	75
L12	4.667	0.492	0.106	66.7
C22	4.667	0.492	0.106	66.7
C25	4.667	0.492	0.106	66.7
B36	4.667	0.492	0.106	66.7
B42	4.667	0.492	0.106	66.7
L4	4.583	0.515	0.112	58.3
B39	4.583	0.515	0.112	58.3
L1	4.5	0.674	0.15	58.3
L17	4.5	0.674	0.15	58.3
B49	4.5	0.674	0.15	58.3
B50	4.5	0.674	0.15	58.3
C23	4.417	0.9	0.204	66.7
B47	4.417	0.669	0.151	50
L7	4.25	0.866	0.204	50
L20	4.25	0.965	0.227	58.3
B45	4.167	1.03	0.247	58.3
L18	4.083	0.9	0.22	41.7
L5	3.833	0.937	0.245	33.3
B40	3.667	1.231	0.336	33.3
L13	3.583	1.165	0.325	25
B41	3.583	0.793	0.221	16.7
B48	3.583	0.793	0.221	0
L3	3.5	1.314	0.376	33.3
L11	3.5	0.522	0.149	0
L14	3.5	1.314	0.376	33.3
B38	3.5	1.382	0.395	41.7
C26	3.417	1.24	0.363	25
C28	3.333	1.303	0.391	25
L6	3.25	0.866	0.266	0
L15	3.25	1.357	0.417	33.3
C24	3.25	1.288	0.396	25
B43	3.25	1.138	0.35	16.7
L16	3.167	1.267	0.4	25
C30	3.167	0.718	0.227	0
B46	3.167	1.193	0.377	16.7
B37	3.083	0.996	0.323	0
L10	3	1.348	0.449	25

Table 2 continued

C21	3	1.044	0.348	16.7
C29	3	0	0	0
B44	3	1.206	0.402	16.7
C33	2.917	1.084	0.372	16.7
C32	2.833	0.389	0.137	0
C34	2.75	0.965	0.351	0
L2	2.667	0.651	0.244	0
C31	2.583	0.515	0.199	0
C35	2.583	0.515	0.199	0
L8	2.5	0.798	0.319	0

2. Stage 2: Delphi method second round

Conduct an in-depth analysis of the results of the first round of expert feedback, tease out the similarities, differences, and major disagreements in the expert opinions, and remove those that are clearly irrelevant to the topic or unreasonable.

Based on the results of the first round of analysis, the second round of questionnaire is designed. In order to allow experts to re-emphasize the basic comprehensiveness, stability, feasibility, safety, etc., on the key points and problems that emerged in the first round.

The first round of the Delphi method was implemented on October 1, 2024, and 12 questionnaires were issued. And on October 7, 12 questionnaires were recovered with a recovery rate of 100%, of which 12 were valid with an effective recovery rate of 100%

As illustrated in Table 3, the reliability analysis results from the second round of the Delphi method. A total of 12 experts evaluated 29 items. The Kendall's coefficient of concordance (W) was 0.622, indicating a substantial level of agreement among the experts. The chi-square (χ^2) value was 208.965, and the result was statistically significant ($p < 0.05$), suggesting a higher degree of consensus compared to the first round. These findings demonstrate that the level of expert agreement improved following the refinement of evaluation items, thereby enhancing the reliability and validity of the data for subsequent analyses.

Table 3. Delphi method second round - Reliability situation

Kendall W			
Number of experts	Number of evaluation items	Kendall	χ^2
12	29	0.622	208.965

As shown in Table 4, the importance scores obtained from the expert questionnaires were further analyzed. In the indicator screening process, the average score was primarily used as the basis for selection. Indicators with a mean score lower than 4.2 and a coefficient of variation (CV) greater than 0.25 were eliminated. Following this screening, 12 indicators were retained for subsequent pretesting.

Table 4. The second round of expert indicators screening results

Item	AVG	SD	CV	CP
L19	5	0	0	100
C27	4.917	0.289	0.059	91.7
B49	4.917	0.289	0.059	91.7
L9	4.75	0.452	0.095	75
B42	4.75	0.452	0.095	75
C22	4.667	0.492	0.106	66.7
B36	4.667	0.492	0.106	66.7
B45	4.667	0.492	0.106	66.7
L1	4.583	0.515	0.112	58.3
L4	4.583	0.515	0.112	58.3
L12	4.583	0.515	0.112	58.3
L7	4.5	0.522	0.116	50
C25	4.25	0.452	0.106	25
B39	4.167	0.389	0.093	16.7
L14	4.083	0.289	0.071	8.3
L17	4.083	0.289	0.071	8.3
L20	4.083	0.289	0.071	8.3
B41	4.083	0.289	0.071	8.3
B50	4.083	0.289	0.071	8.3
L5	3.833	0.389	0.102	0
L11	3.833	0.389	0.102	0
C23	3.833	0.389	0.102	0
L13	3.75	0.622	0.166	8.3
B40	3.75	0.452	0.121	0
B48	3.75	0.452	0.121	0
L3	3.583	0.515	0.144	0
L18	3.583	0.515	0.144	0
B38	3.25	0.452	0.139	0
B47	3.083	0.289	0.094	0

Table 5 presents the pretest results of the evaluation

indicators, including the mean, standard deviation, and completion rates for male and female participants. These findings demonstrate the feasibility of the selected indicators in the target population.

3. Stage 3: The third round of Delphi

Project completion rates reflect the evaluation of feasibility and applicability of the project, where project completion rate is lower than 60%, that is eligibility is poorer, (= can complete the project completion rate index test/test sample by the total number of 100% the number of participants in this study the target movement performance score of 0 is divided into was not completed index test). It can be seen from the data in the table above that the completion rate of the 10 items is greater than 90%, and the test items can be retained. Single factor analysis of variance was performed to test whether each test index had good sensitivity to age. As shown in the table, there are significant differences in the mean values of the test results of the 8 items between males and females, indicating that the test items have good sensitivity to gender.

As shown in Table 6, the retest results provide the mean, standard deviation, and completion rates for male and female participants. These findings confirm that the selected indicators are stable and practical for children aged 5–6 years.

Table 5. Evaluation index pretest situation

Item	Male				Female			
	Mean ±SD	Min	Max	PC	Mean ±SD	Min	Max	PC
L1	6.2 ±2.2	2.7	14.2	95%	6.3 ±2.5	3	15	95%
L4	1.5 ±0.6	0.6	3	100%	1.2 ±0.4	0.5	2.6	95%
L7	10.1 ±2.5	6.1	15.1	100%	9 ±1.5	5.5	13.2	100%
L9	8.2 ±1.8	6.6	10.8	100%	7.5 ±1.5	5.3	9.6	100%
L12	8.1 ±1.2	5.8	9.4	100%	8.2 ±1.3	6.2	10.3	100%
L19	9.1 ±1.6	4.4	11.4	100%	8.8 ±1.5	4.1	12.3	100%
C22	9.3 ±1.3	5.6	12.5	100%	8.4 ±1.6	4.1	10.8	100%
C27	10 ±2	5	15	100%	9 ±3	4	12	100%
B36	13.5 ±6	4.5	20.5	50%	10 ±5	1	18.5	60%
B42	6.3 ±1.7	2	10.2	100%	6.9 ±1.1	4	12	100%
B45	240 ±60	135	450	50%	180 ±60	90	360	70%
B49	6 ±1	4	8	100%	5 ±1	3	7	100%

Table 6. Evaluate the retesting of indicators

Item	Male				Female			
	Mean ±SD	Min	Max	PC	Mean ±SD	Min	Max	PC
L1	6.3 ±2.2	14.3	3	95%	6.4 ±2.5	15	3.1	100%
L4	1.5 ±0.6	0	3	100%	1.3 ±0.4	0.5	2.6	95%
L7	10.2 ±2.5	6.1	15	100%	9.2 ±1.5	5.5	13.2	100%
L9	8.3 ±1.8	6.7	10.8	100%	7.5 ±1.5	5.4	9.6	100%
L12	8.2 ±1.2	9.4	5.7	100%	8.2 ±1.3	10.2	6.2	100%
L19	9.2 ±1.6	11.3	4.4	100%	8.8 ±1.5	12.3	4.1	100%
C22	9.5 ±1.3	5.8	12.5	100%	8.5 ±1.6	4.2	10.8	100%
C27	10 ±2	6	15	100%	9 ±3	4	12	100%
B36	13.5 ±6	4.5	20.5	50%	10 ±5	2	18.5	60%
B42	6.4 ±1.7	10.2	3.9	100%	6.9 ±1.2	12	4	100%
B45	240 ±60	135	450	100%	180 ±60	90	360	65%
B49	6 ±1	4	8	100%	5 ±1	4	7	100%

Spearman's correlation coefficient was the lowest 0.790 and the highest 0.932, all above 0.70, indicating that the reliability of index raters was good and significant at the level of 0.01. It shows that the evaluation index of motor development ability has good stability and high objectivity.

Conclusion: After completing the second round of the Delphi method, we optimized the research direction and indicator setting based on the feedback and suggestions of the experts. Then, the preliminary experiment was carried out to practice selected indicators. In response to the problems exposed in the pre-experiment, we made adjustments and improvements and then carried out a retest. After two rounds of data collection and analysis, 10 core indicators were finally determined. Prepare the third round of the questionnaire.

Table 7 shows the retest reliability analysis of the evaluation indicators, with all coefficients being significant ($p < 0.05$), thereby confirming the consistency and reliability of the measures.

Table 7. Retest the reliability test of the evaluation index

Items	R	P
L1	0.883*	P<0.05
L4	0.931*	P<0.05
L7	0.912*	P<0.05
L9	0.873*	P<0.05
L12	0.871*	P<0.05
L19	0.913*	P<0.05
C22	0.832*	P<0.05
C27	0.843*	P<0.05
B36	0.932*	P<0.05
B42	0.862*	P<0.05
B45	0.921*	P<0.05
B49	0.790*	P<0.05

The detailed scoring system for these indicators is provided in Appendix A, ensuring consistency in evaluation and interpretation across different raters.

4. Stage 4: The third round of Delphi

Design clear and accurate questions based on the principle of equal weight for the different indicators in the second round and the contents that need to be further discussed. Send it to the experts.

The first round of the Delphi law was implemented on November 17, 2024, and 12 questionnaires were issued. And on November 24, 12 questionnaires were recovered with a recovery rate of 100%, of which 12 were valid with an effective recovery rate of 100%.

The results of the third-round Delphi evaluation are summarized in Table 8. Twelve experts evaluated 12 indicators, yielding a Kendall's W of 0.647 ($\chi^2 = 85.348$), which indicates a high level of consistency and agreement among the experts. Following this, the first third of expert indicators was screened, as shown in Table 9, which presents the mean (AVG), standard deviation (SD), coefficient of variation (CV), and completion percentage (CP) for each item. Most indicators received high scores and consensus levels, whereas indicators B36 and B45 scored 0 across all metrics, reflecting minimal consensus and applicability. Based on these results, B36 and B45 were excluded from further consideration, demonstrating the reliability of expert evaluations and the careful selection of indicators for subsequent analysis.

Table 8. Delphi method third round - Reliability situation

Kendall W			
Number of experts	Number of evaluation items	Kendall	χ^2
12	12	0.647	85.348

Table 9. The first third of the expert indicators screening results

Item	AVG	SD	CV	CP
L19	5	0	0	100
B49	5	0	0	100
C27	4.917	0.289	0.059	91.7
B42	4.833	0.389	0.081	83.3
L9	4.75	0.452	0.095	75
L12	4.75	0.452	0.095	75
L1	4.667	0.492	0.106	66.7
L4	4.667	0.492	0.106	66.7
L7	4.667	0.492	0.106	66.7
C22	4.667	0.492	0.106	66.7
B36	0	0	0	0
B45	0	0	0	0

5. Stage 5: Item-Objective Congruence (IOC)

To validate and examine validity, content validity was established through the Project Objective Concordance Index (IOC) to improve the relevance of the questionnaire items to the term definitions defined in Chapter I. The accuracy of the questionnaire content validity was valid when the IOC standard value was greater than or equal to 0.7. The questionnaire was then revised according to the expert advice and recommendations and returned to the consultant for adjustment before trial. The IOC has made the following considerations:

- + 1 Is a question that the experts all agree to answer
- 0 Represents that the elements and indicator content are not appropriate
- -1 Refers to what the expert does not agree with

The researcher pre-tested the questionnaire to find discriminant and Reli competence values. Index of Conformity (IOC)=0.80-1.00. Item classification competence was found by finding the simple correlation coefficient between the item score and the total score from Pearson's simple correlation coefficient. 0.690 or more, by observing the item-to-item total correlation, the researcher takes the question with a discrimination value and uses Cronbach's alpha coefficient method to find the confidence value of the entire version, with the standard being 0.921 and above.

Experts agree on the advantages of using equal weights.

Simple and easy to implement: Assigning the same weight to each evaluation indicator makes the operation very simple, easy to understand and implement, and enables a quick evaluation of children's gross motor skills.

Avoid subjective bias: It reduces the bias caused by the subjective factors of evaluators when assigning different weights to different indicators, making the evaluation relatively more objective. Avoid subjective biases: It reduces the biases caused by evaluators assigning different weights to different indexes due to their personal subjective factors, making the evaluation relatively more objective.

Table 10 reports the Item-Objective Congruence (IOC) results, showing that all items passed the content validity threshold, thereby supporting the appropriateness of the selected indicators.

6. Stage 6: Summary

At this stage, in order to obtain a highly scientific and reliable index system, we adopted a rigorous and systematic research process.

In the first round of Delphi method, we sent questionnaires to authoritative experts from different fields and extensively collected their initial opinions and suggestions on potential indicators. A total of 90 valid feedbacks were collected and 29 indicators were initially sorted out. In the second round, we fed back the summarized results of the first round to the experts, asking them to evaluate and screen the importance of these indicators, and at the same time add new ideas. After in-depth thinking, the experts put forward many constructive views, and we optimized the indicators according to these feedbacks, retaining 12 indicators. In the third round of the Delphi method, the experts evaluated the adjusted indicators again, further improved the indicator system, and initially identified 10 more concentrated indicators.

Another group of experts in related fields were invited to evaluate the retained indicators to determine whether they fully and accurately reflect the research topic. Experts from the content relevance, integrity and other dimensions of consideration, put forward specific suggestions for modification. According to these opinions, we carried out the final improvement and screening of the indicators, and finally determined 10 core indicators.

Table 10. The IOC results of the questionnaire of the evaluation

NO.	Item	Results
1	Do you think the degree of integration of vestibular sense, proprioception and touch can effectively reflect the level of gross motor development of children when they Step forward on the balance beam?	Pass
2	Do you think the degree of integration of vestibular sense, proprioception and touch can effectively reflect the level of gross motor development of children when they Horizontal bar suspension move?	Pass
3	Do you think the degree of integration of vestibular sense, proprioception and touch can effectively reflect the level of gross motor development of children when they v4 legs crawl forward?	Pass
4	Do you think the degree of integration of vestibular sense, proprioception and touch can effectively reflect the level of gross motor development of children when they Crawl forward?	Pass
5	Do you think the degree of integration of vestibular sense, proprioception and touch can effectively reflect the level of gross motor development of children when they 10 meters back run?	Pass
6	Do you think the degree of integration of vestibular sense, proprioception and touch can effectively reflect the level of gross motor development of children when they Double hop?	Pass
7	Do you think the degree of integration of vestibular sense, proprioception and touch can effectively reflect the level of gross motor development of children when they Overhand throw?	Pass
8	Do you think the degree of integration of vestibular sense, proprioception and touch can effectively reflect the level of gross motor development of children when they Throw the ball with both hands?	Pass
9	Do you think the degree of integration of vestibular sense, proprioception and touch can effectively reflect the level of gross motor development of children when they Moving racket?	Pass
10	Do you think the degree of integration of vestibular sense, proprioception and touch can effectively reflect the level of gross motor development of children when they Side roll?	Pass

Note: The finalized 10 core indicators and their scoring system are presented in Appendix A, with detailed indicator–behavior–sensory integration relationships provided in Appendix B.

4. Discussion

This study aimed to establish evaluation criteria for gross motor development of children aged 5-6 based on sensory integration theory. Through three rigorous rounds of the Delphi Method involving expert consensus, three first-level indicators and ten second-level indicators were developed, encompassing upper body actions, core torso movements, and lower extremity functions. Each indicator was systematically linked to sensory integration components, proprioceptive, vestibular, and tactile input, thereby offering a multidimensional view of motor skill performance.

1. Indicator Refinement through the Delphi Method

Initially, 50 indicators were proposed. Through the third round of expert evaluations involving 12 specialists, the list was narrowed down to 10 indicators. The selection criteria were stringent: indicators with a mean score below 4.2 and a coefficient of variation (CV) exceeding 0.25 were excluded. This approach aligns with established practices in Delphi studies, where mean scores and CVs are pivotal in determining consensus and variability among expert opinions.

The increasing agreement among experts across rounds was quantified using Kendall's coefficient of concordance (W), a non-parametric statistic measuring inter-rater agreement. A higher Kendall's W indicates stronger consensus among experts, validating the Delphi method's effectiveness in achieving reliable outcomes.

2. Pretesting and Validation

Subsequent to the Delphi rounds, the retained 12 indicators underwent pretesting to assess their feasibility and applicability. The project completion rate, a measure of feasibility, was notably high, with 10 out of 12 indicators achieving completion rates above 90%. This suggests that the selected indicators are practical and manageable for the target population.

To evaluate the indicators' sensitivity to demographic variables, single-factor analysis of variance (ANOVA) was conducted. The results revealed significant differences in the mean scores of 8 indicators between male and female participants, indicating good sensitivity to gender differences in gross motor skills.

3. Reliability and Content Validity

The study further assessed the reliability of the indicators using Spearman's correlation coefficient, which ranged from 0.790 to 0.932, all exceeding the 0.70 threshold. This high level of correlation signifies strong inter-rater reliability, ensuring consistent evaluations across different rates.

Content validity was established through the Item-Objective Congruence (IOC) index. Experts rated each item, and the IOC values ranged from 0.80 to 1.00, surpassing the acceptable standard of 0.70. This indicates that the indicators are well-aligned with the defined objectives and effectively measure the intended constructs.

4. Equal Weighting of Indicators

Experts concurred on assigning equal weights to each indicator. This approach simplifies the evaluation process, making it more straightforward and less prone to subjective biases. Equal weighting ensures that each indicator contributes equally to the overall assessment, promoting objectivity and fairness in evaluating children's gross motor skills.

This model fills a critical gap in early childhood motor assessment. While numerous existing models emphasize gross motor development, few explicitly incorporate sensory processing dimensions. Traditional tools such as the Peabody Developmental Motor Scales-2 (PDMS-2) [16], and the Test of Gross Motor Development-2 (TGMD-2) [17], primarily assess motor performance without integrating sensory components. In contrast, the present model embeds sensory integration mechanisms at its core, providing a more comprehensive framework aligned with Ayres' Sensory Integration Theory [2], [18].

As outlined in Appendix B, the linkages between each indicator, the targeted motor skills, and the underlying sensory integration processes further support the construct validity of the developed model. For example, the "Forward Balance Beam Walking" task evaluates not only locomotor control but also proprioceptive feedback, vestibular stabilization, and tactile input through foot-ground interaction. Similarly, "Horizontal Bar Suspension" integrates upper limb proprioception with gravitational vestibular adaptation and palmar tactile cues, making it a holistic test of upper body sensory-motor coordination. These multi-dimensional indicators reflect the integrative nature of sensory processing and its role in purposeful movement.

A comparative analysis with existing indicators from China [19], Canada [20], and the United States [17] reveals that the present indicators are more functionally integrated. While previous assessments identify similar movement types, they often lack a structured explanation of how sensory inputs influence motor execution. The inclusion of movements such as "Four-legged Crawl" and "Side Roll" not only reflects motor complexity but also highlights bilateral coordination, cross-midline integration, and postural transitions, key aspects of central nervous system processing in sensory integration [21].

Moreover, the model supports early identification of sensory processing challenges. Research by Roley, Susanne Smith, et al. [22] emphasized that deficits in sensory integration can significantly hinder the development of motor skills in preschool-aged children. For instance, poor performance in the "Double Hop" task may suggest difficulties in vestibular-gravitational security or proprioceptive timing, which might otherwise go unnoticed in conventional evaluations. Such tasks are particularly valuable in identifying subclinical issues

related to Developmental Coordination Disorder (DCD), as emphasized by Zwicker et al. [23].

The structure of the indicators also aligns with the three core domains of gross motor skills: stability (e.g., balance beam walking, crawling), locomotion (e.g., running, hopping), and object manipulation (e.g., overhand throw, catching), as categorized by Gallahue et al. [24]. Their integration with sensory input systems makes them suitable for application in early childhood education, pediatric therapy, and inclusive programs for children with neurodevelopmental delays.

In addition to the strengths of the present model, several areas warrant further consideration. Extending the age range downward to include children aged 3–4 years may be particularly valuable, as this period represents a critical time for acquiring foundational locomotor and stability patterns. Moreover, testing the tool across multiple cultural and geographical contexts would help determine whether localized adaptations are required or whether the indicators demonstrate universal applicability.

Another important dimension is the integration of emotional engagement and play-based movement. Young children’s motor skill performance is often enhanced when tasks are embedded in playful and socially interactive contexts. Considering socio-cultural perspectives may further improve the ecological validity of the tool, ensuring that assessments reflect the diverse ways children engage in movement across different communities.

Finally, feedback from educators, therapists, and children themselves can be instrumental in refining the usability and engagement of the tool. Such qualitative insights would complement the quantitative rigor of the Delphi process and strengthen the tool’s acceptance in practical settings. The potential adaptation of the system into a digital or app-based format also represents a promising direction, offering accessible administration, automated scoring, and longitudinal tracking of motor development.

In summary, the proposed evaluation model provides a neurodevelopmentally informed, multidimensional method for assessing gross motor skills in children aged 5–6. By grounding the assessment in Sensory Integration Theory, it captures not only performance outcomes but also the sensory processes underpinning them. Future research should validate the reliability and construct validity of this model across diverse populations, establish age-normed benchmarks, and explore its utility in intervention planning for children with sensory-motor integration challenges.

5. Conclusions

Through a rigorous and systematic approach, this study successfully developed a concise and reliable set of 10 core indicators for assessing children's gross motor skills. The Delphi method facilitated expert consensus, while subsequent pretesting and validation ensured the indicators' feasibility, sensitivity, reliability, and content validity.

The high completion rates and significant gender sensitivity underscore the indicators' practicality and relevance. The strong inter-rater reliability and content validity further affirm the robustness of the developed tool.

The adoption of equal weighting enhances the tool's simplicity and objectivity, making it suitable for widespread application in educational and clinical settings. Future research should focus on longitudinal studies to assess the indicators' predictive validity and their applicability across diverse populations.

5.1. Limitations

This study is subject to several limitations. First, the age range of participants was restricted to children aged 5–6 years, limiting the generalizability of findings to younger age groups (e.g., 3–4 years) who are in earlier stages of motor pattern acquisition. Second, the data were collected exclusively from preschools in Guangxi Province, China, which may constrain the applicability of the tool across diverse cultural and geographical contexts. Third, while expert consensus and quantitative validation ensured reliability, qualitative feedback from educators, therapists, and children was not incorporated at this stage. Finally, the assessment remains paper-based, and further adaptation to digital or app-based formats is needed to enhance usability and tracking over time.

5.2. Implications

The development of a validated set of indicators for assessing children's gross motor skills carries significant implications across various domains, including educational practice, clinical assessment, and future research.

1. Educational Practice

The finalized set of 10 core indicators offers educators a concise and reliable tool for evaluating gross motor development in children. The simplicity of the equal-weighting approach enhances usability, allowing for straightforward implementation in diverse educational settings. This can facilitate early identification of motor skill deficiencies, enabling timely interventions and support tailored to individual student needs.

2. Clinical Assessment

For clinicians and pediatric therapists, the indicators provide a standardized framework for assessing motor development. The high inter-rater reliability and content validity ensure that assessments are both consistent and reflective of actual motor abilities. This can aid in the diagnosis of developmental delays and the formulation of targeted therapeutic strategies.

3. Policy and Program Development

The robust nature of the indicators, underpinned by

expert consensus and empirical validation, makes them suitable for informing policy decisions and the design of developmental programs. Policymakers can utilize these indicators to establish benchmarks and monitor the effectiveness of interventions aimed at enhancing children's motor skills.

4. Future Research

Future research should address these limitations by extending the applicability of the tool in several directions. First, validation with younger age groups (3–4 years) would capture the developmental trajectory of foundational motor patterns. Second, cross-cultural and multi-regional testing is recommended to examine whether cultural adaptations are necessary or whether the tool demonstrates universal validity. Third, pilot implementations in school settings could provide practical insights into feasibility, educator training requirements, and child engagement. Fourth, incorporating additional dimensions such as emotional engagement, play-based movement, and socio-cultural contexts could enrich the assessment framework. Fifth, qualitative feedback from teachers, therapists, and children should be collected to refine usability and ecological validity. Finally, the development of a digital or mobile app version of the tool could enhance accessibility, streamline scoring, and allow longitudinal tracking of children's motor development.

Acknowledgements

This research is part of a doctoral dissertation conducted at Mahasarakham University, Thailand, with the support of the Department of Health and Sport Science, Faculty of Education, Mahasarakham University. The author extends heartfelt gratitude to my advisor, Dr. Watthanapong Khongsuebsor, for her invaluable guidance and support throughout the research process.

Appendices

Appendix A. Final 10 Core Indicators and Scoring System

Indicator	Description of Child Behavior	Scoring (0–3 scale)
L1. Balance Beam Walking	Child walks forward on a narrow beam with arms extended, integrating proprioceptive and vestibular feedback.	0 = cannot complete; 1 = steps with frequent support/loss of balance; 2 = walks with 1–2 errors; 3 = walks smoothly without errors.
L4. Side Roll	Child performs a lateral roll on a mat, demonstrating trunk control and vestibular orientation.	0 = cannot roll; 1 = partial attempt with pauses; 2 = complete but uncoordinated roll; 3 = smooth, continuous roll.
L7. Four-legged Crawl	Child crawls forward using contralateral arm–leg coordination, engaging tactile and proprioceptive inputs.	0 = refuses or cannot crawl; 1 = asymmetrical crawl; 2 = coordinated crawl but unstable; 3 = coordinated, stable crawl.
L9. Overhand Throw	Child throws a ball overhead with trunk rotation, upper limb coordination, and tactile grip.	0 = cannot throw; 1 = weak throw without trunk movement; 2 = coordinated throw with limited distance; 3 = strong coordinated throw.
L12. Double Hop	Child performs consecutive two-foot hops, testing vestibular balance and lower-limb power.	0 = unable to hop; 1 = one hop only; 2 = multiple hops with loss of balance; 3 = multiple hops smoothly.
L19. Horizontal Bar Suspension	Child hangs on a bar using upper-limb strength and proprioceptive input.	0 = cannot grasp bar; 1 = holds <3 sec; 2 = holds 3–5 sec; 3 = holds >5 sec.
C22. Ten-Meter Back Run	Child runs backward over 10 meters, integrating vestibular stability and spatial orientation.	0 = cannot run backward; 1 = <5 m with stumbling; 2 = completes with errors; 3 = smooth completion.
C27. Forward Balance Beam Step	Child steps forward across balance beam with alternating feet, integrating tactile and proprioceptive cues.	0 = cannot step; 1 = steps with frequent support; 2 = steps with minor errors; 3 = smooth completion.
B42. Two-Handed Ball Throw	Child throws a ball forward using both hands and trunk strength.	0 = cannot throw; 1 = weak throw <1 m; 2 = throw with limited distance/coordination; 3 = strong, coordinated throw.
B49. Racket Movement Task	Child manipulates a racket (e.g., tapping or swinging), requiring hand–eye coordination and proprioceptive feedback.	0 = cannot use racket; 1 = irregular movement; 2 = coordinated but inconsistent; 3 = consistent, controlled movement.

Appendix B. Example of Indicator–Behavior Relationship

Indicator	Target Skill	Example Child Behavior Observed	Related Sensory Integration Components
Balance Beam Walking	Stability	Child maintains upright posture while stepping on a beam	Vestibular equilibrium, proprioceptive joint feedback
Overhand Throw	Object Manipulation	Child rotates trunk and releases ball with coordinated arm swing	Proprioceptive input from shoulder/arm, tactile grip
Four-legged Crawl	Locomotion	Alternating arm–leg crawl across mat	Bilateral coordination, tactile contact, vestibular orientation
Double Hop	Power & Rhythm	Consecutive two-foot hops with balance	Vestibular security, proprioceptive timing
Horizontal Bar Suspension	Upper Body Strength	Child grasps and hangs from bar for sustained seconds	Proprioceptive grip, vestibular gravitational adaptation

REFERENCES

[1] General Administration of Sport of China, "The fifth Chinese citizens' physical fitness monitoring communique," China Sports Press, 2021. http://www.sport.gov.cn (accessed Jul. 1, 2025).

[2] Ayres A. J., *Sensory Integration and Learning Disorders*, Western Psychological Services, 1972.

[3] The State Council of the People's Republic of China, "Report on child health promotion," http://www.gov.cn/zhengce/2022-05/03/content_5688452.htm (accessed Jul. 1, 2025).

[4] Ministry of Education, China, *Kindergarten Work Regulations*, Beijing: Ministry of Education, 1996.

[5] Ministry of Education, China, *Guide to Learning and Development for Children Aged 3–6*, Ministry of Education, 2012.

[6] Ministry of Education, China, *Outline for the Construction of a Sports Power*, Ministry of Education, 2019.

[7] Jin R., "Developmental characteristics of gross motor skills and sensory integration in preschool children aged 5–6," *Chinese Journal of Child Health*, vol. 30, no. 4, pp. 250–255, 2022.

[8] Gu S., "The impact of physical education teacher training on motor skill assessment in kindergartens," *Journal of Early Childhood Education Research*, vol. 12, no. 1, pp. 45–58, 2023.

[9] Ayres A. J., *Sensory Integration and the Child*, Western Psychological Services, 1979.

[10] Luo Y., "Movement as an act of sensory integration: Implications for early childhood development," *Early Childhood Development and Care*, vol. 191, no. 10, pp. 1500–1512, 2021.

[11] Dalkey N., Helmer O., "An experimental application of the Delphi method to the use of experts," *Management Science*, vol. 9, no. 3, pp. 458–467, 1963. DOI: 10.1287/mnsc.9.3.458.

[12] Linstone H. A., Turoff M., Eds., *The Delphi Method: Techniques and Applications*, Addison-Wesley, 2002.

[13] Hsu C.-C., Sandford B. A., "The Delphi technique: Making sense of consensus," *Practical Assessment, Research, and Evaluation*, vol. 12, no. 10, pp. 1–8, 2007. DOI: 10.7275/pdz9-th90.

[14] Okoli C., Pawlowski S. D., "The Delphi method as a research tool: An example, design considerations and applications," *Information & Management*, vol. 42, no. 1, pp. 15–29, 2004. DOI: 10.1016/j.im.2003.11.002.

[15] Keeney S., Hasson F., McKenna H., *The Delphi Technique in Nursing and Health Research*, John Wiley & Sons, 2011. DOI: 10.1002/9781444392029.

[16] Folio M. R., Fewell R. R., *Peabody Developmental Motor Scales*, 2nd ed., PRO-ED, 2000.

[17] Ulrich D. A., Test of Gross Motor Development-2 (TGMD-2), PRO-ED, 2000.

[18] Bundy A. C., Lane S. J., *Sensory Integration: Theory and Practice*, 3rd ed., F.A. Davis Company, 2020.

[19] General Administration of Sport of China, National Physical Fitness Monitoring Standards Manual for Children (Trial), China Sports Press, 2023.

[20] Henderson S. E., Sugden D., Barnett A. L., "Movement assessment battery for children-2," *Research in Developmental Disabilities*, 1992. DOI: 10.1037/t55281-000.

[21] Miller Kuhaneck H., Watling R., *Autism: A Comprehensive Occupational Therapy Approach*, 4th ed., AOTA Press, 2022.

[22] Roley S. S., Mailloux Z., Parham L. D., Schaaf R. C., Lane C. J., Cermak S., "Sensory integration and praxis patterns in children with autism," *The American Journal of Occupational Therapy*, vol. 69, no. 1, pp. 6901220010p1–6901220010p8, 2015. DOI: 10.5014/ajot.2015.012476.

[23] Zwicker J. G., Missiuna C., Harris S. R., Boyd L. A., "Developmental coordination disorder: A review and update," *European Journal of Paediatric Neurology*, vol. 16, no. 6, pp. 573–581, 2012. DOI: 10.1016/j.ejpn.2012.05.005.

[24] Gallahue D. L., Ozmun J. C., Goodway J. D., *Understanding Motor Development: Infants, Children, Adolescents, Adults*, 7th ed., McGraw-Hill, 2012.