國立臺灣大學醫學院物理治療學系暨研究所

碩士論文

School and Graduate Institute of Physical Therapy
College of Medicine
National Taiwan University

Master’s Thesis

遠端嬰幼兒動作評估行動應用程式「寶貝動起來」應用於早產和足月兒的使用性

Usability of the Mobile Application “Baby Go” for Remote Infant Motor Assessment in Preterm and Full-term Infants

謝淳雯
Chun-Wun Hsieh

指導教授:鄭素芳 博士
Advisor: Suh-Fang Jeng, P.T., Sc.D.

中華民國 114 年 7 月

July 2025

doi:10.6342/NTU202503829

誌謝

回首這段碩士研究的歷程，是一段既痛苦又快樂的旅程。從懵懂地踏入研究領域，到逐漸找到屬於自己的方向，研究所期間有時孤單無助，有時靈光乍現，每一個掙扎與突破的時刻，都讓我學到更多，並建立起對研究的自信與熱情。首先，我要衷心感謝我的指導教授—鄭素芳老師，能夠加入老師主持的嬰幼兒動作發展研究室，並參與發人工智慧技術之嬰幼兒動作評估發展行動應用程式的研究計畫，是我莫大的榮幸。老師不僅在學術上給予我深刻的啟發，也在我多次陷入瓶頸、感到壓力的時候，包容我的不足，耐心引導，給我信心和方向。老師從不責備，卻總是鼓勵我、相信我，這份溫柔與堅定，讓我逐漸建立起對研究的熱忱與自信，在研究中找到了成就感，也看見自己未曾發現的潛力，謝謝老師仔細指導我完成每一段論文內容，並帶領我參與研討會，拓展國際視野。誠摯感謝許永真教授、曹伯年教授、陳為堅教授在百忙之中擔任我的碩士論文口試委員，您們的專業建議與寶貴意見，讓我的研究論文更加完整。感謝研究室夥伴 Yohanes、蕭郁靜學姊、陳芊崴學妹，我們一起經歷了研究的酸甜苦辣，在彼此扶持的陪伴下，這段艱難的旅程變得不再孤單，你們是我最堅強的後盾，也是我在研究生活中最珍貴的回憶。我也要特別感謝許永真教授及其帶領的智慧型代理人實驗室研究團隊，感謝所有曾參與這項計畫的學長姐與同學們，包括：Erick Chandra 學長、丁立安學長、林念澤學長、陳逸寧學姊、金明毅學長、林揚昇學長、陳見齊學長、蕭昀豪學長、魏資碩學長、廖金億學長、呂兆凱學長、林鴻儒學長、張仲喆學長、陳韋傑同學、鄭雅勻同學、魏子翔同學、黃日昇學弟、蔡其翰學弟，謝謝你們在人工智慧模型與手機應用程式開發中提供的專業知識與技術，因為有你們的努力與協助，這項研究計畫才能收穫如此豐碩的成果。最後，我最深的感謝獻給我的家人。你們一直是我最溫暖的避風港，在我迷惘無助時，給予我無私的支持與愛，讓我能夠無後顧之憂地全心投入研究。這段碩士旅程即將畫下句點，我將把這份珍貴的經驗與學習傳承給學妹們，期許這項研究能持續進行，幫助更多有發展遲緩風險的嬰幼兒。我也將懷抱著這份初衷與熱忱，繼續攻讀博士學位，持續在人工智慧與嬰幼兒發展領域中深入探索，努力讓研究真正落實於臨床，發揮價值。

doi:10.6342/NTU202503829

Chinese Abstract

背景與目的：早產兒通常伴隨較高的神經發展障礙風險，尤其在出生後第一年常見動作發展遲緩。因此，及早評估其動作發展並提供早期介入至關重要。隨著數位科技的進步，越來越多遠距發展評估工具應運而生，讓家長能透過手機在家中拍攝嬰兒的動作影片並上傳進行評估。人工智慧技術也逐漸被應用於嬰幼兒動作發展分析，但大多限於嬰兒早期的仰姿動作。本研究旨在開發一款結合人工智慧技術的手機應用程式"Baby Go"，協助家長遠距評估嬰兒時期多姿勢下的動作發展，同時探討家長與臨床人員對此應用程式的使用情況與可行性。**方法：**本研究共招募 32 位早產兒與 19 位足月兒及其家長，在嬰兒 2 至 18 個月大之間使用手機應用程式"Baby Go"第二版評估動作發展。家長需進行嬰兒 38 項多姿勢之動作自評，並上傳影片以進行人工智慧模型分析，結果會在一至兩天後呈現。受訓過之物理治療師根據阿爾伯塔嬰兒動作量表的評估標準對影片進行標註，並評估影片品質，同時記錄家長使用手機應用程式的情形、使用過程中是否需要協助，以及人工智慧、家長與物理治療師三者間對嬰兒動作評估結果的一致性。家長會在嬰兒達 6、12 與 18 個月時填寫可用性問卷，物理治療師則在評估結束後填寫可用性問卷。**研究結果：**共有 29 位早產兒（93.5%）和 16 位足月兒（84.2%）至少使用應用程式上傳過一部影片，而 92.3% 的影片品質為中等以上。家長上傳影片的數量隨嬰兒月齡增加而逐漸遞減，3 到 5 個月的動作評估有最高的完整上傳率（50.0%到 54.5%），然而超過 50%的家長在嬰兒滿 13 個月後就不再上傳。排除直向拍攝或嬰兒方向錯誤的影片後，家長、人工智慧與物理治療師在動作評估結果的一致率分別為 79%、79%與 73%。大多數家長認為"Baby Go"使用起來簡單（93%），也幫助他們更了解嬰兒的發展狀況（100%）。但部分家長反映錄影的條件太嚴格及有時難以捕捉嬰兒的動作表現。物理治療師則認為"Baby Go"能夠有效地輔助篩檢嬰兒的動作發展。**結論：**家長使用行動應用程式"Baby Go"第二版上傳至少一段影片的上傳率高，多數家長也認為"Baby Go"行動應用程式容易操作。然而，人工智慧與物理治療師在嬰幼兒動作辨識上的一致率尚有不足，顯示仍有提升人工智慧模型準確度的空間。未來可能須調整部分動作項目，並持續增進應用程式的功能，以進一步提升其可用性與準確度。

關鍵詞：早產兒、行動應用程式、可用性、人工智慧、遠距嬰幼兒動作發展評估

doi:10.6342/NTU202503829

English Abstract

Background and Purposes: Preterm birth is associated with a higher risk of neurodevelopmental disorders, with motor impairment as the major disorder during the first year of life. Early motor assessment is essential for identifying potential motor delays and enabling timely intervention in preterm infants. With the advancement of digital technologies, remote screening tools have been developed to assess infant motor development at home based on parent-recorded videos. Artificial intelligence (AI) has shown promising potential in supporting physiotherapists in evaluating infant motor development; however, the applications were mainly for supine movements in early infancy. This study aims to develop a mobile application (APP), “Baby Go”, that integrates AI technology for remote infant motor assessment in various positions throughout infancy, and to investigate its usability from the perspectives of both parents and clinicians. **Methods:** This study enrolled 32 preterm and 19 full-term infants and their parents using the “Baby Go” APP to assess their infants’ motor development from 2 to 18 months of age. Parents were asked to perform parental perception assessment of 38 movements, and upload videos of the observed movements for AI classification. After uploading, parents received developmental results generated by AI assessment. Physiotherapists annotated the uploaded video using the Alberta Infant Motor Scale (AIMS) criteria. The APP usage, external support for APP use, quality of videos, and agreement between the AI, parents’, and physiotherapists’ labeling results were examined. Parents provided feedback on the APP usability when their infants approached 6, 12, and 18 months of age. Two physiotherapists provided feedback on the APP usability when their patients reached 18 months of age. **Results:** Twenty-nine (93.5%) preterm and 16 (84.2%) full-term infants uploaded at least one video using the “Baby Go” APP version 2. Of the 936 uploaded videos, 864 (92.3%) showed fair to

doi:10.6342/NTU202503829

high quality. Uploaded frequency decreased gradually with age. Most parents completed video uploads during the 3- to 5-month assessments, with more than 50% failing to upload any videos beyond 13 months of age. After excluding those videos with incorrect recording format, the agreement between the AI, parents', and physiotherapists' results for each movement were 79%, 79%, and 73%. The usability survey from parents revealed that 93% considered the APP easy to use, and all parents reported that the APP helped them better understand infant development. However, some parents reported that the recording criterion is rigid and difficult to meet, and the baby's movement is hard to capture. The clinicians indicated that the APP helps screen infant motor development. **Conclusion:** The upload rate by parents of preterm and full-term infants using the "Baby Go" version 2 was high, and most parents reported that the APP was easy to use. However, the moderate agreement between the AI and physiotherapists' results highlights the need to improve the AI model's accuracy. Future work is necessary to enhance the AI model accuracy and adjustment of movement items to further increase the APP's usability.

Keywords: preterm infants, mobile application, usability, artificial intelligence, remote infant motor assessment.

doi:10.6342/NTU202503829

Contents

doi:10.6342/NTU202503829

doi:10.6342/NTU202503829

doi:10.6342/NTU202503829

國立臺灣大學

doi:10.6342/NTU202503829

Contents of Tables, Figures, and Appendices

Tables

doi:10.6342/NTU202503829

doi:10.6342/NTU202503829

Contents of Tables, Figures, and Appendices

Figures

Appendices

doi:10.6342/NTU202503829

Chapter I. Introduction

1.1 Epidemiology of Children with Developmental Delay

A developmental delay refers to children attaining developmental milestones in areas such as motor, language, cognitive, and socio-emotional function at a later age than their peers (Khan & Leventhal, 2024). Children with developmental delay have a greater risk of subsequent suboptimal health, educational attainment, and well-being compared with children with normal development (World Health Organization, 2012). Therefore, early identification and intervention of children who have developmental delays is an important goal for global child health and wellness.

Preterm birth is defined as birth before 37 completed weeks of gestation (World Health Organization, 1977). Approximately 11% of newborns are born prematurely worldwide (Vogel et al., 2018). In Taiwan, the rate of preterm birth has gradually increased from 8.98% in 2008 to 10.88% in 2023 (Ministry of Health and Welfare in Taiwan, 2008; 2024). A population-based study conducted by Su et al. (Su et al., 2016) in Taiwan found that the rising rate of preterm birth may relate to enhanced reproductive technology, such as in *vitro* fertilization, increased age of the mothers, and an increase in Caesarean sections before term age. Furthermore, the more widespread use of antenatal steroids over the last two decades improved the survival rates of very preterm (VPT) infants (Grisaru-Granovsky et al., 2014; Stoll et al., 2015).

However, the surviving preterm infants are more vulnerable to neurodevelopmental disorders than their term peers (Cheong et al., 2017; Pascal et al., 2018). A meta-analytic review showed that the pooled prevalence of cognitive and motor delays was 16.9% and 20.6%, respectively, for preterm infants with very low birth weight from 18 months corrected age until 6 years (Pascal et al., 2018). Furthermore, moderate and late preterm infants showed a higher likelihood of

doi:10.6342/NTU202503829

developmental delays than their term peers, with risk ratios of 1.8 for cognitive delay, 3.1 for language delay, and 2.4 for motor delay at age 2 years (Cheong et al., 2017). These results suggest that parents and healthcare professionals must pay more attention to the risk of developmental problems in preterm infants.

1.1.1 Prevalence of Developmental Delay in Children

The World Health Organization (WHO) statistics in 2016 reported that the prevalence rate of developmental disorders in children is 8.4% globally (Olusanya et al., 2018). However, in Taiwan, among children seeking intervention, 85.5% are identified with developmental disorders after two years old. Only 10.3% of these children are identified because of their parents' awareness and notification (Ministry of Health and Welfare in Taiwan, 2023). Improving the ability of parents to self-screen their baby's development at a young age is necessary for the early detection of developmental disorders and for ensuring timely intervention.

1.1.2 Significance of Motor Development

Several studies have reported that motor development in early childhood affects later development in other domains. For example, writhing and fidgety movements before 16 weeks corrected age assessed by the General Movement Assessment (GMA) predict motor and cognitive outcomes at 24 months in infants with very preterm and very low birth weight (Caesar et al., 2023). Motor skills at 6 and 12 months were positively associated with communication skills at 24 months in preterm and full-term infants (Valla et al., 2020). The emergence of sitting is a significant predictor of subsequent language development at 10 and 14 months of age in full-term infants (Libertus & Violi, 2016). Moreover, early developmental intervention in the first year of life is effective in improving developmental outcomes in preterm infants (Spittle et al., 2024). Thus, early motor assessment of infants provides a critical opportunity to detect

doi:10.6342/NTU202503829

those who may have developmental disorders and could benefit from early intervention.

1.2 Infant Motor Assessment

Infant motor assessments contain screening, diagnostic, and intervention tests. Intervention tests, such as the Early Intervention Developmental Profile (EIDP) and the Portage Guide to Early Education (Portage), are used as references for developing early intervention plans (Rogers, 1975; Shearer et al., 1972). Diagnostic tests, such as the Alberta Infant Motor Scale (AIMS), the Peabody Developmental Motor Scales- Third Edition (PDMS-3), the Bayley Scales of Infant and Toddler Development- 4th Edition (Bayley-4), and the Mullen Scales of Early Learning (MSEL) are standardized tools that can assess infants' motor development with respect to the normative data to form developmental diagnosis (Bayley & Alyward, 2019; Folio & Fewell, 2023; Mullen, 1995; Piper & Darrah, 2021). However, the diagnostic assessments are administered by health professionals in a clinical setting, and are usually costly and time-consuming. As a result, high-risk infants may miss the critical time for early detection of motor delay and early intervention. For example, the AIMS is widely used in clinical settings based on observation of children’s motor development from birth to 18 months of age (Piper & Darrah, 2021). It has clinical feasibility and acceptable psychometric properties; however, early detection of motor disorders using the AIMS is only available when infants have access to hospital settings for standardized assessment by physiotherapists. Furthermore, infants may be distressed when assessed in an unfamiliar environment, leading to unfavorable assessment results.

Screening tests, usually performed by parent report, are used for detecting children who have potential developmental disorders. Parent-report assessments such as The Ages and Stages Questionnaire, Third Edition (ASQ-3) and the Early Motor Questionnaire (EMQ), are cost-effective and easy to perform (Libertus & Landa, 2013;

doi:10.6342/NTU202503829

Squires et al., 2009). The Ages and Stages Questionnaire, Third Edition Gross Motor domain (ASQ-3-GM) showed acceptable agreement with the AIMS results (using score ≤10th centile as the cut-off) [sensitivity 77%, specificity 91%, and positive predictive value 95%] on full-term and preterm infants aged before 18 months (Fauls et al., 2020). Parental reports on the EMQ showed a high concurrent validity (correlation coefficient > 0.9) with the MSEL results in children aged between 3 and 24 months (Libertus & Landa, 2013). Although these questionnaires have acceptable correlations with the diagnostic test results, there are disagreements in the questionnaire results reported by caregivers and healthcare professionals. For example, the agreement between caregivers and healthcare professionals using the ASQ-3 on 9-month-old preterm and full-term infants was fair for the total score (Concordance correlation coefficient = 0.54) in Nepal (Shrestha et al., 2020). Maternal perception of full-term infants' motor development using a questionnaire was overestimated compared to the evaluation results by physiotherapists (Dos Santos et al., 2022). Therefore, assistance from healthcare professionals in conducting screening assessments is necessary.

The Centers for Disease Control and Prevention in the US recommends that child development screening and assessment include three steps: developmental monitoring by parents completing a brief checklist of milestones at home, screening by healthcare providers using a formal questionnaire or checklist at regular well-child visits, and evaluation by trained specialists to identify and diagnose developmental delays (Centers for Disease Control and Prevention, 2024). The Ministry of Health and Welfare in Taiwan launched a new child development screening model on July 1, 2024. Six child development screening services are provided by healthcare professionals for children under the age of 7, aimed at enhancing early detection of infants at risk of neurodevelopmental disorders (Ministry of Health and Welfare in Taiwan, 2024).

doi:10.6342/NTU202503829

1.3 Remote Infant Motor Assessment

With the advent of the internet, digital technologies have started to expand rapidly in the healthcare system. The coronavirus disease 2019 (COVID-19) pandemic has further accelerated the widespread adoption of digital health to provide remote care (Choukou et al., 2021). Digital health refers to the use of information and communications technologies in the healthcare system to manage diseases, mitigate health risks, and promote wellness (Ronquillo et al., 2017). It encompasses a wide range of application categories, including mobile health (mHealth), health information systems, wearable technology, telemedicine, and telehealth (Food and Drug Administration, 2020). It has the potential to improve the health of the pediatric population by expanding access to care, overcoming geographical and time-related constraints, reducing travel costs, and facilitating enhanced monitoring of patients, which is considered a practical alternative to traditional, clinic-based assessments for infant motor screening (Utidjian & Abramson, 2016).

1.3.1 Video-based Digital Technology

According to the WHO, mobile health (mHealth) is the use of mobile phones or wireless technologies to help meet health goals (World Health Organization, 2017). Boonzaaijer et al. designed a home video method to evaluate infants' motor development based on the AIMS (Boonzaaijer et al., 2017). Parents followed the instructions provided by professionals and recorded their infants' movements using a mobile phone or camera. After that, healthcare professionals received the videos and conducted the remote AIMS assessments. This remote AIMS assessment using the home video method was shown to have good inter- and intra-rater reliability (both intraclass correlation coefficient, ICC > 0.99) and good concurrent validity between

doi:10.6342/NTU202503829

video and live observation (ICC = 0.99), and most parents considered it's easy to record the videos (Boonzaaijer et al., 2017).

1.3.2 Smartphone Applications for Infant Motor Assessment

Since the debut of the first iPhone in 2008, smartphone technology has become more advanced and prevalent, offering a promising platform for delivering healthcare information to the general population. One of the key features of smartphones is their ability to operate software programs known as applications. Mobile Health Applications (mHealth APPs) are software designed for health-related purposes, typically on portable devices like smartphones (Kao & Liebovitz, 2017). More than 325,000 mHealth APPs are now available on the APP store globally (Pohl, 2017). Among them, several applications are specifically designed for infant motor assessment, providing parents and healthcare providers with convenient tools to monitor and assess early infant motor development remotely.

1.4 Existing APPs for Infant Motor Assessment

1.4.1 No Embedment of Artificial Intelligence

Several smartphone applications were developed to perform remote infant motor assessments, including the *Baby Moves* (Kwong et al., 2019), *NeuroMotion* (Svensson et al., 2021), and *In-Motion* (Adde et al., 2021). These APPs were all designed to examine infant motor development based on the General Movement Assessment. Parents followed the instruction guidelines summarized in the APP to learn how to record their infant's general movements and then sent the videos through the APP, allowing physiotherapists to perform the remote assessment. Most parents using these APPs reported that they felt safe sending the videos, and the filming instructions were easy to follow. Most of these videos were also classified as "scorable," which means they were of good quality and could be assessed by trained professionals via these

doi:10.6342/NTU202503829

APPs. However, these APPs can only be used for infant motor assessment in the first 20 weeks of age and require human labor to review and assess the videos. Furthermore, parents cannot receive the results and feedback immediately (Adde et al., 2021; Kwong et al., 2019; Svensson et al., 2021).

1.4.2 Embedment of Artificial Intelligence

Artificial Intelligence (AI) has emerged as one of the hottest and most transformative topics in recent years. Its rapid advancements are revolutionizing the healthcare system. It has been implemented in disease diagnosis and prognosis, treatment optimization, outcome prediction, drug development, and public health (Noorbakhsh-Sabet et al., 2019). AI is a field focusing on automating intellectual tasks normally performed by humans (Choi et al., 2020). Machine learning (ML) as a subdiscipline of AI focuses on the learning aspect of AI by developing algorithms that best represent a set of data (Choi et al., 2020). Deep learning is a branch of machine learning that imitates the operation of the human brain by using multiple layers of artificial neural networks to make predictions from large sets of training data automatically. (Choi et al., 2020).

Recently, AI technology has been incorporated into pediatric clinical applications for disease diagnosis (Gensure et al., 2020; Huang et al., 2023). For example, Huang et al. developed a novel automatic system (DDHnet) using AI for fast and highly accurate ultrasound diagnosis of infant developmental dysplasia of the hip (DDH) (Huang et al., 2023). Several research teams developed AI approaches for automated diagnosis of retinopathy of prematurity (ROP) based on medical images and data (Gensure et al., 2020).

Integrating AI techniques into infant motor assessment requires a more complicated algorithm for movement tracking, recognition, and classification. Deep

doi:10.6342/NTU202503829

high-resolution representation learning for human pose estimation is a revolutionary approach that learns high-resolution feature representations without reducing resolution, which has been applied in skeleton-based action tracking on adults (Sun et al., 2019). PoseConv3D is a 3D Convolutional Neural Network (3D-CNN) based method primarily for skeleton-based action recognition that models human key points across time and space, enabling precise movement recognition and analysis in videos (Duan et al., 2022).

1.5 Mobile APP "Baby Go"

In 2021, our research team developed a smartphone APP, "Baby Go," with the ultimate goal of integrating AI technology for remote infant motor assessment based on the framework of the AIMS assessment tool.

1.5.1 Artificial Intelligence and Machine Learning

From 2021 to 2024, our research teams collected infants' movement videos recorded from the laboratory and home to build an AI algorithm. Physiotherapists labeled the presence of movements in the videos as the gold standard and input these labeling results into an AI model for developing an AI algorithm, including action recognition. The preliminary data of the AI algorithm in classifying 31 out of the 58 AIMS movements in full-term and preterm infants from 4 to 18 months of age in a standardized clinical setup showed high agreement (accuracy=0.91, recall=0.91, precision=0.91, and F1 score=0.91) (Lin et al., 2024). The latest data of the AI algorithm in classifying 38 out of the 58 AIMS movements based on lab videos showed high performance (accuracy=0.91, recall=0.91, precision=0.91, and F1 score=0.91) and good performance based on home videos (accuracy=0.84, precision=0.84, recall=0.77, and F1-score=0.78) (Purwanto et al., 2025).

doi:10.6342/NTU202503829

1.5.2 Development of "Baby Go" APP

The first version of "Baby Go" included 58 movements from the AIMS assessment with the movements illustrated with textual descriptions and pictures. Parents previewed, recorded, and uploaded these movements at the corresponding age of their infants. Instructions and guidelines for recording were later added to the APP. Furthermore, we revised the movement description and added demonstration videos to each movement to make it easier for parents to understand and follow.

1.5.3 Usage of the "Baby Go" APP First Version

The upload rate of at least one video was around 67.2% in the first version of "Baby Go," and this may be because the APP was mainly for uploading videos (Tsai et al., 2025). The agreement of parental perception of 58 AIMS movements with the physiotherapist's assessment results was only 61.3%, which may be because parents needed to pay attention to record the video, keep the infants safe, and elicit their movements simultaneously. Furthermore, some confusing movement items were hard to identify, even for physiotherapists. Poor video quality was also a problem for physiotherapists in identifying the movements (Hsiao, 2023). These findings indicated that a new version of the APP is needed to improve the features and usability of "Baby Go".

1.5.4 "Baby Go" APP Second Version

In the second version of "Baby Go," we added the functions of assessment, follow-up, and education. The description of each movement was refined again, and the interface design underwent significant changes. It has become more personalized, with the interface divided into sections for completed actions, goals, and future actions. The original APP included 58 movement items, but the second version of the APP featured

doi:10.6342/NTU202503829

only 38 movement items identified from the AI model and structured them monthly for the age of 3-18 months to simplify the assessment.

1.6 Usability of APP

In addition to providing accurate developmental information, usability is a key factor in developing digital health applications. There is an increased emphasis on usability in the widespread adoption of health applications because people using these applications may encounter various challenges (Zapata et al., 2015). Maramba et al. reported in their systematic review that questionnaires are the most frequently used usability evaluation in mobile Health applications to explore an overall measure of usability. Qualitative methods may be more helpful in identifying the issues that need to be addressed (Maramba et al., 2019). It is necessary to conduct robust evaluation methods to ensure good usability. Jin and Kim developed an evaluation tool for healthcare smartphone applications with good reliability (Cronbach's alpha=0.905) in 2015. The tool used a structured questionnaire, exploring multiple aspects of the APP, including contents, interface design, and technology. Through these dimensions, users can comprehensively evaluate the APP's usability and reliability (Jin & Kim, 2015). Applications designed for remote general movement assessment on infants, such as *Baby Moves* (Kwong et al., 2019), *NeuroMotion* (Svensson et al., 2021), and *In-Motion* (Adde et al., 2021) applied the modified questionnaires based on Jin and Kim's method to investigate the usability of APP from infants' parents. Parents received the questionnaire by email after using the APP to provide feedback, including several dimensions, such as parents' experience of the APP, perceptions of security, recording instructions, and technique. Most parents felt safe while uploading the videos and found the APP easy to use. Common features in the APP, such as instructions for recording, reminder notifications, a baby silhouette, and set recording length time, helped enhance

doi:10.6342/NTU202503829

usability. Instructions for recording provide clear guidance on how to position the infant, how to hold the camera, appropriate clothing, and the infant's behavioral state, all aimed at enabling better-quality recordings of the infant's general movements for scoring. Reminder notifications reminded users to record videos of fidgety general movements at specific times (time point 1: 12^{+1} to 13^{+6} weeks, time point 2: 14^{+1} to 17^{+6}) to ensure timely data collection, essential for tracking development and comprehensive analysis. A baby silhouette was a visual guide, such as an outline of a baby, displayed on the APP to help users position their baby properly and capture movements in the required format. The required recording length was set up to ensure videos are long enough to capture activity of interest without being unnecessarily long, making video assessment easier for the GMA examiner.

Professional perspective is another vital aspect of developing digital health technologies. Lewis et al. explored the perspectives of caregivers and healthcare providers toward a digital health application called "Nemours" APP for managing asthma in 5-11-year-old children using questionnaires (Lewis et al., 2023). Parents completed surveys about APP experience regarding medication reminders, symptom tracker, access to telehealth visits, and education resources. Clinical providers completed surveys about the frequency of accessing patient APP data, type of patient data accessed (tracker entries, messages, symptoms), barriers to data access, and whether they inform clinical decisions about patient care through the use of the APP. The study found that around half of caregivers and clinical providers were willing to use the APP and access APP data for managing child asthma. Several strategies for enhancing the utilization of the APP, such as creating alerts for using the APP, video tutorials about how to use the APP and potential benefits of the APP, and medication adherence features for encouraging continued APP usage, were suggested (Lewis et al.,

doi:10.6342/NTU202503829

2023). These findings suggest that by including usability surveys, developers of applications can gain a deeper understanding of users' utilization, thereby enhancing the overall functionality of the applications.

1.7 Rationales of the Study

The rationales of this study are as follows. First, research evidence indicates that motor development before age two is closely linked to various aspects of later development. However, the rate of parental awareness of their children's developmental status under two years of age remains low. Current remote motor assessment applications primarily target infants younger than 20 weeks. This highlights the critical need for an application to include motor development in children under two years to enable regular remote monitoring and early detection of developmental disorders. Second, existing remote infant motor assessment applications rely on human evaluators to analyze videos, which is labor-intensive. Integrating the AI technology into the application may reduce human workload while maintaining clinical utility and accuracy. Third, in the previous version of the "Baby Go" APP, the agreement between parental perceptions of 58 AIMS movements (assessed before 18 months) and physiotherapists' evaluations was unsatisfactory (61.3%). This demonstrates the need for improved APP design and educational features to enhance parental understanding and interpretation of motor development. Finally, evaluating APP usability is a fundamental step in developing applications for family and professional use. To promote widespread adoption of the "Baby Go" APP in the clinic, it is essential to assess its feasibility and usability through comprehensive surveys involving parents and healthcare providers.

1.8 Purposes of the Study

The purpose of the study is three-fold: (1) to develop the "Baby Go" APP version 2.0 with the features of assessment, follow-up, and education, (2) to compare the

doi:10.6342/NTU202503829

parental perception of infant motor assessment with the AI-driven screening results via the "Baby Go" APP, and (3) to investigate the feasibility and usability of the "Baby Go" APP by parents of full-term and preterm infants and clinical providers.

1.9 Hypotheses of the Study

The corresponding hypotheses of the study are:

(1) The upload rate of at least one video using the "Baby Go" APP is greater than 80% of the parents.

(2) The proportion of high-quality videos is greater than 80%.

(3) The agreement between parental perception and AI-driven assessment results is greater than 80%.

(4) All the clinical providers and over 80% of the parents agree that the "Baby Go" APP is feasible and informative to use through an online survey.

doi:10.6342/NTU202503829

Chapter II. Method

2.1 Participants

Preterm infants and full-term infants were recruited from a well-baby and neonatal follow-up clinic at the National Taiwan University Children's Hospital (NTUCH). Two pediatric physiotherapists who provided developmental care to preterm infants at the study hospital were recruited. Preterm infants were recruited if their gestational age is of < 37 weeks and birth body weight is of < 2500 grams. Full-term infants were recruited if their gestational age ranged 37 - 42 weeks and their birth body weight was of > 2500 grams. Infants with congenital genetic abnormalities were also included to expand the variability of subjects. All the infants were recruited before 18 months of age. The percentage of male infants was similar to that of female infants in each group. Before participating in the study, the research assistant informed infants' parents and physiotherapists about the study's details and obtained a written consent form. This study has been approved by the Institutional Review Board at the National Taiwan University Hospital (202311095RIND and NCT06521918).

2.2 Testing Procedure

Information regarding the infant's date of birth, gestational age, and birth body weight was abstracted from the electronic medical records system at the study hospital. Parents provided socio-demographic data in a telephone interview regarding their age, education level, occupation, family income, marital status, primary caregiver, and family members. Pediatric physiotherapists provided information regarding their age, sex, and work experience through a telephone interview by the research assistant.

A research assistant assisted the parents and physiotherapists in installing the "Baby Go" APP and taught them how to use it by sending the instructional manual and demonstration videos. Parents used the APP when their infants were 3 months old or

doi:10.6342/NTU202503829

above. Parents were encouraged to observe the motor behaviors of their infants and completed the parental checklist in the APP for infant motor development biweekly. However, we later found that the APP design did not allow illustration of parental perceptions for each movement item. We therefore sent a monthly questionnaire about parental perception of infant movements (observed or not observed) via Google Forms. Then, parents recorded the observed movements and uploaded the videos to the APP. The AI model in the APP automatically assessed the videos showing the "presence" or "absence" of each assessed movement and provided the overall developmental results to the parents. If the APP shows the result of development within the expected range, parents will continue uploading videos in the following month. If the APP shows the result of development not within the expected range, physiotherapists will review the uploaded videos to determine whether there is an AI issue or an indication of developmental delay. In cases where poor video quality is identified as a potential issue, parents will be encouraged to re-film the movement. If the APP still shows the result of development not within the expected range by the last week of the assessment month, infants will be invited to the laboratory for the AIMS assessment. If the AIMS result shows motor delay, infants should visit a neonatologist for further examination and follow-up. A flowchart for the APP developmental assessment, AIMS assessment, and follow-up is shown in Figure 1. The procedure for using the APP for remote motor assessment is shown in Figure 3.

When the infants approached 6, 12, and 18 months, an online link to a questionnaire about user experience would be sent by Gmail to the parents. Physiotherapists who had provided early intervention services to the babies whose parents had used the APP "Baby Go" received a questionnaire by Gmail when the infants approached 18 months.

doi:10.6342/NTU202503829

2.3 Development of APP

The development of the APP is shown in Figure 2. The first version of the "Baby Go" APP was designed by a graduate student of the Department of Computer Science & Information Engineering, NTU, and released in 2021. The APP was mainly intended to collect home videos. It consisted of three components: (1) video uploading, (2) instructions for video recording, and (3) illustration of 58 AIMS movements with text descriptions and pictures of movements. The research assistant helped parents set up an account, and then parents could begin to use the APP. Parents followed the instructions and recorded the videos of the infant's movements listed in the APP. The uploaded videos were used to develop the AI algorithm for home videos. Version 1.1 and 1.2 of the "Baby Go" APP were later updated in 2022 with the revision of the instructions for video recording and illustrating 58 AIMS movements. Example videos were added to the recording guidelines to assist parents in understanding how to record clearly. Wordings were revised, and additional video demonstrations of each movement were added to enhance the parents' perception and knowledge of the infant's movements.

To expand the APP into clinical use, the second version of the "Baby Go" APP underwent drastic revisions and was launched in October of 2024. The APP contained the assessment integrated with an AI model, follow-up, and education features. Unlike the previous versions, parents can set up the account and log into the APP alone, not by a research assistant. Fifty-eight AIMS movements used to assess infant motor development were reduced to 38 movements in the second version of the APP due to the varying amounts of data collected for each movement. Thirty-eight of the 58 movements were finally selected for the final version of the AIMS movement set in the application with sufficient data, frequent occurrence, and developmental significance. Concurrent validity of the 38 items of AIMS compared to the 58 items of AIMS

doi:10.6342/NTU202503829

dispatched into age-based sets (Appendix 1) was high in preterm infants and full-term infants across ages (agreement = 0.99, sensitivity = 0.92, specificity =0.99, positive predictive value = 0.86, and negative predictive value =1.00) (Purwanto et al., 2025). Additional features were added, including reminders for uploading videos and a video editor, to enhance usability. Uploading interference was changed from one page with all the movements grouped into three dimensions: completed, in progress, and future movements over the 3-18 months age band. Version 2.1 was released in December 2024 with a minor revision regarding the appropriate recording length. Version 2.2, launched in January 2025, included a logic revision for determining passing or not passing the movement assessment: if an infant passed the mature movement assessment, less mature movements would be automatically treated as passing as well. A video upload guide was also added to assist parents in identifying confusing items and ensuring the correct videos were uploaded. Later, we identified an issue where videos recorded and uploaded by parents were automatically rotated vertically, leading to errors in AI recognition. This issue was addressed and resolved in version 2.3, which was updated in March 2025. The APP features are shown in Table 1, and the revised evaluation logic in version 2.2 is shown in Table 2.

2.4 Development of AI Action Recognition Model

2.4.1. Previous Data Collection of Lab and Home Videos

From 2021 to 2024, our research team recruited preterm and full-term infants between 4 and 18 months of age to the laboratory for the AIMS assessment every 2 months to collect their movement video records. Simultaneously, their parents were asked to use the first version of the "Baby Go" APP to upload their infants' movements on the AIMS milestones at home. We collected infant movement videos during in-person AIMS assessments using five cameras from different angles in the laboratory and

doi:10.6342/NTU202503829

home videos from the APP. Parents were asked to send home videos temporarily through LINE during a small time period in which the APP was not working due to technical issues. Videos were all subsequently annotated by physiotherapists to develop and validate an AI model in the laboratory and home setting, respectively. The AI model for home setting was then incorporated into the second version of “Baby Go” APP (Purwanto et al., 2025).

2.4.2. Current Data Collection of Home Videos

The second version of “Baby Go” APP with the previously established AI model was launched in 2024 October and provided to parents for remote infant motor assessment from 3 to 18 months of age in this study. Parents recorded and uploaded videos to the APP to follow up their infant motor development at home every 2 weeks. They followed the recording guideline and previewed the description and demo video in the APP for each assessing movement before uploading. They obtained the AI motor assessment results within two days of video uploading. All the home videos were sent to physiotherapists to annotate the 38 movements to keep improving the accuracy of the AI model.

2.4.3. Data Processing

All videos were anonymized using coded filenames to protect privacy and were stored on a computer server in MP4 format for processing. Each movement segment in the recordings was annotated by physiotherapists using the labeling tool, developed by our research team, to mark the start and end times. Three physiotherapists were trained to reach an acceptable level of reliability (ICC > 0.85) in all the 58 AIMS movements from 5 camera views of lab videos in 3 infants aged 4, 8, and 38 AIMS movements from home videos in infants between 4 to 18 months of age before beginning annotation work. Annotators first previewed the videos and then labeled the movements according

doi:10.6342/NTU202503829

to the AIMS criteria and our labeling instructions (Appendix 2). For laboratory videos, movements were annotated as either perfect presence (the movement characteristics could be fully identified) or acceptable presence (the movement features were difficult to distinguish between confusing items due to suboptimal camera angles). For home videos, movements were annotated as: perfect presence, acceptable presence, or partial presence (the movement was visible, but its duration was shorter than the required criteria). The annotated results were then sent to the computer science team for AI model training and validation. Reliability of second and third rater with first rater on home videos in preterm and full-term infants between 4-18 months of age is good (ICC=0.89-1.00) and excellent (ICC=0.90-1.00).

2.4.4. Training and Validation of AI Models

Developing the AI action recognition model required both model training and validation. Training and validation set were divided into ratio of 8:2 to avoid overfitting in model development (Shahroudy et al., 2016). The first AI model incorporated into the second version of "Baby Go" APP during the time period of October 25, 2024 to January 13, 2025 used the annotated results of 1,396 home videos (G00001-G00324 and H00000-H01072) uploaded via the first version of "Baby Go" APP and LINE as the training and validation set. The first AI algorithm in classifying 38 out of the 58 AIMS movements showed good performance based on home videos (accuracy=0.84, precision=0.84, recall=0.77, and F1-score=0.78) (Purwanto et al., 2025)(Table 3). The second AI model was incorporated into the APP after January 14, 2025, which used the 36,805 annotated results from 474 lab videos (S001-S099 and P001-P375) and 2,257 annotated results from 1,419 home videos (G00001-G00324 and H00000-H01095) collected via the first version of "Baby Go" APP and LINE as the training and validation set. A new training strategy was used in the second AI model development,

doi:10.6342/NTU202503829

including pre-training on the lab video dataset, then fine-tuning with the home video data. This method allowed the model to transfer robust features learned in the laboratory videos to the more variable home videos, which have diverse angles and motion artifacts like camera shake. It greatly boosted training efficiency, cutting epochs from 240 to 50 and enabling a smaller learning rate (0.001). The second AI algorithm in classifying 38 out of the 58 AIMS movements showed a higher accuracy than the first AI model (accuracy=0.87, precision=0.85, recall=0.82, and F1-score=0.82) (Table 3).

2.5 Outcome Assessments

2.5.1 APP Usage

APP usage was used to evaluate parents' engagement and adherence to the APP. It was examined using four measures. First, the rate of APP usage was defined as the proportion of users who had uploaded at least one video among the total participating users during the study period. Second, the upload rate for monthly assessment was defined as the ratio of the number of uploaded parents to the total expected uploaded parents at the corresponding month. Third, the upload frequency from the first to the last month of usage, and the average number of videos uploaded by parents per month. Fourth, the upload rate per movement was defined as the ratio of actual uploads to the expected movement video uploads for a specific movement (some movements required at more than one age point). All the data was abstracted from the background of the APP and documented in Excel files.

2.5.2 External Support for APP Use

External support for the APP was aimed at investigating the level of parental assistance required and identifying features to be enhanced for future APP development. It was examined through two primary metrics. The proportion of parents requiring additional reminders from the research assistant for video upload was defined as the

doi:10.6342/NTU202503829

percentage of users who received reminders from the research assistant during the final week of a monthly assessment out of the total number of participating users in the study period. The proportion of parents who contacted the research assistant regarding APP-related questions was defined as the percentage of users who reached out to the research assistant via message or phone call for APP-related inquiries out of the total number of participating infants during the study period. All data was documented in Excel files following each contact.

2.5.3 Quality of Video Recordings

The quality of video recordings was assessed to determine whether the videos uploaded by parents met the minimum quality requirements for remote assessment and to determine the feasibility of APP assessments. Quality video assessment was performed using the criteria for infant body parts, camera angle, and camera movement previously described by Hsiao (Hsiao, 2023). One physiotherapist checked the uploaded videos in the background of the APP and labeled the videos as “poor,” “fair,” or “high” in the Excel files. The criteria for video quality are shown in Table 4. Furthermore, any videos recorded with incorrect recording format, such as portrait mode (i.e., when parents held the cellphone in a vertical position during video recording) or with incorrect infant orientation (in which the infant’s head and feet were not properly aligned) were also noted for further assessment if affecting the AI-based movement recognition.

2.5.4 Agreement between Parents and PT Results

The agreement between the parents’ and physiotherapists’ labeling results for uploaded movements was used to evaluate the extent to which parents can recognize and correctly identify each infant's movement when using the APP. Results of parental perception were shown in the parents’ APP accounts. Parental perception was examined

doi:10.6342/NTU202503829

using the agreement between the movement parents choose to upload and the labeling results in the Excel file from the physiotherapists for each movement. Movement parents choose to upload were collected from the APP's background (Figure 4).

2.5.5 Agreement between AI and PT Results

The agreement between the AI and physiotherapists' labeling results for uploaded movements showed the performance of the AI model applied on the uploaded videos. The AI movement recognition results were obtained from the APP's background for analysis (Figure 4). Two physiotherapists checked the video in the APP's background and labeled each movement from the uploaded videos in the Excel files using labeling criteria in the AIMS assessment (Appendix 2). A consensus decision would be made by the third physiotherapist if there was disagreement between the two physiotherapists' labeling results. All three physiotherapists conducted an interrater reliability test and achieved an acceptable level of reliability (ICC > 0.85) in labeling the movements from the video uploaded by the parents using the labeling tool before participating in this work (same training for labeling work). The accuracy of the AI results was examined by comparing the agreement between the AI classification results from APP's background and labeling results in the Excel files from the physiotherapists for each movement.

2.5.6 Agreement between AI and Parents' Results

The agreement between the AI and parents results for uploaded movement indicated whether the movement assessment that parents choose to upload was classified as "passed" or "not passed" by the AI assessment. These results were examined by comparing the agreement between the parent-selected movements and the AI classification outcomes. All data were abstracted from the APP's background (Figure 4) and documented in an Excel file.

doi:10.6342/NTU202503829

2.5.7 Usability from Parents and Clinicians

Parents' and clinicians' perspectives on the use of the APP "Baby Go" were collected using Google Forms to investigate the feasibility of APP (For parents: https://forms.gle/BPL7CWopaB7qk3388 , for clinicians: https://forms.gle/3L7nGKe2ZTT39t7n6) (Table 5 & 6). The questionnaire for parents was customized for the APP "Baby Go" in this study and adapted from a survey used in the "*Baby Moves*" APP previously described by Kwong et al (Kwong et al., 2019). The questionnaire for the parents consisted of three sections: (1) frequency of the APP use, (2) benefits of the APP use, and (3) user's experience. The questionnaire for clinical providers was customized for the APP "Baby Go" in this study and adapted from a healthcare professional interview previously described by AlMahadin et al (AlMahadin et al., 2020). The questionnaire for clinicians contained two sections: (1) perspectives on APP use and (2) recommendations for APP use. Both questionnaires contained three types of measures: (1) multiple choice, (2) five-point Likert scale (1: Strongly disagree, 2: Disagree, 3: Neither agree nor disagree, 4: Agree, 5: Strongly agree), and (3) open-ended questions.

2.6 Statistical Analysis

To analyze each infant's basic and demographic data, continuous variables was presented as mean ± standard deviation and categorical variables as numbers (percentages). Analysis of continuous variables was performed using an independent T-test or Mann–Whitney U test. Comparisons of categorical variables was analyzed using the Chi-square test. Intraclass Correlation Coefficient (ICC) was used to examine the interrater reliabilities of the movement annotation (continuous variables) (Shrout & Fleiss, 1979). Inter-rater reliability for labeling all the movements from the video uploaded by the parents must reached an acceptable level (ICC > 0.85) before the three

doi:10.6342/NTU202503829

physiotherapists join the study.

The formula for the rate of APP usage was users who have uploaded at least one video divided by the total expected users who had uploaded at least one video, represented as a percentage. The formula for the upload rate for monthly assessment was the number of uploaded movements/numbers of expected uploaded movements in assessment per month, which was represented as a percentage. The formula for the upload rate per specific movement was the number of uploads/expected number of uploads for a specific movement, which was represented as a percentage. The agreement between the AI and physiotherapists' labeling results, parents' and physiotherapists' labeling results, AI and parents' results for uploaded movements was presented as percentage. The standard for agreement percentage was defined as good (>0.80), moderate (0.60–0.79), and poor (<0.60). The results of the APP's feasibility were analyzed using descriptive statistics through parents' and clinicians' questionnaires. Responses for each level or choice were reported as the percentages for multiple-choice questions or questions using a five-point Likert scale. Qualitative analysis was used to describe responses to open-ended questions. Responses were sorted into specific themes for a more concise summary and presented as percentages for each theme. Cross-tabulations was used to analyze the results of APP usage and responses from the parents' questionnaire across different groups, including full-term versus preterm infants. The chi-square test was used to determine if there was a significant relationship.

Statistical analyses were performed using the SAS (version 9.4, SAS Institute Inc., Cary, NC, USA).

doi:10.6342/NTU202503829

Chapter III. Results

3.1 Study Samples

In total, 72 preterm and 19 full-term infants were recruited in the study (Figure 5). Infant data and sociodemographic data were shown in Table 7 and 8. In the group whose parents had uploaded at least one video, preterm infants had lower birth body weight ($p<0.001$), gestational age ($p<0.001$), 1-minute and 5-minute Apgar scores ($p=0.037$ and 0.014), and a more severe level of respiratory distress syndrome ($p<0.001$) and bronchopulmonary dysplasia ($p=0.045$) compared with full-term infants. Mothers of preterm infants were older than full-term infants ($p=0.034$). A borderline significant difference was found in full-time maternal caregiving between the groups ($p=0.053$). Numbers of family members ($p = 0.005$) and numbers of children ($p = 0.006$) were significantly higher in the preterm group. The proportion of sex was similar in the preterm and full-term groups.

3.2 APP Usage

Fifty parents downloaded and registered the APP, and 45 (90.0%) of them uploaded at least one video. Among them, the proportions of parents who uploaded at least one video were 93.5% for preterm infants and 84.2% for full-term infants (Figure 5). The majority of parents began using the APP when their infants were between 3 and 5 months of age (Table 9). For the follow-up rate for each month assessment, the 3- to 5-month movement assessment had the highest complete upload rate (50.0%-54.5%), whether passed by APP or not (except the 17-month assessment with only one trial). Most parents uploaded at least one video before the 12-month assessment but failed to complete the full set of movement videos uploading. In the 13- to 15-month assessments, above 50% of parents did not upload any videos at all (Table 11). parents of preterm and full-term infants were likely to upload videos in the 1st month of APP

doi:10.6342/NTU202503829

usage and then decreased their video uploads in the following months (Table 10). Parents of full-term infants uploaded more videos during the 1st to 3rd month since APP registration than at other ages, while parents of preterm infants uploaded more videos during the 4th to 6th month since APP registration than at other ages. Parents of full-term infants uploaded more videos in the 5- and 7-month assessment, while parents of preterm infants uploaded more videos in the 8- and 10-month assessment. The upload frequency decreased after 10 months of age for both groups. For the upload rate for each movement assessment, forearm support (1), forearm support (2), supine lying (3), supine lying (4), and pull to sit had the top 5 highest upload rates (70.0%-92.3%). The video upload rate for each movement decreased as the developmental age of the movements increased. The overall upload rate ranged from the lowest to highest was standing (53.7%), sitting (58.2%), prone (63.6%), and supine movements (64.1%) (Table 12).

3.3 External Support for APP

During the study, 20 parents (45.5%) contacted the research assistant regarding why certain movements were not classified by the AI as passed and being unfamiliar with the operation of the APP, and 20 parents (45.5%) required additional reminders from the research assistant to implement the monthly video uploading.

3.4 Quality of Video Recordings

A total of 951 home videos were uploaded by the parents via the "Baby Go" APP during the study from October 2024 to June 2025. Of the uploaded videos, 891 were determined by the annotators as containing the 38 AIMS movements; whereas, 60 were not included for analysis because movement in the videos didn't fulfill the AIMS criteria (N=9), movement unrecognizable due to body parts missing (N=5), incomplete movement features (N=18), insufficient time duration (N=4), movements out of the 38

doi:10.6342/NTU202503829

items (N=9), and duplicated uploads (N=15). After removing the duplicated videos, 864 (92.3%) out of 936 videos were scored by the physiotherapists for their quality and presence of portrait mode or incorrect infant orientation (Table 13). The most common issues affecting video quality were missing body parts and unstable camera movement. Seventy-two (7.7%) videos showed missing more than 2 key joints, leading to 7 videos could not be labeled by physiotherapists, and 100 (10.7%) videos showed missing 1 key joint. Parents moved the camera slowly during the recording in 203 (21.7%) uploaded videos. Sixteen (1.7%) videos were recorded from an inappropriate angle, which made it hard for the physiotherapists to differentiate between confusing movement items. Among the 936 videos, 303 (32.4%) were recorded in portrait mode, and 142 (15.2%) of them showed incorrect infant orientation.

3.5 Agreement between Parents' and PT Results for Uploaded Movements

A total of 936 videos were included in the data analysis for assessing the agreement between parents' and physiotherapists' labeling results for 38 movements. The overall agreement between the parents' perception and physiotherapists' labeling results for 38 movements was 76.6% (Table 14). The agreement improved from APP version 2.0 to 2.1 (75.6% to 82.4%) but decreased in the later APP versions (79.5% in version 2.2 and 74.7% in version 2.3). High agreement between parents' and physiotherapists' labeling results was observed in the supine and standing subscales across all versions (all above 80%), except for the standing subscale in version 2.2, which showed a lower agreement rate (50%), likely due to the small sample size. In contrast, the prone subscale consistently showed the lowest agreement across versions, ranging from 63.6% to 75.0% (Table 14). Several movements [e.g., Forearm support (1), Extended arm support, Reaching from forearm support, Reciprocal crawling, Sitting with propped arms, Reach with rotation in sitting, Sitting without arm support (2),

doi:10.6342/NTU202503829

Cruising with rotation] demonstrated poor overall parental agreement (<60%).

3.6 Agreement between AI and PT Results for Uploaded Movements

A total of 936 videos were included in the data analysis for assessing the agreement between AI and physiotherapists' labeling results for 38 movements. The overall agreement between the AI and physiotherapists' labeling results for the 38 movements was 66.7% (Table 15). The agreement for all movement sets was higher in APP versions 2.0 (74.4%) and 2.2 (84.6%), but lower in versions 2.1 (65.3%) and 2.3 (65.0%). The supine subscale showed the highest agreement between the AI and physiotherapists' labeling results across different versions of APP (all above 80%), whereas the sitting subscale had the lowest agreement between the AI and physiotherapists' labeling results in version 2.0 (56.3%) and standing subscale had the lowest agreement in version 2.1 (52.9%), 2.2 (75.0%), and 2.3 (54.1%). Several movements [e.g. Reaching from forearm support, Four-point kneeling (1), Four-point kneeling to sitting, Reaching from extended arm support, Sitting with support, Reach with rotation in sitting, Sitting to four-point kneeling, Supported standing (3), Pull to stand with support, Pull to stand, Supported standing with rotation, Cruising with rotation, Stand alone, Standing from modified squat] got poor agreement (<60%). Subgroup analysis of agreement between the AI and physiotherapists' labeling results for each movement showed that videos with portrait mode or incorrect infant orientation had low agreement (52.3%), and videos with incorrect infant orientation had the lowest agreement (46.5%) (Table 16). After removing the videos with incorrect recording format, such as portrait mode or incorrect infant orientation, agreement improved in each subscale and overall movements (agreement for correct recording format videos: 79.3%). However, several movements [e.g., Reaching from forearm support, Reaching from extended arm support, Sitting with support, Reach with rotation in sitting,

doi:10.6342/NTU202503829

Supported standing (3), Pull to stand with support, Supported standing with rotation] still showed poor agreement in correct recording format videos (<60%).

3.7 Agreement between AI and Parents' Results for Uploaded Movements

A total of 936 videos were included in the data analysis for assessing the agreement between the AI and parental results for 38 movements. The overall agreement for all movement assessments was 61.6% (Table 17). The highest overall agreement was observed in APP version 2.2 (74.4%). The supine subscale showed the highest agreement across versions (all above 80%). On the other hand, the other subscales—prone, sitting, and standing—demonstrated poor agreement (ranging from 50.7% to 66.7%), except for the standing subscale in versions 2.0 and 2.2, which showed higher agreement rates (75%–100%), likely due to small sample sizes (Table 17). Half of 38 AIMS movements [e.g. Forearm support (1), Extended arm support, Reaching from forearm support, Four-point kneeling (1), Reciprocal crawling, Four-point kneeling to sitting, Reaching from extended arm support, Sitting with support, Reach with rotation in sitting, Sitting without arm support (2), Supported standing (3), Pull to stand with support, Pull to stand, Supported standing with rotation, Cruising with rotation, Stand alone, Standing from modified squat] showed poor agreement between parents' and AI results.

3.8 APP usability by Parents and Physiotherapists

Online survey responses were received from 30 parents who had uploaded at least one video through the APP, with two parents completing the survey twice. Details of the parental responses are presented in Table 18. Approximately 90% of parents found the "Baby Go" APP easy to use, and reported that the information provided in the parental checklist and assessment results was helpful. Over 80% also reported the home-based educational information in the APP to be beneficial. All respondents agreed or strongly

doi:10.6342/NTU202503829

agreed that the APP's design helped them better understand their infant's development.

The majority of parents (over 85%) reported spending 0–30 minutes recording the movement videos at each age interval and preferred to conduct infant movement assessments through both remote and in-person methods. Over 90% reported requiring assistance with video recording from other family members. Furthermore, 93.3% of parents found the recording instructions easy to understand. However, 73.3% of parents agreed that the video recording standards were easy to meet, and 53.4% of parents reported that capturing the required infant movements was easy. Common technical issues included incorrect recording angles (prior to the update to version 2.3), occasional inaccessible demonstration videos due to server disconnection, and the inability to review or modify uploaded videos within the APP. Parents who registered for the APP but did not upload videos reported being too busy caring for two children or already receiving early intervention services at healthcare institutes.

Online survey responses were obtained from two pediatric physiotherapists, each having over five years of clinical experience and providing early intervention services to two participants in our study. Details of the physiotherapists' responses are presented in Table 19. Both physiotherapists agreed that the "Baby Go" APP facilitated infant motor assessments in clinical settings, reduced working hours, demonstrated good accuracy in identifying motor delays, and increased parental awareness. They also expressed willingness to recommend the APP to other parents. They suggested improvements in the ability to store video recordings within the APP, the provision of an immediate feedback system, and the inclusion of multiple language options—at minimum, an English version.

doi:10.6342/NTU202503829

Chapter IV. Discussion

4.1 Usage of "Baby Go" APP

This is the first video-based smartphone application for remote motor assessment in preterm and full-term infants from 3 to 18 months of age. In this study, the proportion of parents of preterm and full-term infants who had registered "Baby Go" APP version 2 and uploaded at least one video (90.0%, SE = 4.2%, 95% CI [81.7%, 98.3%]) was significantly higher than our previous study using "Baby Go" APP version 1 (67.2%, SE = 4.3%, 95% CI [58.9%, 75.5%]) ($p<0.05$) (Tsai et al., 2025). The upload rate is also higher than that reported in studies using other remote General Movement Assessment (GMA) applications, such as the *Baby Moves* (76.3%) (Kwong et al., 2019) and *In-Motion* APPs (80.2%) (Adde et al., 2021), but lower than that reported for the *NeuroMotion* APP (100%) (Svensson et al., 2021).

Although most parents completed video uploads during the 3- to 5-month assessments, participation declined over time, with more than 30% failing to upload any videos beyond 10 months of age. This reflects reduced adherence after initial engagement with the APP. Some parents noted that certain movements were no longer present at the assessed age in the baby who showed advanced development, which made it difficult to complete the full set of movement assessments. Another possible reason is that movements at older developmental stages are more dynamic and therefore more difficult for parents to capture. This is supported by our findings showing that upload rates were higher for static movements during early infancy compared to dynamic movements in later months. For upload frequency, we found that both parents of preterm and full-term infants uploaded the most videos in the first month of APP usage, likely because they were still familiarizing themselves with the process of recording the videos. Furthermore, some parents who began using the APP after their infants reached

doi:10.6342/NTU202503829

4 months of age uploaded videos corresponding to previous monthly assessments. Parents of preterm infants tended to upload more videos when their infants reached 8 and 10 months of age, while parents of full-term infants uploaded more videos at 5 and 7 months of age. This may be because parents of full-term infants observed developmental milestones earlier than those of preterm infants. To enhance the usage of the APP, it is recommended to remove the movements that are difficult to capture without jeopardizing the validity of the age-based assessment, and to add a feature that prompts parents of babies with advanced development to upload videos of more mature movements in the future.

4.2 Quality of Video Recordings

The proportion of high-quality videos uploaded using the second version of the "Baby Go" APP was 78.3%-98.3%, representing an improvement over the video quality observed in the first version of the APP (72.3%-96%) (Hsiao, 2023). Unlike the first version, the second version provided users with immediate results from the AI action recognition model, which may have motivated parents to record higher-quality videos in order to ensure successful movement assessments. Additionally, all parents were required to view the recording guidelines before using the APP, which likely improved their understanding of how to record videos correctly. However, a small proportion of videos still showed issues such as missing body parts, shaky camera movement, or inappropriate camera angles, all of which could affect movement identification by the AI and physiotherapists. Missing body parts typically occurred when infants were partially covered by toys or furniture, or when the camera failed to capture the entire body during manual camera movement. Shaky footage was often the result of recording dynamic movements or unstable handheld shooting. We recommend that parents record from a sufficient distance to include the infant's full body, whether during static or

doi:10.6342/NTU202503829

dynamic movements. Additionally, maintaining a clean and tidy environment without unnecessary toys or furniture can help reduce visual obstructions. For certain movements, such as Reaching from forearm support, Reaching from extended arm support, and Reaching with rotation in sitting, the infant's body parts were often obscured by toys provided by the recorder, and the resulting videos frequently showed shaky footage. It is recommended that two individuals be involved in the recording process—one to handle the camera and the other to interact with the infant or provide toys—to ensure better video quality. Only a few videos were recorded from incorrect angles, which may be attributed to the additional angle-specific instructions included in the second version of the APP for certain movements.

Other issues identified in the uploaded videos included the use of portrait mode and incorrect infant orientation. Although parents were instructed to record in horizontal landscape mode, many still recorded in portrait mode, particularly when recording sitting and standing movements. According to parent feedback, this was because they were more accustomed to recording vertically with their smartphones. Some videos with incorrect infant orientation were observed prior to APP version 2.3 due to a technical issue. In earlier versions, the APP automatically flipped videos to vertical format even when parents had recorded them in horizontal mode. This issue was resolved in version 2.3. However, videos with incorrect infant orientation were still present in version 2.3. This issue may be attributed to several factors. First, some users may have continued using older versions of the APP without updating to version 2.3, in which the orientation bug was resolved. Second, malfunctions in the phone’s angle detection system during video recording may have also contributed to the problem. We found that videos with incorrect orientation had lower agreement between AI and physiotherapist results, due to poor AI recognition accuracy. It is therefore

doi:10.6342/NTU202503829

recommended that the APP incorporates a feature allowing users to manually adjust video orientation prior to upload, or alternatively, implement an automatic detection and correction system for correct recording format videos before AI-based movement identification.

4.3 Agreement between Parents' and Physiotherapists' Results

The overall parental agreement with physiotherapists' labeling results (76.6%, SE = 1.4%, 95% CI [73.9%, 79.3%]) was significantly higher than that reported in the first version of the APP (61.3%, SE = 2.8%, 95% CI [55.7%, 66.8%]) ($p<0.05$) (Hsiao, 2023). This improvement may be attributed to the refinement of movement items from 58 to 38 in the second version of the APP, retaining those that are more easily recognized by both parents and physiotherapists. However, several movements still showed moderate to poor agreement. A common pattern observed was that parents tended to upload less mature movement items, even when the infant had already demonstrated a more advanced movement in the video. For example, videos chosen by parents as Forearm support (1) were actually Forearm support (2), Extended arm support was actually Four-point kneeling (1), Reaching from forearm support was actually Reaching from four-point kneeling, and Four-point kneeling was often Reciprocal creeping. Similarly, Reciprocal crawling was frequently misclassified as Reciprocal creeping, and Sitting with propped arms was often actually Sitting with arm support or Sitting without arm support. In other cases, Reaching with rotation in sitting was often misclassified with Sitting to four-point kneeling or Sitting without arm support (2). These misclassifications may be partly due to the APP design. Most participants in this study exhibited typical or even advanced motor development. Parents of infants with advanced development tended to upload less mature movements that were shown as "in progress" on the upload page of APP. Although guiding

doi:10.6342/NTU202503829

descriptions were added to encourage parents to select and upload the more mature movement in the future, in the upload page after APP version 2.2, many parents still failed to follow this guidance. Additionally, Forearm support (2) was often misidentified by parents as Extended arm support. This may be because some infants showed transitional postures between movement categories, making it difficult for parents to distinguish between them. Another movement is Sitting without arm support (2), which was often confused with Sitting without arm support (1) or Sitting to four-point kneeling. This movement is particularly challenging to observe as it requires infants to spontaneously shift their leg positions in sitting, rather than simply changing posture. Pull to stand and Pull to stand with support were frequently misclassified. The distinction between these two movements, whether the infant uses more arm or leg force to stand, is subtle and difficult for parents to recognize. Furthermore, although the APP included reminders in the movement descriptions that the movement must include the full transition from the floor to standing, some parents uploaded incomplete movements showing only the standing phase. Lastly, several other standing movements showed moderate or poor agreement primarily due to small sample sizes, which should be further monitored with more data.

4.4 Accuracy of the AI Results Against the Physiotherapists' Results

The overall agreement between AI and physiotherapists' results was only 66.7% in our study, and several factors may have contributed to the misclassification of infant movements. Subgroup analysis revealed that videos recorded in portrait mode or with incorrect infant orientation resulted in notably lower agreement between AI and physiotherapists' labeling. This may be partly because our AI model was trained exclusively on horizontally recorded videos with correct infant orientation. However, even after removing those incorrect recording format videos, several movements

doi:10.6342/NTU202503829

continued to be misclassified by the AI model. First, the ambiguity of fine-grained movements posed a major challenge for the AI model to distinguish between visually and kinematically similar actions that differed only in subtle aspects. For example, Forearm support (1) was easily misidentified as Forearm support (2), Reach with rotation in sitting confused with other sitting movements, Supported standing (3) was misidentified as Supported standing (2), and Cruising with rotation was misclassified as Cruising without rotation. Second, occlusion in home videos adversely affected recognition. Unlike the clean lab-based videos used in training, home videos often included occlusion caused by toys, furniture, or the caregiver's hands. Such occlusion frequently occurs in movements like Reaching from forearm support, Reaching from extended arm support, and Reach with rotation in sitting, resulting in decreased AI accuracy. Third, long-tail data distribution affected the model's ability to learn certain movements. Several standing-related movements, such as Supported standing (3), Cruising with rotation, and Stand alone, had very few training samples (4-13 data samples), limiting the model's exposure and learning capacity for these categories. Fourth, the exclusion of non-infant objects during AI identification also contributed to misclassification. While this design minimized background interference, it impaired the model's ability to correctly interpret movements that inherently require external support. For instance, Sitting with support movement typically involves a caregiver holding the infant's trunk from behind. Since the supporting caregiver was excluded from detection, the AI model failed to recognize the support, often misclassifying the movement as other sitting postures.

To improve the performance of the AI model in the future, several enhancements are recommended. First, addressing data imbalance by increasing the number of training samples for underrepresented movements is essential. In addition, introducing data

doi:10.6342/NTU202503829

augmentation techniques, such as random rotation, projection, and artificial occlusion, could enhance the model's robustness against variability in real-world home video conditions. Second, enhancements in the model design may further improve recognition accuracy. One technical limitation in our current pipeline is that the 2D keypoints extracted by HRNet lack viewpoint invariance, which directly contributes to reduced accuracy in home video recognition, where camera angles and recording perspectives vary widely. To address this, one potential direction is to integrate a 2D-to-3D pose lifting model, which would transform 2D key points into more stable and viewpoint-robust 3D pose representations, serving as stronger inputs for movement classification (Pham et al., 2020). Another promising approach involves learning a prototype representation for each movement class in the feature space, where a new input would be classified based on its similarity to these learned prototypes, potentially improving the model's generalizability across varying conditions. Third, combining certain confused movement items [e.g., Forearm support (1) and (2), Cruising with and without rotation] into a single assessment category, or excluding movements that are particularly prone to occlusion [e.g., Reaching from forearm support, Reaching from extended arm support, Reaching with rotation in sitting] may be considered, providing that such modifications do not compromise the overall accuracy of the developmental assessment. Fourth, including a feature in the APP that allows parents to describe the movements they are recording, either through voice or text, may assist the AI in identifying the movements.

4.5 Accuracy of the AI Model Against the Parents' Results

The overall agreement between AI and parental results was only 61.6%, which means that only 61.6% of videos were passed by AI based on the movement assessment parents chose, but participants in our study mostly have typical motor development. It is

doi:10.6342/NTU202503829

attributed to suboptimal agreement between AI and parents' results compared with physiotherapists. The mechanism of determining whether to pass or not pass for each movement assessment in the current APP is mainly based on parents' selection of the movement and AI accuracy for each movement, which leads to many infants with typical development failing to pass the remote assessment. AI accuracy for movement assessment is the major challenge that needs to be addressed. Future work may consider simplifying the movement assessment by designing movement sets and providing video recording instructions for parents (e.g., instructing parents to place a toy in front of the baby and record a side view of the infant crawling toward it). Additionally, the AI could focus on recognizing only the movements within each group, which would help improve its accuracy.

4.6 APP Usability by Parents and Physiotherapists

To our knowledge, the "Baby Go" APP is the only application that applies remote AIMS assessments for infants aged 3 to 18 months, and this study is the first to evaluate its usability from parents and clinicians. Overall, parental responses regarding the APP's usability were satisfactory. The online survey revealed that over 90% of parents found the "Baby Go" APP easy to use, which is in accordance with findings from previous studies utilizing APP with remote General Movement Assessment (GMA) (Adde et al., 2021; Kwong et al., 2019; Svensson et al., 2021). Parents reported that the information provided through the parental checklist, the assessment results of their infant's movement videos, and the home-based educational content were helpful and using the APP helped alleviate their concerns. All parents (100%) reported that the app's design enhanced their understanding of their baby's motor development, which is significantly better than the results reported for the GMA APP (24%–48%) (Adde et al., 2021; Kwong et al., 2019; Svensson et al., 2021). The main differences between the "Baby

doi:10.6342/NTU202503829

Go" APP and other GMA-based APPs lie in the assessment tools used and the inclusion of educational content. While GMA is limited to assessing infants in a supine position and only up to 20 weeks of age, our APP enables the assessment of infant motor development throughout the first 18 months of life. It also provides detailed movement characteristics for each month. Furthermore, the home-based educational activities included in the APP help parents understand and promote their infant's motor development. The majority of parents expressed a preference for conducting motor assessments through a combination of in-person and remote approaches. This finding is consistence with a previous study (Suir et al., 2022), in which parents of preterm infants reported that home videos could serve as a supplement to follow-up visits but were not a substitute for in-clinic assessments.

However, there were some challenges that parents encountered during the APP usage. Although most of the parents reported that the instructions for recording videos are easy to understand, some parents expressed the difficulties of meeting the standard of recordings and capturing the movements from their infants. Parents provided the feedback that it is sometimes difficult to capture the movement because they are unable to know the timing of when infants' movement show up and record the video simultaneously, especially for dynamic movement. Some movements need at least two persons to prompt and record the movement separately. These limitations have been noted in the previous study using home video method (Boonzaaijer et al., 2019). Another big challenge from parents' recommendations is that no feedback system in the APP to explain why the uploaded movement is not passed by AI. This made it difficult for parents to know whether the reason for not passing is due to the infant or video quality immediately. In our study, physiotherapists reviewed the uploaded videos weekly and provided feedback to parents via messages or phone calls for movements

doi:10.6342/NTU202503829

that were not passed by the AI, which is inconvenient for physiotherapists. Meanwhile, although parents received information about which movements had not passed, they were unable to view the previously uploaded video files due to the current APP design, making it difficult to fully understand the feedback. To address these issues, it may be beneficial to implement a feedback system within the APP, allowing physiotherapists to provide information on the reasons why the movements were not passed by the AI, along with individual child's developmental suggestions, to help parents better understand the assessment results. In addition, enabling video storage or playback functions would allow parents to review their videos, facilitating a clearer understanding and potential improvement in future recordings.

Several technical issues encountered during the study negatively impacted the usability of the APP. One common problem was the unstable display of demonstration videos, which were sometimes inaccessible. This made it difficult for parents to review the movement requirements prior to recording and uploading, potentially affecting video quality and assessment outcomes. The underlying cause of this issue requires further investigation to develop an effective solution. Another limitation reported by parents was the inability to delete or modify previously uploaded videos within the APP. Addressing this concern would require the development of a secure video storage function, which could serve as a foundation for implementing editing and management features in future APP versions.

In this study, we recruited three infants with motor delays who were already receiving early intervention services through hospital or outpatient clinics. However, two of them demonstrated low adherence to APP usage. Their parents reported that physiotherapists had already provided more detailed and tailored motor practice strategies during regular in-person visits, which made the additional use of the APP

doi:10.6342/NTU202503829

seem less necessary. Moreover, parents also reported that infants with hypertonia had limited ability to perform certain movements, making it particularly difficult for the parents to facilitate or capture the required movements on video.

Clinicians' response from online survey showed overall satisfaction to the APP. Both physiotherapists agreed that the APP effectively supported infant motor assessments, reduced the time required for in-person evaluations, and enhanced parental understanding of their infant's development. It highlighted the clinical applicability of the "Baby Go" APP in real-world settings. However, two of three parents of infants with motor delays who were already receiving early intervention services through hospital or outpatient clinics demonstrated low adherence to APP usage. Their parents reported that physiotherapists had already provided more detailed and tailored motor practice strategies during regular in-person visits, which made the additional use of the APP seem less necessary. Moreover, parents also reported that infants with hypertonia had limited ability to perform certain movements, making it particularly difficult for the parents to facilitate or capture the required movements on video. These findings suggest that the APP may be better positioned as a complementary tool for early screening, ongoing parental education, and progress monitoring in infants without developmental diagnosis, or as a follow-up tool between clinic visits.

In addition to positive feedback, the physiotherapists also noted suggestions for future improvement. These included adding a video storage function within the APP, implementing a feedback system, and expanding language options to improve accessibility for diverse populations. These recommendations are aligned with the challenges reported by parents and point to the need for continued development to enhance user experience and efficiency of the "Baby Go" APP.

doi:10.6342/NTU202503829

Chapter V. Conclusion

This is the first study to investigate the usability of the smartphone application "Baby Go" version 2, which combines AI techniques for remote infant motor assessment in preterm and full-term infants from 3 to 18 months of age. A high proportion of parents uploaded at least one video using the APP, and most reported it was easy to use. However, the moderate agreement between the AI and physiotherapists' results in recognizing infant movements indicates a need to improve the AI accuracy. Future work should focus on implementing an automatic detection and correction system for incorrect recording format videos prior to AI-based movement identification. In addition, modifying the movement assessment by removing or combining certain movement items without compromising the AI model's validity may help enhance the overall usability of the APP.

doi:10.6342/NTU202503829

References

Adde, L., Brown, A., van den Broeck, C., DeCoen, K., Eriksen, B. H., Fjørtoft, T., Groos, D., Ihlen, E. A. F., Osland, S., Pascal, A., Paulsen, H., Skog, O. M., Sivertsen, W., & Støen, R. (2021). In-Motion-App for remote General Movement Assessment: A multi-site observational study. *British Medical Journal Open*, *11*(3). https://doi.org/10.1136/bmjopen-2020-042147

AlMahadin, G., Lotfi, A., Zysk, E., Siena, F. L., Carthy, M. M., & Breedon, P. (2020). Parkinson's disease: Current assessment methods and wearable devices for evaluation of movement disorder motor symptoms - a patient and healthcare professional perspective. *BioMed Central Neurology*, *20*(1). https://doi.org/10.1186/s12883-020-01996-7

Bayley, N., & Alyward, G. P. (2019). *Bayley scales of infant and toddler development.* (4th ed.). Bloomington, MN: Pearson Assessments.

Boonzaaijer, M., van Dam, E., van Haastert, I. C., & Nuysink, J. (2017). Concurrent validity between live and home video observations using the Alberta Infant Motor Scale. *Pediatric Physical Therapy*, *29*(2), 146-151. https://doi.org/10.1097/pep.0000000000000363

Boonzaaijer, M., van Wesel, F., Nuysink, J., Volman, M. J. M., & Jongmans, M. J. (2019). A home-video method to assess infant gross motor development: parent perspectives on feasibility. *BioMed Central Pediatrics*, *19*(1). https://doi.org/10.1186/s12887-019-1779-x

Caesar, R. A., Boyd, R. N., Cioni, G., Ware, R. S., Doherty, J., Jackson, M. P., Salthouse, K. L., & Colditz, P. B. (2023). Early detection of developmental delay in infants born very preterm or with very low birthweight. *Developmental Medicine and Child Neurology*, *65*(3), 346-357.

doi:10.6342/NTU202503829

https://doi.org/10.1111/dmcn.15381

Centers for Disease Control and Prevention. (2024). *Developmental monitoring and screening*. Retrieved September 29, 2024, from https://www.cdc.gov/ncbddd/actearly/screening.html

Cheong, J. L., Doyle, L. W., Burnett, A. C., Lee, K. J., Walsh, J. M., Potter, C. R., Treyvaud, K., Thompson, D. K., Olsen, J. E., Anderson, P. J., & Spittle, A. J. (2017). Association between moderate and late preterm birth and neurodevelopment and social-emotional development at age 2 years. *Journal of the American Medical Association Pediatrics*, *171*(4). https://doi.org/10.1001/jamapediatrics.2016.4805

Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F., & Campbell, J. P. (2020). Introduction to machine learning, neural networks, and deep learning. *Translational Vision Science and Technology*, *9*(2). https://doi.org/10.1167/tvst.9.2.14

Choukou, M. A., Taha, A., Qadeer, A., & Monnin, C. (2021). Digital health technology for remote care in response to the COVID-19 pandemic: A scoping review. *European Review for Medical and Pharmacological Sciences*, *25*(8), 3386-3394. https://doi.org/10.26355/eurrev_202104_25751

Dos Santos, R. S., de Lima Barreto, A. O. O., da Silva, M., HOV, S. N., Peres, R. T., Ribeiro, M. G., & Ferreira, H. C. (2022). Agreement between Alberta Infant Motor Scale assessment and maternal perception of motor development in full-term infants. *Developmental Neurorehabilitation*, *25*(6), 417-425. https://doi.org/10.1080/17518423.2022.2052373

Duan, H., Zhao, Y., Chen, K., Lin, D., & Dai, B. (2022). Revisiting skeleton-based action recognition. *Proceedings of the IEEE/CVF conference on computer vision*

doi:10.6342/NTU202503829

and pattern recognition (pp. 2969-2978). New Orleans, LA: IEEE. https://doi.org/10.1109/CVPR52688.2022.00298

Fauls, J. R., Thompson, B. L., & Johnston, L. M. (2020). Validity of the Ages and Stages Questionnaire to identify young children with gross motor difficulties who require physiotherapy assessment. *Developmental Medicine and Child Neurology*, *62*(7), 837-844. https://doi.org/10.1111/dmcn.14480

Folio, M. R., & Fewell, R. R. (2023). *Peabody developmental motor scales.* (3rd ed.). Austin, TX: Pro-Ed, Inc.

Food and Drug Administration. (2020). *What is Digital Health?* Retrieved September 29, 2024, from https://www.fda.gov/medical-devices/digital-health-center-excellence/what-digital-health

Gensure, R. H., Chiang, M. F., & Campbell, J. P. (2020). Artificial intelligence for retinopathy of prematurity. *Current Opinion in Ophthalmology*, *31*(5), 312-317. https://doi.org/10.1097/icu.0000000000000680

Grisaru-Granovsky, S., Reichman, B., Lerner-Geva, L., Boyko, V., Hammerman, C., Samueloff, A., & Schimmel, M. S. (2014). Population-based trends in mortality and neonatal morbidities among singleton, very preterm, very low birth weight infants over 16 years. *Early Human Development*, *90*(12), 821-827. https://doi.org/10.1016/j.earlhumdev.2014.08.009

Hsiao, Y. C. (2023). *Application of artificial intelligence in mobile application for infant movement assessment*. [Unpublished master's thesis]. National Taiwan University. https://doi.org/10.6342/NTU202303122

Huang, B., Xia, B., Qian, J., Zhou, X., Zhou, X., Liu, S., Chang, A., Yan, Z., Tang, Z., Xu, N., Tao, H., He, X., Yu, W., Zhang, R., Huang, R., Ni, D., & Yang, X. (2023). Artificial intelligence-assisted ultrasound diagnosis on infant

doi:10.6342/NTU202503829

developmental dysplasia of the hip under constrained computational resources. *Journal of Ultrasound in Medicine*, *42*(6), 1235-1248. https://doi.org/10.1002/jum.16133

Jin, M., & Kim, J. (2015). Development and evaluation of an evaluation tool for healthcare smartphone applications. *Telemedicine and E-Health*, *21*(10), 831-837. https://doi.org/10.1089/tmj.2014.0151

Kao, C. K., & Liebovitz, D. M. (2017). Consumer mobile health apps: Current state, barriers, and future directions. *Physical Medicine and Rehabilitation*, *9*(5S), S106-S115. https://doi.org/10.1016/j.pmrj.2017.02.018

Khan, I., & Leventhal, B. L. (2024). *Developmental Delay.* Treasure Island, FL: StatPearls Publishing, LLC.

Kwong, A. K., Eeles, A. L., Olsen, J. E., Cheong, J. L., Doyle, L. W., & Spittle, A. J. (2019). The Baby Moves smartphone app for General Movements Assessment: Engagement amongst extremely preterm and term-born infants in a state-wide geographical study. *Journal of Paediatrics and Child Health*, *55*(5), 548-554. https://doi.org/10.1111/jpc.14240

Lewis, K., Zettler-Greeley, C. M., Blake, K. V., & Milkes, A. (2023). A digital health application for managing pediatric asthma: Use and benefits. *Informatics for Health and Social Care*, *48*(4), 370-386. https://doi.org/10.1080/17538157.2023.2250435

Libertus, K., & Landa, R. J. (2013). The Early Motor Questionnaire (EMQ): A parental report measure of early motor development. *Infant Behavior and Development*, *36*(4). https://doi.org/10.1016/j.infbeh.2013.09.007

Libertus, K., & Violi, D. A. (2016). Sit to talk: Relation between motor skills and language development in infancy. *Frontiers in Psychology*, *7*.

doi:10.6342/NTU202503829

https://doi.org/10.3389/fpsyg.2016.00475

Lin, S. C., Chandra, E., Tsao, P. N., Liao, W. C., Chen, W. J., Yen, T. A., Hsu, J. Y., & Jeng, S. F. (2024). Application of artificial intelligence in infant movement classification: A reliability and validity study in infants who were full-term and preterm. *Physical Therapy*, *104*(2). https://doi.org/10.1093/ptj/pzad176

Maramba, I., Chatterjee, A., & Newman, C. (2019). Methods of usability testing in the development of eHealth applications: A scoping review. *International Journal of Medical Informatics*, *126*, 95-104. https://doi.org/10.1016/j.ijmedinf.2019.03.018

Ministry of Health and Welfare in Taiwan. (2008). *Annual report on birth notification statistics*. Retrieved August 11, 2024, from https://www.hpa.gov.tw/Pages/Detail.aspx?nodeid=649&pid=1257

Ministry of Health and Welfare in Taiwan. (2023). *Annual data of early intervention services for children with developmental disabilities*. Retrieved August 11, 2024, from https://crc.sfaa.gov.tw/Statistics/Detail/125

Ministry of Health and Welfare in Taiwan. (2024). *Annual report on birth notification statistics*. Retrieved August 11, 2024, from https://www.hpa.gov.tw/Pages/Detail.aspx?nodeid=649&pid=18408

Ministry of Health and Welfare in Taiwan. (2024). *Child development screening service program*. Retrieved August 25, 2024, from https://www.hpa.gov.tw/Pages/Detail.aspx?nodeid=129&pid=18030

Mullen, E. M. (1995). *Mullen scales of early learning.* San Antonio, TX: Pearson Assessments.

Noorbakhsh-Sabet, N., Zand, R., Zhang, Y., & Abedi, V. (2019). Artificial intelligence transforms the future of health care. *The American Journal of Medicine*, *132*(7),

doi:10.6342/NTU202503829

795-801. https://doi.org/10.1016/j.amjmed.2019.01.017

Olusanya, B. O., Davis, A. C., Wertlieb, D., Boo, N.-Y., Nair, M., Halpern, R., Kuper, H., Breinbauer, C., De Vries, P. J., & Gladstone, M. (2018). Developmental disabilities among children younger than 5 years in 195 countries and territories, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Global Health*, *6*(10), e1100-e1121.

Pascal, A., Govaert, P., Oostra, A., Naulaers, G., Ortibus, E., & Van den Broeck, C. (2018). Neurodevelopmental outcome in very preterm and very-low-birthweight infants born over the past decade: A meta-analytic review. *Developmental Medicine & Child Neurology*, *60*(4), 342-355. https://doi.org/10.1111/dmcn.13675

Pham, H. H., Salmane, H., Khoudour, L., Crouzil, A., Velastin, S. A., & Zegers, P. (2020). A Unified Deep Framework for Joint 3D Pose Estimation and Action Recognition from a Single RGB Camera. *Sensors*, *20*(7). https://doi.org/10.3390/s20071825

Piper, M., & Darrah, J. (2021). *Motor assessment of the developing infant.* (2nd, ed.). St. Louis, MO: Elsevier.

Pohl, M. (2017). *325,000 mobile health apps available in 2017–Android now the leading mHealth platform*. Research 2 Guidance. Retrieved September 29, 2024, from https://research2guidance.com/325000-mobile-health-apps-available-in-2017/

Purwanto, Y., Chandra, E., Tsao, P. N., Yen, T. A., Liao, W. C., Chen, W. J., Liao, C.-Y., Hsieh, C. W., Hsu, J. Y., & Jeng, S. F. (2025, May 29-31). *Artificial intelligence algorithms accurately classify 38 movements in infants born full-term and preterm: video recorded in the laboratory and at home* [Platform]. World

doi:10.6342/NTU202503829

Physiotherapy Congress 2025, Tokyo, Japan.

Rogers, S. J. (1975). *Early intervention developmental profile.* Austin, TX: Pro-Ed, Inc.

Ronquillo, Y., Meyers, A., & Korvek, S. J. (2017). *Digital health.* Treasure Island, FL: StatPearls Publishing, LLC.

Shahroudy, A., Liu, J., Ng, T.-T., & Wang, G. (2016). Ntu rgb+ d: A large scale dataset for 3d human activity analysis. *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1010-1019). Las Vegas, NV: IEEE. https://doi.org/10.1109/CVPR.2016.115

Shearer, D., Billingsley, J., Frohman, A., Hilliard, J., Johnson, F., & Shearer, M. (1972). *Portage guide to early education.* Portage, WI: Cooperative Educational Service Agency.

Shrestha, M., Schwinger, C., Hysing, M., Chandyo, R. K., Ulak, M., Ranjitkar, S., Kvestad, I., Sharma, S., Shrestha, L., & Strand, T. A. (2020). Agreement between mothers and fieldworkers while assessing child development using Ages and Stages Questionnaires, Third Edition in Nepal. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.579412

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428. https://doi.org/10.1037//0033-2909.86.2.420

Spittle, A., Orton, J., Anderson, P. J., Boyd, R., & Doyle, L. W. (2024). Early developmental intervention programmes provided post hospital discharge to prevent motor and cognitive impairment in preterm infants. *Cochrane Database of Systematic Reviews*, *2*(2). https://doi.org/10.1002/14651858.CD005495.pub5

Squires, J., Bricker, D. D., & Twombly, E. (2009). *Ages & stages questionnaires.* Baltimore, MD: Brookes Publishing Co.

doi:10.6342/NTU202503829

Stoll, B. J., Hansen, N. I., Bell, E. F., Walsh, M. C., Carlo, W. A., Shankaran, S., Laptook, A. R., Sánchez, P. J., Van Meurs, K. P., Wyckoff, M., Das, A., Hale, E. C., Ball, M. B., Newman, N. S., Schibler, K., Poindexter, B. B., Kennedy, K. A., Cotten, C. M., Watterberg, K. L., D'Angio, C. T., et al. (2015). Trends in care practices, morbidity, and mortality of extremely preterm neonates, 1993-2012. *Journal of the American Medical Association*, *314*(10), 1039–1051. https://doi.org/10.1001/jama.2015.10244

Su, Y. Y., Wang, S. H., Chou, H. C., Chen, C. Y., Hsieh, W. S., Tsao, P. N., & Tsou, K. I. (2016). Morbidity and mortality of very low birth weight infants in Taiwan-Changes in 15 years: A population based study. *Journal of the Formosan Medical Association*, *115*(12), 1039-1045. https://doi.org/10.1016/j.jfma.2016.10.011

Suir, I., Oosterhaven, J., Boonzaaijer, M., Nuysink, J., & Jongmans, M. (2022). The AIMS home-video method: parental experiences and appraisal for use in neonatal follow-up clinics. *BioMed Central Pediatrics*, *22*(1). https://doi.org/10.1186/s12887-022-03398-9

Sun, K., Xiao, B., Liu, D., & Wang, J. (2019). Deep high-resolution representation learning for human pose estimation. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5693-5703). Long Beach, CA: IEEE. https://doi.org/10.1109/CVPR.2019.00584

Svensson, K. A., Örtqvist, M., Bos, A. F., Eliasson, A. C., & Sundelin, H. E. (2021). Usability and inter-rater reliability of the NeuroMotion app: A tool in General Movements Assessments. *European Journal of Paediatric Neurology*, *33*, 29-35. https://doi.org/10.1016/j.ejpn.2021.05.006

Tsai, Y. H., Purwanto, Y., Chandra, E., Hsieh, C. W., Tsao, P. N., Yen, T. A., Liao, W. C.,

doi:10.6342/NTU202503829

Chen, W. J., Liao, C. Y., Hsu, Y. J., & Jeng, S. F. (2025). A Feasibility Study of a Video-Based Application by Parents of Infants Born Full-Term and Preterm. *Pediatric Physical Therapy*, *37*(3), 317-325. https://doi.org/10.1097/pep.0000000000001203

Utidjian, L., & Abramson, E. (2016). Pediatric telehealth: Opportunities and challenges. *Pediatric Clinics of North America*, *63*(2), 367-378. https://doi.org/10.1016/j.pcl.2015.11.006

Valla, L., Slinning, K., Kalleson, R., Wentzel-Larsen, T., & Riiser, K. (2020). Motor skills and later communication development in early childhood: Results from a population-based study. *Child: Care, Health and Development*, *46*(4), 407-413. https://doi.org/10.1111/cch.12765

Vogel, J. P., Chawanpaiboon, S., Moller, A. B., Watananirun, K., Bonet, M., & Lumbiganon, P. (2018). The global epidemiology of preterm birth. *Best Practice & Research Clinical Obstetrics & Gynaecology*, *52*, 3-12. https://doi.org/10.1016/j.bpobgyn.2018.04.003

World Health Organization. (1977). WHO: Recommended definitions, terminology and format for statistical tables related to the perinatal period and use of a new certificate for cause of perinatal deaths. *Acta Obstetricia et Gynecologica Scandinavica*, *56*, 247-253.

World Health Organization. (2012). *Developmental difficulties in early childhood: Prevention, early identification, assessment and intervention in low-and middle-income countries: a review*. https://iris.who.int/bitstream/handle/10665/97942/9789241503549_eng.pdf

World Health Organization. (2017). *Use of appropriate digital technologies for public health*. Retrieved September 29, 2024, from

doi:10.6342/NTU202503829

https://iris.who.int/bitstream/handle/10665/276430/A71_20-en.pdf?sequence=1&isAllowed=y

Zapata, B. C., Fernández-Alemán, J. L., Idri, A., & Toval, A. (2015). Empirical studies on usability of mHealth apps: A systematic literature review. *Journal of Medical Systems*, *39*(2). https://doi.org/10.1007/s10916-014-0182-2

doi:10.6342/NTU202503829

Tables and Figures

Table 1. The features of the "Baby Go" APP version 2.3

App features	Description
Upload videos	Allows parents to upload videos for remote infant motor assessment.
Demonstration videos and description (each movement)	Provides parents with of the knowledge regarding the characteristics of the movements and the appropriate filming technology.
In-progress movements	Displays developmental movements expected for the baby's current age (correct age), helping parents track age-appropriate milestones.
Future movements	Displays developmental movements that are likely to appear in the upcoming months, helping parents prepare for observation and recording of future milestones.
Completed movements	Displays developmental movements successfully identified as passed by the AI assessment.
Parental checklist	After selecting a movement for evaluation, parents should observe whether their baby has previously demonstrated this movement and select "Yes" or "No."
Recording	Video records the movement for assessment.
Video editor	Allows parents to trim unnecessary parts of the video, reducing the likelihood of AI misinterpretation.
Recording guidelines	Guidelines to help parents record movement videos with high quality.
User manual (Q&A)	Helps users understand how to use the APP and solve common problems.
Assessment results	Results for each month, including both parent perception and AI evaluation results.
Knowing developmental milestones	Information on infant motor milestones from 3 – 18 months.
Information for follow-up	Information for tracking outcome through in-person developmental assessments.
Recommendation for activity	Provides parents with age-appropriate motor development activities for babies at 4, 6, 9, and 12 months to do at home.
Reminder	Prompts parents to complete the remote motor assessment every two months.

doi:10.6342/NTU202503829

Table 2. Revised evaluation logic in the "Baby Go" APP version 2.2

	Movement
Less mature	Forearm supported (1)
	Forearm supported (2)
	Extended arm support
	Reaching from Forearm Support
	Four-point kneeling (1)
More mature	Reaching from Extended Arm Support
Less mature	Reciprocal Crawling
	Four-Point Kneeling to Sitting or Half-Sitting
	Reciprocal Creeping (1)
More mature	Reciprocal Creeping (2)
Less mature	Supine Lying (3)
	Supine Lying (4)
	Hands to knees
More mature	Hands to feet
Less mature	Sitting with Support
	Sitting with Propped Arms
	Sitting with Arm Support
	Sitting without Arm Support (1)
	Reach with Rotation in Sitting
More mature	Sitting without Arm Support (2)
Less mature	Supported Standing (2)
	Supported Standing (3)
	Pulls to Stand with Support
More mature	Pulls to Stand/Stands
Less mature	Supported Standing with Rotation
	Cruising without Rotation
	Cruising with Rotation
	Controlled Lowering through Standing
	Early Stepping
	Standing from Modified Squat
	Standing from Quadruped Position
More mature	Walks Alone

In each movement section (distinguished by color), if the more mature movement is assessed as passed by the APP, any less mature movements will be automatically treated as passed, whether they are uploaded or not.

doi:10.6342/NTU202503829

Table 3. Validation of two AI models incorporated into the second version of "Baby Go" APP for classifying 38 movements

Movement	AI model I				AI model II			
	Test sample	Pre	Rec	F1	Test sample	Pre	Rec	F1
Forearm support (1)	9	0.71	0.56	0.63	9	0.71	0.56	0.63
Forearm support (2)	50	0.88	0.86	0.87	50	0.88	0.86	0.87
Extend arm support	18	1.00	0.94	0.97	18	1.00	0.94	0.97
Reach from forearm support	17	0.82	0.82	0.82	17	0.82	0.82	0.82
Pivoting	9	0.62	0.56	0.59	9	0.62	0.56	0.59
Four-point kneeling (1)	11	0.77	0.91	0.83	11	0.77	0.91	0.83
Reciprocal crawling	16	0.82	0.88	0.85	16	0.82	0.88	0.85
Four-point kneeling to sitting	19	1.00	0.84	0.91	19	1.00	0.84	0.91
Reciprocal creeping (1)	17	0.71	1.0	0.83	18	0.71	1.0	0.83
Reaching from extended arm support	5	1.00	0.40	0.57	5	1.00	0.40	0.57
Reciprocal creeping (2)	9	1.0	0.78	0.88	10	1.0	0.78	0.88
Supine lying (3)	9	0.80	0.89	0.84	9	0.80	0.89	0.84
Supine lying (4)	7	0.86	0.86	0.86	7	0.86	0.86	0.86
Hands to knees	8	0.89	1.00	0.94	8	0.89	1.00	0.94
Hands to feet	13	1.00	0.92	0.96	13	1.00	0.92	0.96
Rolling supine to prone with rotation	8	1.00	0.75	0.86	8	1.00	0.75	0.86
Sitting with support	2	1.00	0.50	0.67	2	1.00	0.50	0.67
Sitting with propped arms	5	1.00	0.80	0.89	5	1.00	0.80	0.89
Pull to sit	1	1.00	1.00	1.00	1	1.00	1.00	1.00
Sitting with arm support	21	0.88	1.00	0.93	21	0.88	1.00	0.93
Sitting without arm support (1)	43	0.86	0.98	0.91	44	0.86	0.98	0.91
Reach with rotation in sitting	2	1.00	1.00	1.00	2	1.00	1.00	1.00
Sitting to four-point kneeling	23	0.83	0.87	0.85	24	0.83	0.87	0.85
Sitting without arm support (2)	6	0.50	0.17	0.25	6	0.50	0.17	0.25
Supported standing (2)	4	0.30	0.75	0.43	4	0.30	0.75	0.43
Supported standing (3)	1	0.00	0.00	0.00	1	0.00	0.00	0.00
Pull to stand with support	5	1.00	0.8	0.89	5	1.00	0.8	0.89
Pulls to stand/stands	11	0.75	0.82	0.78	11	0.75	0.82	0.78
Supported standing with rotation	7	0.80	0.57	0.67	8	0.80	0.57	0.67
Cruising without rotation	15	0.70	0.93	0.80	15	0.70	0.93	0.80
Controlled lowering through standing	9	1.00	0.78	0.88	9	1.00	0.78	0.88
Cruising with rotation	5	1.00	0.40	0.57	5	1.00	0.40	0.57
Stands alone	13	0.92	0.85	0.88	13	0.92	0.85	0.88
Early stepping	21	0.85	0.81	0.83	21	0.85	0.81	0.83
Standing from modified squat	5	1.00	0.60	0.75	5	1.00	0.60	0.75
Standing from quadruped position	13	0.87	1.00	0.93	13	0.87	1.00	0.93
Walks alone	19	0.79	0.79	0.79	20	0.79	0.79	0.79
Squat	5	1.00	0.80	0.89	5	1.00	0.80	0.89
All Movements								
Accuracy	461			0.84	467			0.87
Macro average	461	0.84	0.77	0.78	467	0.85	0.82	0.82

The AI model I was incorporated into the APP before January 14, 2025, trained on 1,419 home videos (Purwanto et al., 2025).
The AI model II was incorporated into the APP after January 14, 2025, trained on 474 lab videos and 1,419 home videos.
Pre: Precision; Rec: Recall; F: F1 score

doi:10.6342/NTU202503829

Table 4. Criteria for assessment of video quality

	Infant body parts	Camera angle	Camera movement
Poor	More than two key points were missing in the video or covered by objects.	Video was recorded at an inappropriate angle and can't be recognized by PT.	Camera movement was shaky and couldn't be recognized by PT.
Fair	Less than two infant key points were missing in the video or covered by objects.	PT recognized the movement but hardly differentiated between confusing items due to the incorrect angle.	Camera movement was slowly moved but could be recognized by PT.
High	Complete infant key points were shown in the video.	Appropriate camera angle	Camera movement was stable.

PT, Physical Therapist; AI, Artificial Intelligence

Confusing items include forearm support (1) and (2), supported standing (2) and (3), reciprocal creeping (1) and (2)

Human key points include nose, right/left ear, right/left shoulder, right/left elbow, right/left wrist, right/left hip, right/left knee, right/left ankle, top/base of the spine

doi:10.6342/NTU202503829

Table 5. Contents of the “Baby Go” usability questionnaire for parents

基本資料	
● 寶寶受試者代號：	
● 寶寶性別：	☐ 男 ☐ 女
● 寶寶為足月兒或早產兒？	☐ 足月兒 ☐ 早產兒
A 應用程式的使用狀況	
1. 自參加計畫以來，我經常使用本款應用程式來評估寶寶的發展。	☐ 曾經使用（進入 1-1.） ☐ 不曾使用（進入 1-2.）
1-1. 使用本款應用程式的優點是：(可多選)	☐有趣 ☐方便 ☐專業
1-2. 您未曾使用本款應用程式的原因是：(可多選)	☐工作太忙 ☐家務太忙 ☐寶寶動作影片難掌握 ☐寶寶發展正常，無須相關資訊 ☐其他，請具體說明：__________
B 使用應用程式的幫助和風險	
2. 家長自評表的資訊有幫助。	☐非常不同意 ☐不同意 ☐普通 ☐同意 ☐非常同意
3. 寶寶動作影片的評估結果有幫助。	☐非常不同意 ☐不同意 ☐普通 ☐同意 ☐非常同意
4. 居家衛教的資訊有幫助。	☐非常不同意 ☐不同意 ☐普通 ☐同意 ☐非常同意
5. 整體設計對於促進寶寶發展的了解有幫助。	☐非常不同意 ☐不同意 ☐普通 ☐同意 ☐非常同意
6. 應用程式提供的資訊讓我更擔心寶寶的發展。	☐非常不同意 ☐不同意 ☐普通 ☐同意 ☐非常同意
7. 影片資料的傳送方式是安全可靠的。	☐非常不同意 ☐不同意 ☐普通 ☐同意 ☐非常同意
8. 針對寶寶的動作發展評估，我比較傾向採用以下何種方式？	▫ 僅由醫療院所的專業人員進行實體評估 ▫ 僅由本款應用程式的遠端評估 ▫ 實體和遠端兩者同時進行評估
C 操作應用程式的心得	
9. 本款應用程式容易使用。	☐非常不同意 ☐不同意 ☐普通 ☐同意 ☐非常同意
10. 我平均花多少時間錄製寶寶一個年紀的動作影片？	☐ 0-30 分鐘 ☐ 30-60 分鐘 ☐ 60-90 分鐘 ☐ >90 分鐘
11. 當我拍攝寶寶動作影片時，是否有其他人在旁協助？	☐ 無 ☐ 爸爸或媽媽 ☐ 其他家人
12. 影片錄製的指導說明容易理解。	☐非常不同意 ☐不同意 ☐普通 ☐同意 ☐非常同意
13. 影片錄製的標準容易達到。	☐非常不同意 ☐不同意 ☐普通 ☐同意 ☐非常同意
14. 影片中的姿勢和動作技能容易捕捉。	☐非常不同意 ☐不同意 ☐普通 ☐同意 ☐非常同意
15. 錄製寶寶影片時，何項因素常影響其動作表現？(可多選)	▫ 無 ▫ 環境噪音 ▫ 寶寶身體不適 ▫ 寶寶心情不佳 ▫ 有家人在場 ▫ 其他，請具體說明：__
16. 操作本款應用程式是否曾遇到技術問題？	☐ 無 ☐ 有，請具體說明：__
17. 提醒上傳影片的時間點有幫助。	☐非常不同意 ☐不同意 ☐普通 ☐同意 ☐非常同意
18. 提醒上傳影片的次數是合適的。	☐非常不同意 ☐不同意 ☐普通 ☐同意 ☐非常同意
19. 您對這款應用程式還有其他建議或意見嗎？	

doi:10.6342/NTU202503829

Table 6. Contents of the “Baby Go” usability questionnaire for physiotherapists

基本資料	
● 您的職稱是？	☐ 物理治療師 ☐ 醫師 ☐ 其他，請說明：__
● 您的工作經驗是？	☐ 1 年以下 ☐ 1-3 年 ☐ 3-5 年 ☐ 5 年以上
● 您主要服務的部門是？	☐ 小兒部 ☐ 物理治療中心 ☐ 復健部 ☐ 其他，請說明：__
A 對本款應用程式的看法	
1. 這款應用程式對評估寶寶動作發展的工作有幫助。	☐非常不同意 ☐不同意 ☐普通 ☐同意 ☐非常同意
2. 這款應用程式提供的評估結果準確性高。	☐非常不同意 ☐不同意 ☐普通 ☐同意 ☐非常同意
3. 這款應用程式能夠減少臨床嬰幼兒動作發展評估的時間和工作量。	☐非常不同意 ☐不同意 ☐普通 ☐同意 ☐非常同意
4. 這款應用程式能夠提升家長對嬰幼兒動作發展的認識與重視。	☐非常不同意 ☐不同意 ☐普通 ☐同意 ☐非常同意
5. 您是否願意推薦這款應用程式給其他家長使用？	☐ 是 ☐ 否
B 其他建議	
6. 您認為本款應用程式有哪些方面還需要改進？(可多選)	☐ 拍攝指南 ☐ 動作敘述 ☐ 發展結果 ☐ 介面設計 ☐ 其他，請具體說明：__
7. 您認為本款應用程式還可以增加哪些方面的功能？(可多選)	☐ 無 ☐ 保留影片紀錄 ☐ 線上即時通訊 ☐ 多語言版本 ☐ 其他，請具體說明：__
8. 您對這款應用程式還有其他建議或意見嗎？	

doi:10.6342/NTU202503829

Table 7. Infants' perinatal characteristics

Characteristics	Infants whose parent uploaded at least one video		Infants whose parent did not upload video	
	Preterm (N=29)	**Full-Term (N=16)**	**Preterm (N=2)**	**Full-Term (N=3)**
Female sex	16 (55.2%)	10 (62.5%)	0 (0%)	1 (33.3%)
Birth body weight (grams)	1756.1±361.6*	2949.1±268.9*	1268.0±265.9	3110.0±295.5
Gestational age (weeks)	33.4 ± 2.1*	38.6 ± 0.8*	32.6 ± 2.0	39.1 ± 0.8
Small gestational age	4 (13.8%)	2 (12.5%)	1 (50.0%)	0 (0.0%)
1-minute Apgar score	6.9 ± 2.2*	8.2 ± 1.5*	5.0 ± 0.0	9.0 ± 0.0
5-minute Apgar score	8.3 ± 1.3*	8.9 ± 0.3*	7.0 ± 0.0	9.0 ± 0.0
Respiratory distress syndrome				
None	7 (24.1%)*	16 (100%)*	1 (50.0%)	3 (100.0%)
Grade I	16 (55.2%)*	0 (0.0%)*	0 (0.0%)	0 (0.0%)
Grade II	4 (13.8%)*	0 (0.0%)*	1 (50.0%)	0 (0.0%)
Grade III	2 (6.9%)*	0 (0.0%)*	0 (0.0%)	0 (0.0%)
Bronchopulmonary dysplasia				
None	20 (69.0%)*	16 (100%)*	1(50.0%)	3 (100.0%)
Mild	4 (13.8%)*	0 (0.0%)*	1 (50.0%)	0 (0.0%)
Moderate	5 (17.2%)*	0 (0.0%)*	0 (0.0%)	0 (0.0%)
Severe	0 (0.0%)*	0 (0.0%)*	0 (0.0%)	0 (0.0%)
Intraventricular hemorrhage				
Normal	28 (96.6%)	16 (100%)	2 (100.0%)	3 (100.0%)
Grade I-II	1 (3.4%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Grade III-IV	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Periventricular leukomalacia	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Patent ductus arteriosus with ligation	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Retinopathy of prematurity				
Stage 0	29 (100%)	16 (100%)	2 (100.0%)	3 (100.0%)
Stage I-II	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
>Stage III	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Necrotizing enterocolitis	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Neonatal jaundice				
Normal	19 (65.5%)	14 (93.3%)	0 (0.0%)	2 (66.7%)
Under monitor	7 (24.1%)	1 (6.7%)	2 (100.0%)	1 (33.3%)
Use of phototherapy	3 (10.4%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

Data presented as mean ± SD or number (%).
* Preterm infants vs. full-term infants, $p<0.05$.
The data was updated until 2025.06.30.

doi:10.6342/NTU202503829

Table 8. Infants' sociodemographic data

Characteristics	Infants whose parents uploaded at least one video		Infants whose parents did not upload video	
	Preterm (N=29)	**Full-Term (N=16)**	**Preterm (N=2)**	**Full-Term (N=3)**
Paternal age (years)	41.0 ± 7.2	37.7 ± 3.6	37.0 ± 1.4	37.3 ± 13.3
Paternal education				
College or above	28 (96.6%)	15 (93.8%)	1 (50.0%)	3 (100.0%)
Senior high school	1 (3.4%)	1 (6.2%)	1 (50.0%)	0 (0.0%)
Middle school or below	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Paternal occupation				
Professional	22 (75.9%)	14 (87.5%)	1 (50.0%)	1 (33.3%)
Technological	5 (17.2%)	2 (12.5%)	1 (50.0%)	2 (66.7%)
Labor or homemaker	2 (6.9%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Maternal age (years)	38.5 ± 4.3*	35.8 ± 3.5*	36.5 ± 2.1	34.7 ± 11.2
Maternal education				
College or above	27 (93.1%)	16 (100.0%)	2 (100.0%)	2 (66.7%)
Senior high school	2 (6.9%)	0 (0.0%)	0 (0.0%)	1 (33.3%)
Middle school or below	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Maternal occupation				
Professional	17 (58.6%)	10 (62.4%)	2 (100.0%)	3 (100.0%)
Technological	3 (10.3%)	3 (18.8%)	0 (0.0%)	0 (0.0%)
Labor or homemaker	9 (31.1%)	3 (18.8%)	0 (0.0%)	0 (0.0%)
Annual household income (NTD)				
> 1,500,000	10 (34.5%)	10 (62.5%)	1 (50.0%)	2 (66.7%)
1,000,000-1,500,000	11 (37.9%)	4 (25.0%)	0 (0.0%)	1 (33.3%)
< 1,000,000	8 (27.6%)	2 (12.5%)	1 (50.0%)	0 (0.0%)
Parent as caregiver	25 (86.2%)	12 (75.0%)	2 (100.0%)	2 (66.7%)
Full-time maternal caregiving	23 (79.3%)§	8 (50.0%)§	1 (50.0%)	2 (66.7%)
Family size	4.9 ± 1.7*	3.6 ± 1.0*	4.0 ± 1.4	4.3 ± 2.3*
Number of children	2.0 ± 0.7*	1.4 ± 0.5*	1.0 ± 1.4	0.3 ± 0.6*
First child	12 (41.4%)	10 (62.5%)	1 (50.0%)	2 (66.7%)
Parental leave taken	21 (72.4%)	10 (62.5%)	1 (50.0%)	2 (66.7%)
Infant age when parents return to work (month)	4.7 ± 2.9	9.3 ± 3.9	-	-

Data presented as mean ± SD or number (%).

* Preterm infants vs. full-term infants, $p<0.05$. §Preterm infants vs. full-term infants, $p=0.05$.

One father of a twin in the preterm group had passed away during the mother's pregnancy.

The data was updated until 2025.06.30.

doi:10.6342/NTU202503829

Table 9. Number of first-month APP users

Assessment age	Number of first-month APP users		
	Preterm	**Full-term**	**All**
3	9	4	13
4	8	3	11
5	5	4	9
6	2	3	5
7	1	1	2
8	0	0	0
9	3	0	3
10	1	1	2
Total	**29**	**16**	**45**

No first-month user after 10 months of age.

doi:10.6342/NTU202503829

Table 10. Average number of uploaded videos from parents

Time	Preterm			Full-term			All		
	No. of users	No. of uploaded videos	Min-Max	No. of users	No. of uploaded videos	Min-Max	No. of users	No. of uploaded videos	Min-Max
Since APP registration									
1st	29	5.6 ± 4.3	1-19	16	6.3 ± 5.3	1-19	45	5.8 ± 4.6	1-19
2nd	24	2.6 ± 3.0	0-11	15	3.5 ± 2.5	0-8	39	3.0 ± 2.8	0-11
3rd	21	3.1 ± 3.4	0-12	14	3.9 ± 4.5	0-17	35	3.5 ± 3.8	0-17
4th	19	4.1 ± 5.7	0-19	13	3.2 ± 3.0	0-8	33	3.7 ± 4.7	0-19
5th	17	3.9 ± 3.7	0-12	12	3.2 ± 2.9	0-8	29	3.6 ± 3.4	0-12
6th	14	5.5 ± 8.2	0-21	11	3.1 ± 4.4	0-13	25	4.4 ± 6.8	0-21
7th	9	2.6 ± 3.2	0-9	7	2.7 ± 2.6	0-6	16	2.6 ± 2.9	0-9
8th	3	1.0 ± 1.7	0-3	3	5.3 ± 3.8	1-8	6	3.2 ± 3.5	0-8
Monthly assessment									
3	9	3.3 ± 1.8	1-6	4	3.3 ± 1.5	1-4	13	3.3 ± 1.7	1-6
4	15	3.8 ± 2.9	0-9	7	4.3 ± 2.0	2-8	22	4.0 ± 2.6	0-9
5	19	4.6 ± 3.6	0-14	11	7.5 ± 5.2	0-17	30	5.7 ± 4.4	0-17
6	18	3.8 ± 4.1	0-12	14	3.8 ± 3.0	0-9	32	3.8 ± 3.6	0-12
7	16	4.2 ± 4.3	0-12	14	4.9 ± 5.0	0-19	30	4.5 ± 4.6	0-19
8	10	7.4 ± 8.3	0-21	12	3.1 ± 2.8	0-8	22	5.0 ± 6.2	0-21
9	12	2.9 ± 2.8	0-8	9	3.0 ± 4.3	0-13	21	3.0 ± 3.4	0-13
10	10	8.0 ± 8.5	0-19	8	3.4 ± 3.1	0-7	18	5.9 ± 6.9	0-19
11	8	2.4 ± 2.7	0-8	6	2.7 ± 3.5	0-8	14	2.5 ± 2.9	0-8
12	5	1.4 ± 2.1	0-5	2	0.5 ± 0.7	0-1	7	1.1 ± 1.8	0-5
13	5	0.4 ± 0.9	0-2	1	0.0	0-0	6	0.3 ± 0.8	0-2
14	4	0.3 ± 0.5	0-1	1	2.0	2-2	5	0.6 ± 0.9	0-2
15	2	1.0 ± 1.4	0-2	1	0.0	0-0	3	0.7 ± 1.2	0-2
16	2	3.0 ± 4.2	0-6	1	0.0	0-0	3	2.0 ± 3.5	0-6
17	1	0.0	0-0	0	-		1	0.0	0-0

Data presented as mean ± SD. The data was updated until 2025.06.30.

doi:10.6342/NTU202503829

Table 11. Follow-up rate for each monthly assessment in APP

Monthly assessment	Number of first-month APP users	Number of trials	Upload all and passed	Upload all but not passed	Incomplete upload	Did not upload at all
3	13	13	3 (23.1%)	4 (30.8%)	5 (38.5%)	1 (7.7%)
4	11	22	7 (31.8%)	5 (22.7%)	9 (40.9%)	1 (4.5%)
5	9	30	12 (40.0%)	3 (10.0%)	13 (43.3%)	2 (6.7%)
6	5	32	8 (25.0%)	2 (6.3%)	16 (50.0%)	6 (18.8%)
7	2	30	7 (23.3%)	5 (16.7%)	10 (33.3%)	8 (26.7%)
8	0	22	3 (13.6%)	5 (22.7%)	10 (45.5%)	4 (18.2%)
9	3	21	3 (14.3%)	5 (23.8%)	7 (33.3%)	6 (28.6%)
10	2	18	3 (16.7%)	3 (16.7%)	6 (33.3%)	6 (33.3%)
11	0	14	1 (7.1%)	3 (21.4%)	4 (28.6%)	6 (42.9%)
12	0	7	2 (28.6%)	0 (0.0%)	2 (28.6%)	3 (42.9%)
13	0	6	1 (16.7%)	0 (0.0%)	2 (33.3%)	3 (50.0%)
14	0	5	1 (20.0%)	0 (0.0%)	0 (0.0%)	4 (80.0%)
15	0	3	1 (33.3%)	0 (0.0%)	0 (0.0%)	2 (66.7%)
16	0	3	1 (33.3%)	0 (0.0%)	1 (33.3%)	1 (33.3%)
17	0	1	1 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Total	45	227	54 (24.1%)	35 (15.4%)	85 (37.3%)	53 (23.2%)

This table included the parents who had uploaded at least one video to the APP.
Infants who haven't reached the end day of the current monthly assessment were not included in this table.
The data was updated until 2025.06.30.

doi:10.6342/NTU202503829

Table 12. Upload rate for each movement assessment in APP

Movement item	Number of APP users	Number of trials	Upload rate
Prone subscale (movements)			
Forearm support (1)	24	35	32 (91.4%)
Forearm support (2)	30	30	26 (86.7%)
Extended arm support	32	32	21 (65.6%)
Reaching from forearm support	30	30	17 (56.7%)
Pivoting	22	22	12 (54.5%)
Four-point kneeling (1)	21	21	10 (47.6%)
Reciprocal crawling	18	18	10 (55.6%)
Four-point kneeling to sitting or half-sitting	18	18	9 (50.0%)
Reciprocal creeping (1)	14	14	7 (50.0%)
Reaching from extended arm support	14	14	5 (35.7%)
Reciprocal creeping (2)	8	13	8 (61.5%)
Prone subscale (total)	**231**	**247**	**157 (63.6%)**
Supine subscale (movements)			
Supine lying (3)	13	13	12 (92.3%)
Supine lying (4)	22	22	17 (77.3%)
Hands to knees	30	30	20 (66.7%)
Hands to feet	42	62	40 (64.5%)
Rolling supine to prone with rotation	30	43	20 (46.5%)
Supine subscale (total)	**137**	**170**	**109 (64.1%)**
Sitting subscale (movements)			
Sitting with support	24	35	23 (65.7%)
Sitting with propped arms	30	30	20 (66.7%)
Pull to sit	30	30	21 (70.0%)
Sitting with arm support	42	62	35 (56.5%)
Sitting without arm support (1)	22	22	14 (63.6%)
Reach with rotation in sitting	22	22	8 (36.4%)
Sitting to four-point kneeling	21	21	9 (42.9%)
Sitting without arm support (2)	26	39	22 (56.4%)
Sitting subscale (total)	**217**	**261**	**152 (58.2%)**
Standing subscale (movements)			
Supported standing (2)	51	114	75 (65.8%)
Supported standing (3)	21	21	10 (47.6%)
Pulls to stand with support	21	21	12 (57.1%)
Pulls to stand	18	18	9 (50.0%)
Supported standing with rotation	18	18	8 (44.4%)
Cruising without rotation	15	21	10 (47.6%)
Controlled lowering through standing	7	7	3 (42.9%)
Cruising with rotation	6	6	0 (0.0%)
Stands alone	8	11	2 (18.2%)
Early stepping	5	5	1 (20.0%)
Standing from modified squat	3	3	1 (33.3%)
Standing from quadruped position	3	3	1 (33.3%)
Walks alone	6	6	4 (66.7%)
Squat	4	3	2 (66.7%)
Standing subscale (total)	**186**	**257**	**138 (53.7%)**

The upload rate was defined as the proportion of videos uploaded on or before the scheduled age for each movement assessment. Because some movements are scheduled for assessment at more than one time (month), the APP user may need to upload these movements multiple times.
Data presented as number (%). The data was updated until 2025.06.30.

doi:10.6342/NTU202503829

Table 13. Quality of video recordings (N=936)

Quality	Infant body part	Camera angle	Camera movement	Portrait mode	Incorrect infant orientation
High	764 (81.6%)	920 (98.3%)	733 (78.3%)		
Fair	100 (10.7%)	16 (1.7%)	203 (21.7%)		
Poor	72 (7.7%)	0 (0.0%)	0 (0.0%)		
Absent				633 (67.6%)	794 (84.8%)
Present				303 (32.4%)	142 (15.2%)

Data presented as number (%).
The data was updated until 2025.06.30.

doi:10.6342/NTU202503829

Table 14. Agreement between parents' and physiotherapists' labeling results for each movement across different versions of APP

Movement item	Ver 2.0		Ver 2.1		Ver 2.2		Ver 2.3 (Current)		Overall	
	Number of trials	Agreement	Number of trials	Agreement	Number of trials	Agreement	Number of trials	Agreement	Number of trials	Agreement
Forearm support (1)	16	8 (50.0%)	23	15 (65.2%)	6	4 (66.7%)	18	5 (27.8%)	63	32 (50.8%)
Forearm support (2)	1	1 (100.0%)	14	12 (85.7%)	3	3 (100.0%)	22	17 (77.3%)	40	33 (82.5%)
Extended arm support	8	4 (50.0%)	14	8 (57.1%)	2	2 (100.0%)	27	12 (44.4%)	51	26 (51.0%)
Reaching from forearm support	1	1 (100.0%)	4	3 (75.0%)	3	2 (66.7%)	21	11 (52.4%)	29	17 (58.6%)
Pivoting	1	1 (100.0%)	2	1 (50.0%)	1	1 (100.0%)	15	12 (80.0%)	19	15 (78.9%)
Four-point kneeling (1)	1	1 (100.0%)	2	2 (100.0%)	0	-	23	15 (65.2%)	26	18 (69.2%)
Reciprocal crawling	1	1 (100.0%)	5	2 (40.0%)	1	0(0.0%)	14	6 (42.9%)	21	9 (42.9%)
Four-point kneeling to sitting or half-sitting	1	1 (100.0%)	1	1 (100.0%)	0	-	22	20 (90.9%)	24	22 (91.7%)
Reciprocal creeping (1)	1	1 (100.0%)	1	1 (100.0%)	0	-	15	12 (80.0%)	17	14 (82.4%)
Reaching from extended arm support	1	1 (100.0%)	2	2 (100.0%)	0	-	21	16 (76.2%)	24	19 (79.2%)
Reciprocal creeping (2)	1	1 (100.0%)	1	1 (100.0%)	0	-	16	14 (87.5%)	18	16 (88.9%)
Prone subscale (total)	**33**	**21 (63.6%)**	**69**	**48 (69.6%)**	**16**	**12(75.0%)**	**214**	**140 (65.4%)**	**332**	**221 (66.6%)**
Supine lying (3)	10	9 (90.0%)	15	15 (100.0%)	2	2(100.0%)	9	6 (66.7%)	36	32 (88.9%)
Supine lying (4)	6	3 (50.0%)	15	15 (100.0%)	4	4(100.0%)	16	15 (93.8%)	41	37 (90.2%)
Hands to knees	2	2 (100.0%)	7	4 (57.1%)	2	2(100.0%)	16	15 (93.8%)	27	23 (85.2%)
Hands to feet	4	4 (100.0%)	7	6 (85.7%)	1	1(100.0%)	21	21 (100.0%)	33	32 (97.0%)
Rolling supine to prone with rotation	2	2 (100.0%)	2	2 (100.0%)	1	1(100.0%)	21	20 (95.2%)	26	25 (96.2%)
Supine subscale (total)	**24**	**20 (83.3%)**	**46**	**42 (91.3%)**	**10**	**10(100.0%)**	**83**	**77 (92.8%)**	**163**	**149 (91.4%)**
Sitting with support	9	8 (88.9%)	23	23 (100.0%)	6	6(100.0%)	14	14 (100.0%)	52	51 (98.1%)
Sitting with propped arms	2	1 (50.0%)	14	8 (57.1%)	1	0(0.0%)	24	14 (58.3%)	41	23 (56.1%)
Pull to sit	1	1 (100.0%)	9	9 (100.0%)	0	-	20	19 (95.0%)	30	29 (96.7%)
Sitting with arm support	1	1 (100.0%)	9	8 (88.9%)	0	-	28	24 (85.7%)	38	33 (86.8%)
Sitting without arm support (1)	1	0 (0.0%)	8	7 (87.5%)	1	1(100.0%)	30	25 (83.3%)	40	33 (82.5%)
Reach with rotation in sitting	0	-	1	1(100.0%)	0	-	31	16 (51.6%)	32	17 (53.1%)
Sitting to four-point kneeling	1	1(100.0%)	2	2(100.0%)	0	-	23	18 (78.3%)	26	21 (80.8%)
Sitting without arm support (2)	1	1(100.0%)	1	0 (0.0%)	1	0(0.0%)	20	8 (40.0%)	23	9 (39.1%)
Sitting subscale (total)	**16**	**13 (81.3%)**	**67**	**58(86.6%)**	**9**	**7(77.8%)**	**190**	**138 (72.6%)**	**282**	**216 (76.6%)**

Major changes for each version: wording revision for recording time duration in version 2.1; revision for evaluation logic and wording of some movements in version 2.2; issue of incorrect infant orientation due to automatic video rotation, which caused the infant to be misaligned, was fixed in version 2.3.
Data presented as number (%). The data was updated until 2025.06.30.
(To be continued)

doi:10.6342/NTU202503829

Table 14. Agreement between parents' and physiotherapists' labeling results for each movement across different versions of APP (Con.)

Movement item	Ver 2.0		Ver 2.1		Ver 2.2		Ver 2.3 (Current)		Overall	
	Number of trials	Agreement	Number of trials	Agreement	Number of trials	Agreement	Number of trials	Agreement	Number of trials	Agreement
Supported standing (2)	1	1(100.0%)	9	9(100.0%)	2	1(50.0%)	18	18 (100.0%)	30	29 (96.7%)
Supported standing (3)	1	1(100.0%)	0	-	1	0(0.0%)	11	9 (81.8%)	13	10 (76.9%)
Pulls to stand with support	0	-	3	3(100.0%)	0	-	22	15 (68.2%)	25	18 (72.0%)
Pulls to stand	1	1(100.0%)	0	-	0	-	26	19 (73.1%)	27	20 (74.1%)
Supported standing with rotation	1	1(100.0%)	1	1(100.0%)	0	-	14	11 (78.6%)	16	13 (81.3%)
Cruising without rotation	1	1(100.0%)	0	-	0	-	11	8 (72.7%)	12	9 (75.0%)
Controlled lowering through standing	0	-	2	1(50.0%)	0	-	9	8 (88.9%)	11	9 (81.8%)
Cruising with rotation	0	-	0	-	1	1(100.0%)	3	1 (33.3%)	4	2 (50.0%)
Stands alone	0	-	2	2(100.0%)	0	-	5	5 (100.0%)	7	7 (100.0%)
Early stepping	0	-	0	-	0	-	4	4 (100.0%)	4	4 (100.0%)
Standing from modified squat	0	-	0	-	0	-	3	3 (100.0%)	3	3 (100.0%)
Standing from quadruped position	0	-	0	-	0	-	3	3 (100.0%)	3	3 (100.0%)
Walks alone	0	-	0	-	0	-	2	2 (100.0%)	2	2 (100.0%)
Squat	0	-	0	-	0	-	2	2 (100.0%)	2	2 (100.0%)
Standing subscale (total)	**5**	**5(100.0%)**	**17**	**16(94.1%)**	**4**	**2(50.0%)**	**133**	**108 (81.2%)**	**159**	**131 (82.4%)**
Overall	**78**	**59(75.6%)**	**199**	**164(82.4%)**	**39**	**31(79.5%)**	**620**	**463 (74.7%)**	**936**	**717 (76.6%)**

Major changes for each version: wording revision for recording time duration in version 2.1; revision for evaluation logic and wording of some movements in version 2.2; issue of incorrect infant orientation due to automatic video rotation, which caused the infant to be misaligned, was fixed in version 2.3.
Data presented as number (%). The data was updated until 2025.06.30.

doi:10.6342/NTU202503829

Table 15. Agreement between AI and physiotherapists' labeling results for each movement across different versions of APP

Movement item	Ver 2.0		Ver 2.1		Ver 2.2		Ver 2.3 (Current)		Overall	
	Number of trials	Agreement	Number of trials	Agreement	Number of trials	Agreement	Number of trials	Agreement	Number of trials	Agreement
Forearm support (1)	16	11 (68.8%)	23	14 (60.9%)	6	5 (83.3%)	18	13 (72.2%)	63	43 (68.3%)
Forearm support (2)	1	1 (100.0%)	14	12 (85.7%)	3	3 (100.0%)	22	15 (68.2%)	40	31 (77.5%)
Extended arm support	8	5 (62.5%)	14	8 (57.1%)	2	2 (100.0%)	27	25 (92.6%)	51	40 (78.4%)
Reaching from forearm support	1	1 (100.0%)	4	3 (75.0%)	3	1 (33.3%)	21	9 (42.9%)	29	14 (48.3%)
Pivoting	1	1 (100.0%)	2	2 (100.0%)	1	1 (100.0%)	15	13 (86.7%)	19	17 (89.5%)
Four-point kneeling (1)	1	0 (0.0%)	2	2 (100.0%)	0	-	23	12 (52.2%)	26	14 (53.8%)
Reciprocal crawling	1	1 (100.0%)	5	3 (60.0%)	1	0 (0.0%)	14	9 (64.3%)	21	13 (61.9%)
Four-point kneeling to sitting or half-sitting	1	1 (100.0%)	1	1 (100.0%)	0	-	22	12 (54.5%)	24	14 (58.3%)
Reciprocal creeping (1)	1	1 (100.0%)	1	0 (0.0%)	0	-	15	12 (80.0%)	17	13 (76.5%)
Reaching from extended arm support	1	0 (0.0%)	2	0 (0.0%)	0	-	21	4 (19.0%)	24	4 (16.7%)
Reciprocal creeping (2)	1	1 (100.0%)	1	1 (100.0%)	0	-	16	11 (68.8%)	18	13 (72.2%)
Prone subscale (total)	**33**	**23 (69.7%)**	**69**	**46 (66.7%)**	**16**	**12 (75.0%)**	**214**	**135 (63.1%)**	**332**	**216 (65.1%)**
Supine lying (3)	10	8 (80.0%)	15	14 (93.3%)	2	2 (100.0%)	9	9 (100.0%)	36	33 (91.7%)
Supine lying (4)	6	5 (83.3%)	15	10 (66.7%)	4	4 (100.0%)	16	14 (87.5%)	41	33 (80.5%)
Hands to knees	2	2 (100.0%)	7	5 (71.4%)	2	2 (100.0%)	16	14 (87.5%)	27	23 (85.2%)
Hands to feet	4	4 (100.0%)	7	6 (85.7%)	1	1 (100.0%)	21	21 (100.0%)	33	32 (97.0%)
Rolling supine to prone with rotation	2	2 (100.0%)	2	2 (100.0%)	1	1 (100.0%)	21	16 (76.2%)	26	21 (80.8%)
Supine subscale (total)	**24**	**21 (87.5%)**	**46**	**37 (80.4%)**	**10**	**10 (100.0%)**	**83**	**74 (89.2%)**	**163**	**142 (87.1%)**
Sitting with support	9	4 (44.4%)	23	6 (26.1%)	6	5 (83.3%)	14	8 (57.1%)	52	23 (44.2%)
Sitting with propped arms	2	2 (100.0%)	14	10 (71.4%)	1	1 (100.0%)-	24	21 (87.5%)	41	34 (82.9%)
Pull to sit	1	0 (0.0%)	9	8 (88.9%)	0	-	20	16 (80.0%)	30	24 (80.0%)
Sitting with arm support	1	1 (100.0%)	9	8 (88.9%)	0	-	28	21 (75.0%)	38	30 (78.9%)
Sitting without arm support (1)	1	0 (0.0%)	8	5 (62.5%)	1	1 (100.0%)	30	19 (63.3%)	40	25 (62.5%)
Reach with rotation in sitting	0	-	1	0 (0.0%)	0	-	31	14 (45.2%)	32	14 (43.8%)
Sitting to four-point kneeling	1	1 (100.0%)	2	1 (50.0%)	0	-	23	11 (47.8%)	26	13 (50.0%)
Sitting without arm support (2)	1	1 (100.0%)	1	0 (0.0%)	1	1 (100.0%)	20	12 (60.0%)	23	14 (60.9%)
Sitting subscale (total)	**16**	**9 (56.3%)**	**67**	**38 (56.7%)**	**9**	**8 (88.9%)**	**190**	**122 (64.2%)**	**282**	**177 (62.8%)**

Major changes for each version: wording revision for recording time duration in version 2.1; revision for evaluation logic and wording of some movements in version 2.2; issue of incorrect infant orientation due to automatic video rotation, which caused the infant to be misaligned, was fixed in version 2.3.
Data presented as number (%). The data was updated until 2025.06.30.
(To be continued)

doi:10.6342/NTU202503829

Table 15. Agreement between AI and physiotherapists' labeling results for each movement across different versions of APP (Con.)

Movement item	Ver 2.0		Ver 2.1		Ver 2.2		Ver 2.3 (Current)		Overall	
	Number of trials	**Agreement**	**Number of trials**	**Agreement**	**Number of trials**	**Agreement**	**Number of trials**	**Agreement**	**Number of trials**	**Agreement**
Supported standing (2)	1	1 (100.0%)	9	7 (77.8%)	2	1 (50.0%)	18	15 (83.3%)	30	24 (80.0%)
Supported standing (3)	1	1 (100.0%)	0	-	1	1 (100.0%)	11	1 (9.1%)	13	3 (23.1%)
Pulls to stand with support	0	-	3	0 (0.0%)	0	-	22	11 (50.0%)	25	11 (44.0%)
Pulls to stand	1	1 (100.0%)	0	-	0	-	26	11 (42.3%)	27	12 (44.4%)
Supported standing with rotation	1	1 (100.0%)	1	0 (0.0%)	0	-	14	6 (42.9%)	16	7 (43.8%)
Cruising without rotation	1	1 (100.0%)	0	-	0	-	11	7 (63.6%)	12	8 (66.7%)
Controlled lowering through standing	0	-	2	1 (50.0%)	0	-	9	6 (66.7%)	11	7 (63.6%)
Cruising with rotation	0	-	0	-	1	1 (100.0%)	3	1 (33.3%)	4	2 (50.0%)
Stands alone	0	-	2	1 (50.0%)	0	-	5	3 (60.0%)	7	4 (57.1%)
Early stepping	0	-	0	-	0	-	4	3 (75.0%)	4	3 (75.0%)
Standing from modified squat	0	-	0	-	0	-	3	1 (33.3%)	3	1 (33.3%)
Standing from quadruped position	0	-	0	-	0	-	3	3 (100.0%)	3	3 (100.0%)
Walks alone	0	-	0	-	0	-	2	2 (100.0%)	2	2 (100.0%)
Squat	0	-	0	-	0	-	2	2 (100.0%)	2	2 (100.0%)
Standing subscale (total)	**5**	**5 (100.0%)**	**17**	**9 (52.9%)**	**4**	**3 (75.0%)**	**133**	**72 (54.1%)**	**159**	**89 (56.0%)**
Overall	**78**	**58 (74.4%)**	**199**	**130 (65.3%)**	**39**	**33 (84.6%)**	**620**	**403 (65.0%)**	**936**	**624 (66.7%)**

Major changes for each version: wording revision for recording time duration in version 2.1; revision for evaluation logic and wording of some movements in version 2.2; issue of incorrect infant orientation due to automatic video rotation, which caused the infant to be misaligned, was fixed in version 2.3.
Data presented as number (%). The data was updated until 2025.06.30.

doi:10.6342/NTU202503829

Table 16. Agreement between AI and physiotherapists' labeling results for each movement in videos with portrait mode or incorrect infant orientation

Movement item	Portrait mode		Incorrect infant orientation		Portrait or incorrect orientation		Correct recording format		Overall	
	Number of trials	Agreement	Number of trials	Agreement	Number of trials	Agreement	Number of trials	Agreement	Number of trials	Agreement
Forearm support (1)	9	5 (55.6%)	20	13 (65.0%)	28	18 (64.3%)	35	25 (71.4%)	63	43 (68.3%)
Forearm support (2)	3	2 (66.7%)	5	4 (80.0%)	8	6 (75.0%)	32	25 (78.1%)	40	31 (77.5%)
Extended arm support	7	7 (100.0%)	7	2 (28.6%)	14	9 (64.3%)	37	31 (83.8%)	51	40 (78.4%)
Reaching from forearm support	8	3 (37.5%)	1	1 (100.0%)	9	4 (44.4%)	20	10 (50.0%)	29	14 (48.3%)
Pivoting	6	5 (83.3%)	2	2 (100.0%)	8	7 (87.5%)	11	10 (90.9%)	19	17 (89.5%)
Four-point kneeling (1)	10	1 (10.0%)	3	1 (33.3%)	13	2 (15.4%)	13	12 (92.3%)	26	14 (53.8%)
Reciprocal crawling	5	3 (60.0%)	5	0 (0.0%)	10	3 (30.0%)	11	10 (90.9%)	21	13 (61.9%)
Four-point kneeling to sitting or half-sitting	11	2 (18.2%)	1	1 (100.0%)	12	3 (25.0%)	12	11 (91.7%)	24	14 (58.3%)
Reciprocal creeping (1)	4	3 (75.0%)	2	1 (50.0%)	6	4 (66.7%)	11	9 (81.8%)	17	13 (76.5%)
Reaching from extended arm support	6	0 (0.0%)	4	0 (0.0%)	10	0 (0.0%)	14	4 (28.6%)	24	4 (16.7%)
Reciprocal creeping (2)	4	2 (50.0%)	2	0 (0.0%)	5	2 (40.0%)	13	11 (84.6%)	18	13 (72.2%)
Prone subscale (total)	**73**	**33 (45.2%)**	**52**	**25 (48.1%)**	**123**	**58 (47.2%)**	**209**	**158 (75.6%)**	**332**	**216 (65.1%)**
Supine lying (3)	12	12 (100.0%)	8	6 (75.0%)	20	18 (90.0%)	16	15 (93.8%)	36	33 (91.7%)
Supine lying (4)	11	8 (72.7%)	8	5 (62.5%)	19	13 (68.4%)	22	20 (90.9%)	41	33 (80.5%)
Hands to knees	9	6 (66.7%)	2	2 (100.0%)	11	8 (72.7%)	16	15 (93.8%)	27	23 (85.2%)
Hands to feet	9	9 (100.0%)	2	2 (100.0%)	11	11 (100.0%)	22	21 (95.5%)	33	32 (97.0%)
Rolling supine to prone with rotation	9	4 (44.4%)	3	3 (100.0%)	12	7 (58.3%)	14	14 (100.0%)	26	21 (80.8%)
Supine subscale (total)	**50**	**39 (78.0%)**	**23**	**18 (78.3%)**	**73**	**57 (78.1%)**	**90**	**85 (94.4%)**	**163**	**142 (87.1%)**
Sitting with support	15	7 (46.7%)	15	3 (20.0%)	30	10 (33.3%)	22	13 (59.1%)	52	23 (44.2%)
Sitting with propped arms	13	11 (84.6%)	9	4 (44.4%)	22	15 (68.2%)	19	19 (100.0%)	41	34 (82.9%)
Pull to sit	5	4 (80.0%)	4	3 (75.0%)	9	7 (77.8%)	21	17 (81.0%)	30	24 (80.0%)
Sitting with arm support	16	10 (62.5%)	1	0 (0.0%)	17	10 (58.8%)	21	20 (95.2%)	38	30 (78.9%)
Sitting without arm support (1)	21	12 (57.1%)	8	3 (37.5%)	29	15 (51.7%)	11	10 (90.9%)	40	25 (62.5%)
Reach with rotation in sitting	12	5 (41.7%)	3	0 (0.0%)	15	5 (33.3%)	17	9 (52.9%)	32	14 (43.8%)
Sitting to four-point kneeling	10	3 (30.0%)	3	1 (33.3%)	13	4 (30.8%)	13	9 (69.2%)	26	13 (50.0%)
Sitting without arm support (2)	11	5 (45.5%)	3	2 (66.7%)	13	6 (46.2%)	10	8 (80.0%)	23	14 (60.9%)
Sitting subscale (total)	**103**	**57 (55.3%)**	**46**	**16 (34.8%)**	**148**	**72 (48.6%)**	**134**	**105 (78.4%)**	**282**	**177 (62.8%)**

There were 7 videos recorded in portrait mode with incorrect infant orientation.

Correct recording format videos: Videos that were recorded in a horizontal (landscape) orientation with the infant's head and feet properly aligned.

Data presented as number (%). The data was updated until 2025.06.30.

(To be continued)

doi:10.6342/NTU202503829

Table 16. Agreement between AI and physiotherapists' labeling results for each movement in videos with portrait mode or incorrect infant orientation (Con.)

Movement item	Portrait mode		Incorrect infant orientation		Portrait or incorrect orientation		Correct recording format		Overall	
	Number of trials	Agreement	Number of trials	Agreement	Number of trials	Agreement	Number of trials	Agreement	Number of trials	Agreement
Supported standing (2)	13	9 (69.2%)	3	3 (100.0%)	16	12 (75.0%)	14	12 (85.7%)	30	24 (80.0%)
Supported standing (3)	8	3 (37.5%)	0	-	8	3 (37.5%)	5	0 (0.0%)	13	3 (23.1%)
Pulls to stand with support	7	3 (42.9%)	5	1 (20.0%)	12	4 (33.3%)	13	7 (53.8%)	25	11 (44.0%)
Pulls to stand	15	3 (20.0%)	1	0 (0.0%)	16	3 (18.8%)	11	9 (81.8%)	27	12 (44.4%)
Supported standing with rotation	9	5 (55.6%)	4	0 (0.0%)	12	5 (41.7%)	4	2 (50.0%)	16	7 (43.8%)
Cruising without rotation	6	2 (33.3%)	1	0 (0.0%)	6	2 (33.3%)	6	6 (100.0%)	12	8 (66.7%)
Controlled lowering through standing	4	2 (50.0%)	4	1 (25.0%)	7	3 (42.9%)	4	4 (100.0%)	11	7 (63.6%)
Cruising with rotation	3	1 (33.3%)	1	0 (0.0%)	3	1 (33.3%)	1	1 (100.0%)	4	2 (50.0%)
Stands alone	4	2 (50.0%)	0	-	4	2 (50.0%)	3	2 (66.7%)	7	4 (57.1%)
Early stepping	3	2 (66.7%)	0	-	3	2 (66.7%)	1	1 (100.0%)	4	3 (75.0%)
Standing from modified squat	2	0 (0.0%)	0	-	2	0 (0.0%)	1	1 (100.0%)	3	1 (33.3%)
Standing from quadruped position	1	1 (100.0%)	1	1 (100.0%)	2	2 (100.0%)	1	1 (100.0%)	3	3 (100.0%)
Walks alone	1	1 (100.0%)	1	1 (100.0%)	2	2 (100.0%)	0	-	2	2 (100.0%)
Squat	1	1 (100.0%)	0	-	1	1 (100.0%)	1	1 (100.0%)	2	2 (100.0%)
Standing subscale (total)	**77**	**35 (45.5%)**	**21**	**7 (33.3%)**	**94**	**42 (44.7%)**	**65**	**47 (72.3%)**	**159**	**89 (56.0%)**
Overall	**303**	**164 (54.1%)**	**142**	**66 (46.5%)**	**438**	**229 (52.3%)**	**498**	**395 (79.3%)**	**936**	**624 (66.7%)**

There were 7 videos recorded in portrait mode with incorrect infant orientation.

Correct recording format videos: Videos that were recorded in a horizontal (landscape) orientation with the infant's head and feet properly aligned.

Data presented as number (%). The data was updated until 2025.06.30.

doi:10.6342/NTU202503829

Table 17. Agreement between AI and parents' results for each movement across different versions of APP

Movement item	Ver 2.0		Ver 2.1		Ver 2.2		Ver 2.3 (Current)		Overall	
	Number of trials	Agreement	Number of trials	Agreement	Number of trials	Agreement	Number of trials	Agreement	Number of trials	Agreement
Forearm support (1)	16	8 (50.0%)	23	9 (39.1%)	6	3 (50.0%)	18	7 (38.9%)	63	27 (42.9%)
Forearm support (2)	1	1 (100.0%)	14	13 (92.9%)	3	3 (100.0%)	22	15 (68.2%)	40	32 (80.0%)
Extended arm support	8	2 (25.0%)	14	5 (35.7%)	2	2 (100.0%)	27	16 (59.3%)	51	25 (49.0%)
Reaching from forearm support	1	1 (100.0%)	4	3 (75.0%)	3	1 (33.3%)	21	8 (38.1%)	29	13 (44.8%)
Pivoting	1	1 (100.0%)	2	1 (50.0%)	1	1 (100.0%)	15	11 (73.3%)	19	14 (73.7%)
Four-point kneeling (1)	1	0 (0.0%)	2	2 (100.0%)	0	-	23	12 (52.2%)	26	14 (53.8%)
Reciprocal crawling	1	1 (100.0%)	5	3 (60.0%)	1	0 (0.0%)	14	8 (57.1%)	21	12 (57.1%)
Four-point kneeling to sitting or half-sitting	1	1 (100.0%)	1	1 (100.0%)	0	-	22	12 (54.5%)	24	14 (58.3%)
Reciprocal creeping (1)	1	1 (100.0%)	1	0 (0.0%)	0	-	15	13 (86.7%)	17	14 (82.4%)
Reaching from extended arm support	1	0 (0.0%)	2	0 (0.0%)	0	-	21	2 (9.5%)	24	2 (8.3%)
Reciprocal creeping (2)	1	1 (100.0%)	1	1 (100.0%)	0	-	16	11 (68.8%)	18	13 (72.2%)
Prone subscale (total)	**33**	**17 (51.5%)**	**69**	**38 (55.1%)**	**16**	**10 (62.5%)**	**214**	**115 (53.7%)**	**332**	**180 (54.2%)**
Supine lying (3)	10	8 (80.0%)	15	14 (93.3%)	2	2(100.0%)	9	6 (66.7%)	36	30 (83.3%)
Supine lying (4)	6	5 (83.3%)	15	10 (66.7%)	4	4(100.0%)	16	13 (81.3%)	41	32 (78.0%)
Hands to knees	2	2 (100.0%)	7	7 (100.0%)	2	2(100.0%)	16	14 (87.5%)	27	25 (92.6%)
Hands to feet	4	4 (100.0%)	7	6 (85.7%)	1	1(100.0%)	21	21 (100.0%)	33	32 (97.0%)
Rolling supine to prone with rotation	2	2 (100.0%)	2	2 (100.0%)	1	1(100.0%)	21	17 (81.0%)	26	22 (84.6%)
Supine subscale (total)	**24**	**21 (87.5%)**	**46**	**39 (84.8%)**	**10**	**10(100.0%)**	**83**	**71 (85.5%)**	**163**	**141 (86.5%)**
Sitting with support	9	5 (55.6%)	23	6 (26.1%)	6	5(83.3%)	14	8 (57.1%)	52	24 (46.2%)
Sitting with propped arms	2	1 (50.0%)	14	7 (50.0%)	1	0 (0.0%)	24	19 (79.2%)	41	27 (65.9%)
Pull to sit	1	0 (0.0%)	9	8 (88.9%)	0	-	20	16 (80.0%)	30	24 (80.0%)
Sitting with arm support	1	1 (100.0%)	9	8 (88.9%)	0	-	28	21 (75.0%)	38	30 (78.9%)
Sitting without arm support (1)	1	1 (100.0%)	8	4 (50.0%)	1	1(100.0%)	30	18 (60.0%)	40	24 (60.0%)
Reach with rotation in sitting	0	-	1	0 (0.0%)	0	-	31	9 (29.0%)	32	9 (28.1%)
Sitting to four-point kneeling	1	1(100.0%)	2	1 (50.0%)	0	-	23	14 (60.9%)	26	16 (61.5%)
Sitting without arm support (2)	1	1(100.0%)	1	0 (0.0%)	1	0 (0.0%)	20	12 (60.0%)	23	13 (56.5%)
Sitting subscale (total)	**16**	**10 (62.5%)**	**67**	**34 (50.7%)**	**9**	**6(66.7%)**	**190**	**117 (61.6%)**	**282**	**167 (59.2%)**

Major changes for each version: wording revision for recording time duration in version 2.1; revision for evaluation logic and wording of some movements in version 2.2; issue of incorrect infant orientation due to automatic video rotation, which caused the infant to be misaligned, was fixed in version 2.3.
Data presented as number (%). The data was updated until 2025.06.30.
(To be continued)

doi:10.6342/NTU202503829

Table 17. Agreement between AI and parents' results for each movement across different versions of APP (Con.)

Movement item	Ver 2.0		Ver 2.1		Ver 2.2		Ver 2.3 (Current)		Overall	
	Number of trials	Agreement	Number of trials	Agreement	Number of trials	Agreement	Number of trials	Agreement	Number of trials	Agreement
Supported standing (2)	1	1(100.0%)	9	7 (77.8%)	2	2 (100.0%)	18	15 (83.3%)	30	25 (83.3%)
Supported standing (3)	1	1(100.0%)	0	-	1	0 (0.0%)	11	2 (18.2%)	13	3 (23.1%)
Pulls to stand with support	0	-	3	0 (0.0%)	0	-	22	11 (50.0%)	25	11 (44.0%)
Pulls to stand	1	1(100.0%)	0	-	0	-	26	11 (42.3%)	27	12 (44.4%)
Supported standing with rotation	1	1(100.0%)	1	0(0.0%)	0	-	14	5 (35.7%)	16	6 (37.5%)
Cruising without rotation	1	1(100.0%)	0	-	0	-	11	8 (72.7%)	12	9 (75.0%)
Controlled lowering through standing	0	-	2	1 (50.0%)	0	-	9	6 (66.7%)	11	7 (63.6%)
Cruising with rotation	0	-	0	-	1	1(100.0%)	3	0 (0.0%)	4	1 (25.0%)
Stands alone	0	-	2	1 (50.0%)	0	-	5	3 (60.0%)	7	4 (57.1%)
Early stepping	0	-	0	-	0	-	4	3 (75.0%)	4	3 (75.0%)
Standing from modified squat	0	-	0	-	0	-	3	1 (33.3%)	3	1 (33.3%)
Standing from quadruped position	0	-	0	-	0	-	3	3 (100.0%)	3	3 (100.0%)
Walks alone	0	-	0	-	0	-	2	2 (100.0%)	2	2 (100.0%)
Squat	0	-	0	-	0	-	2	2 (100.0%)	2	2 (100.0%)
Standing subscale (total)	**5**	**5(100.0%)**	**17**	**9 (52.9%)**	**4**	**3(75.0%)**	**133**	**72 (54.1%)**	**159**	**89 (56.0%)**
Overall	**78**	**53 (67.9%)**	**199**	**120 (60.3%)**	**39**	**29 (74.4%)**	**620**	**375 (60.5%)**	**936**	**577 (61.6%)**

Major changes for each version: wording revision for recording time duration in version 2.1; revision for evaluation logic and wording of some movements in version 2.2; issue of incorrect infant orientation due to automatic video rotation, which caused the infant to be misaligned, was fixed in version 2.3.
Data presented as number (%). The data was updated until 2025.06.30.

doi:10.6342/NTU202503829

Table 18. Illustration of parental reports of the APP usability (N=30)

Statement	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
A. Frequency of the APP use					
Since participating in the project, I have frequently used this App to assess my baby's development.	Frequently used：30 (100.0%) / Never used：0 (0.0%)				
The advantages of using this App are: (Multiple choices)	Interesting：17 (56.7%) / Convenient：23 (76.7%) / Professional：22 (73.3%)				
B. Benefits of the APP use					
The information from the parental checklist is helpful.	0 (0.0%)	0 (0.0%)	3 (10.0%)	17 (56.7%)	10 (33.3%)
The assessment results of the baby's movement videos are helpful.	0 (0.0%)	0 (0.0%)	2 (6.7%)	14 (46.7%)	14 (46.7%)
The information from home-based education is helpful.	0 (0.0%)	0 (0.0%)	5 (16.7%)	15 (50.0%)	10 (33.3%)
The overall design helps me better understand my baby's development.	0 (0.0%)	0 (0.0%)	0 (0.0%)	14 (46.7%)	16 (53.3%)
The information provided by the App makes me more worried about my baby's development.	5 (16.7%)	19 (63.3%)	5 (16.7%)	1 (0.3%)	0 (0.0%)
The way of transmitting video data is secure and reliable.	0 (0.0%)	0 (0.0%)	3 (10.0%)	17 (56.7%)	10 (33.3%)
Which method do I prefer for assessing my baby's motor development?	Assessment conducted only in person by healthcare professionals：3 (10.0%) Assessment conducted only remotely via this app：1 (3.3%) Assessment conducted both in person and remotely：26 (86.7%)				
The App is easy to use.	0 (0.0%)	0 (0.0%)	2 (6.7%)	14 (46.7%)	14 (46.7%)
How much time do I spend recording videos of my baby's movements at each age?	0-30 minutes：26 (86.7%) / 30-60 minutes：3 (10.0%) / 60-90 minutes：1 (3.3%) / >90 minutes：0 (0%)				
Is someone assisting me when I record videos of my baby's movements?	None：7 (23.3%)/ Father or mother：19 (63.3%)/ Other family members：9 (30.0%)				
The instructions for recording videos are easy to understand.	0 (0.0%)	0 (0.0%)	2 (6.7%)	14 (46.7%)	14 (46.7%)
The standards for recording videos are easy to meet.	0 (0.0%)	2 (6.7%)	6 (20.0%)	13 (43.3%)	9 (30.0%)
The postures and motor skills in the videos are easy to capture.	0 (0.0%)	5 (16.7%)	9 (30.0%)	11 (36.7%)	5 (16.7%)
What factors commonly affect your baby's performance during video recording? (Multiple choices)	None：9 (30.0%)/ Environmental noise：3 (10.0%)/ Baby was in poor physical condition：2 (6.7%) Baby was in a bad mood：17 (56.7%)				
Have you encountered any technical problems while using the App?	(descriptive results)				
The reminders about when to upload videos are helpful.	0 (0.0%)	1 (3.3%)	5 (16.7%)	12 (40.0%)	12 (40.0%)
The frequency of reminders to upload videos is appropriate.	0 (0.0%)	0 (0.0%)	3 (10.0%)	13 (50.0%)	12 (40.0%)

Data presented as number (%).
A total of 28 parents completed the survey for the first time, and 2 parents completed it for the second time.
The data was updated until 2025.06.30.

doi:10.6342/NTU202503829

Table 19. Illustration of clinicians' reports of the APP usability (N=2)

Statement	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
What is your occupation?	Physiotherapy: 2 (100.0%)				
What is your work experience?	Above 5 years: 2 (100.0%)				
What is your main service department?	Physiotherapy center: 1 (50.0%) /Department of Physical Medicine and Rehabilitation: 1 (50.0%)				
A. Perspective of using APP					
It is helpful for performing infant motor assessment.	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (100.0%)
The APP results have good accuracy for assessing motor delay.	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (50.0%)	1 (50.0%)
This APP can decrease the working hours for infant motor assessment in the clinic.	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (100.0%)	0 (0.0%)
This APP increases the parent's awareness of their infant's motor development.	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (100.0%)
Are you willing to recommend this APP to other parents?	Yes: 2 (100.0%)				
B. Other recommendations					
What aspects of this APP do you think need improvement?	No: 2 (100.0%)				
What other features do you think this APP can add?	Keep the video record on the APP: 2 (100.0%) Immediate feedback system: 2 (100.0%) Multiple languages available: 1 (50.0%)				

Data presented as number (%).
The data was updated until 2025.06.30.

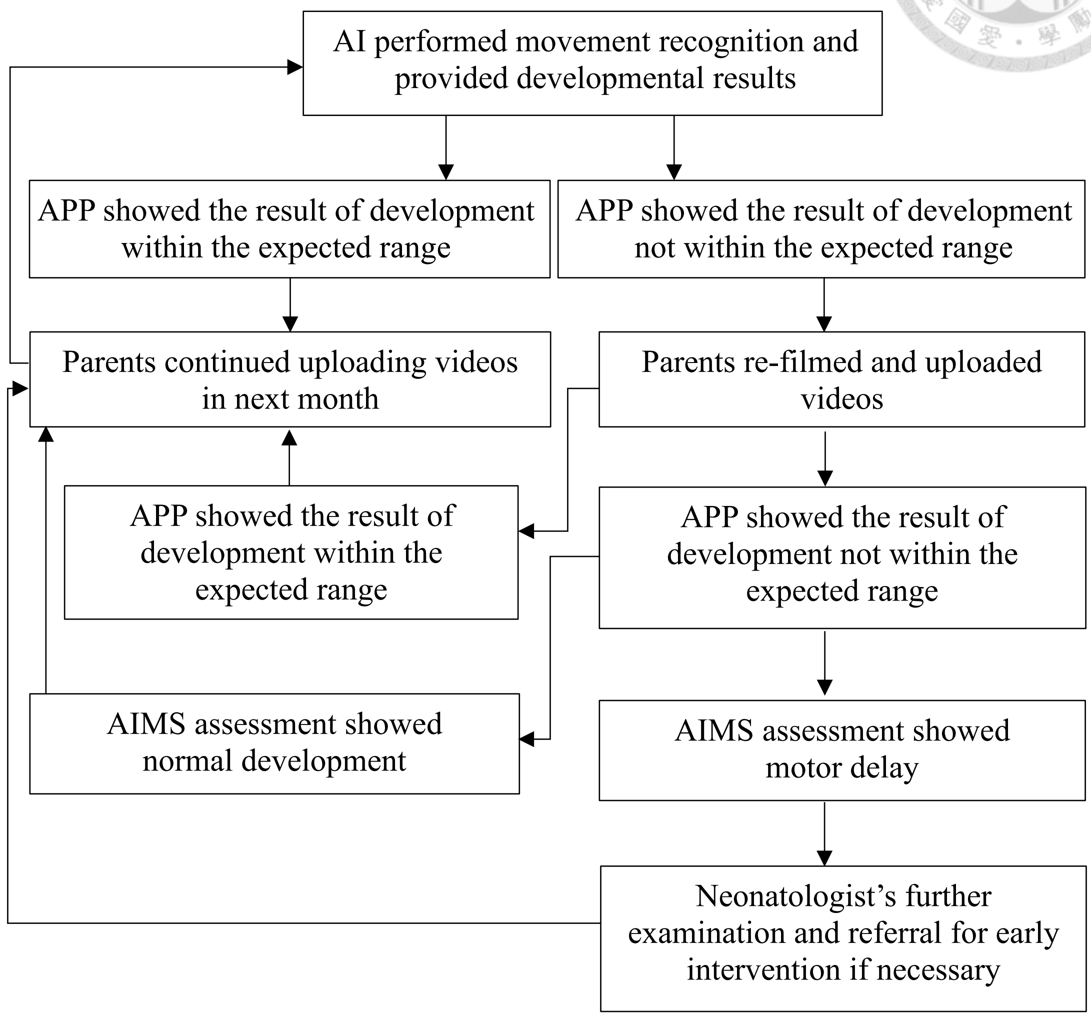
doi:10.6342/NTU202503829

Table 20. Movement with low agreement between AI and physiotherapists' labeling results

Movement	Agreement	Issues observed from PT side	Cause identified by CS team	Evidence supports
Prone				
Forearm support (1)	68.3%	AI identify as other mature prone movement	Ambiguity of fine-grained movement	Confusion matrix showed misclassification with forearm support (2)
Reaching from forearm support	48.3%	Cover by toy	Occlusion	Users pointed out toy-related occlusion
Reaching from extended arm support	16.7%	Unknown	Occlusion Long-tail distribution	Movements involving reaching often resulted in self-occlusion of the infant's body, and the number of video samples may have been relatively low.
Sitting				
Sitting with support	44.2%	Vertical video AI identify as other sitting movement	Incorrect Video Orientation Ambiguity of fine-grained movement	Users pointed out issues related to orientation and camera angles
Reach with rotation in sitting	43.8%	Unknown	Ambiguity of fine-grained movement Long-tail distribution	Movement involves subtle motion and may have a relatively small number of samples
Standing				
Supported standing (3)	23.1%	AI identify as Supported Standing (2)	Ambiguity of fine-grained movement Long-tail distribution	Differences between this movement and other supported standing actions were subtle, and the sample size was extremely small (only 5 samples)
Cruising with rotation	50.0%	AI identify as Cruising without Rotation	Long-tail distribution Ambiguity of fine-grained movement	The number of samples was extremely limited (e.g., only 2), and the differences from similar actions such as lateral cruising without rotation were subtle.
Stands alone	57.1%	AI identify as other standing movements	Ambiguity of fine-grained movement Long-tail distribution	Minimal distinction from other standing movements

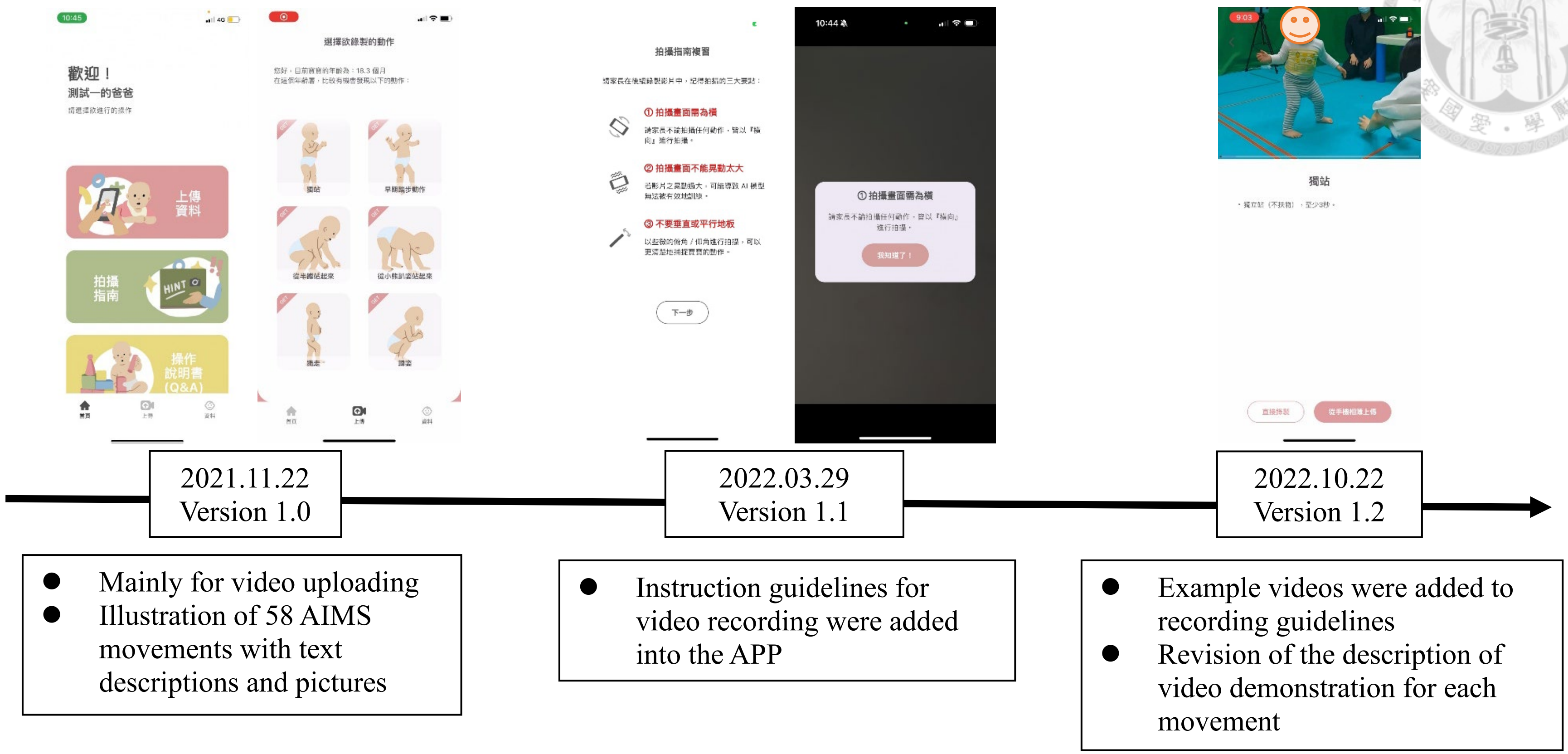
doi:10.6342/NTU202503829

Figure 1. Flowchart for the APP developmental assessment, AIMS assessment, and follow-up



doi:10.6342/NTU202503829

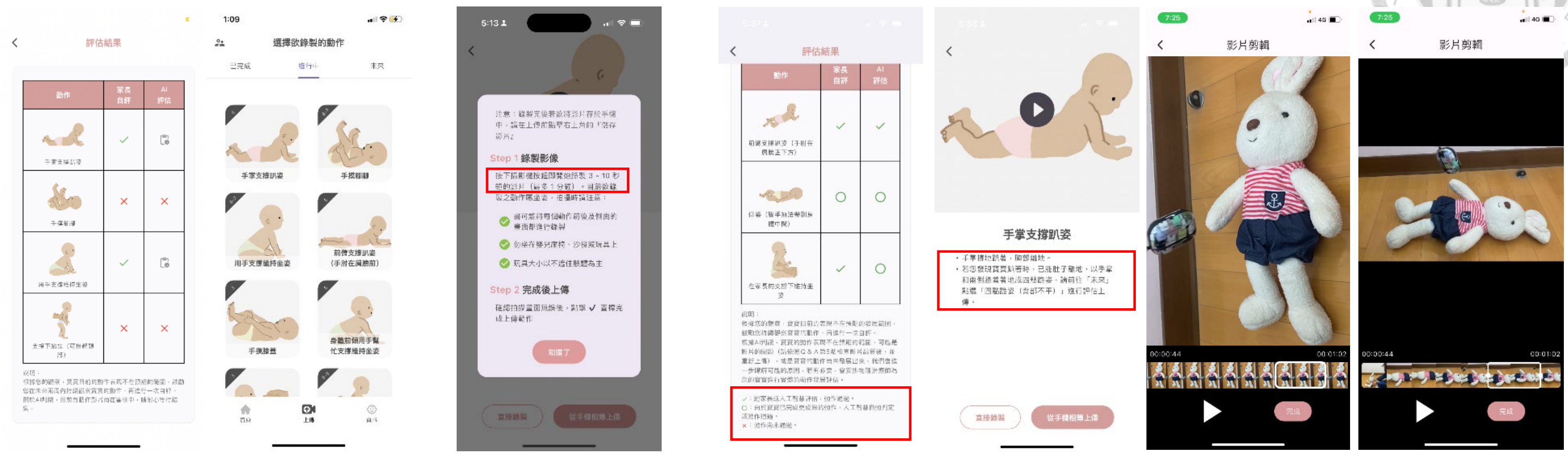
Figure 2. Development of “Baby Go” APP



(To be continued)

doi:10.6342/NTU202503829

Figure 2. Development of “Baby Go” APP (Con.)



2024.10.04 Version 2.0	2024.12.05 Version 2.1	2025.02.18 Version 2.2	2025.03.08 Version 2.3
• Added assessment, follow-up, and education features • Added reminders and video editor • Uploading interference was changed	• Wording revision for recording time duration	• Change in the APP's evaluation logic if a baby passed advanced movement assessment, movements before that would be automatically marked as passed • Upload guide was added to help parents identify confusing items and upload the correct videos	• Issue of incorrect infant orientation due to automatic video rotation, which caused the infant to be misaligned, was fixed

doi:10.6342/NTU202503829

Figure 3. Procedure for remote infant motor assessment using the “Baby Go” APP

(To be continued)

doi:10.6342/NTU202503829

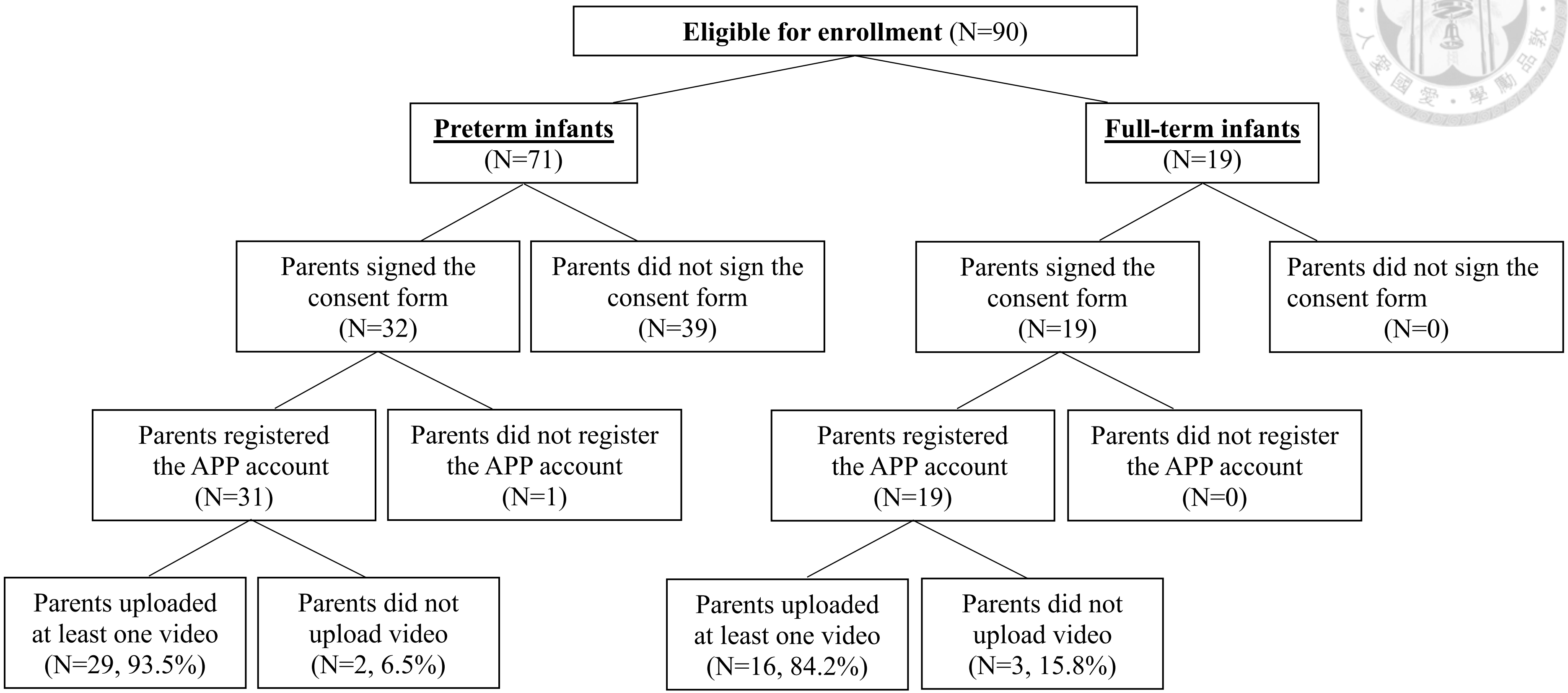
Figure 3. Procedure for remote infant motor assessment using the “Baby Go” APP (Con.)

doi:10.6342/NTU202503829

Figure 4. Procedure for collecting raw data of parental perception and AI classification in the “Baby Go” APP background

doi:10.6342/NTU202503829

Figure 5. Flowchart for subject recruitment and data collection



doi:10.6342/NTU202503829

Appendices

Appendix 1. Dispatch of the 38 movements into age-based sets for infant motor assessment during 3 to 18 months of age

Age (Month)	Prone	Supine	Sitting	Standing
3	Forearm support (1)	Supine lying (3)	Sitting with support	
4	Forearm support (1)	Supine lying (4)	Sitting with support	
5	Forearm support (2)	Hands to knees	Sitting with propped arms Pull to sit	Supported standing (2)
6	Extended arm support	Hands to feet	Sitting with arm support	Supported standing (2)
7	Reaching from arm support	Hands to feet	Sitting with arm support	Supported standing (2)
8	Pivoting	Rolling to prone with rotation	Sitting without arm support (1) Reach with rotation in sitting	Supported standing (2)
9	Four points kneeling (1)	Rolling to prone with rotation	Sitting to four points kneeling	Supported standing (3) Pull to stand with support
10	Reciprocal crawling Four points kneeling to sitting		Sitting without arm support (2)	Pull to stand/stance Supported standing with rotation
11	Reciprocal creeping (1) Reaching from extended arm support		Sitting without arm support (2)	Cruising without rotation
12	Reciprocal creeping (2)		Sitting without arm support (2)	Cruising without rotation Controlled lowering through standing
13	Reciprocal creeping (2)			Cruising with rotation Stands alone
14				Stands alone Early stepping
15				Standing from quadruped position Standing from modified squat Walks alone
16-18				Walks alone Squatting

doi:10.6342/NTU202503829

Appendix 2. Labeling instructions for 38 AIMS movements

AIMS movements	Type	Demonstration
PRONE		
Forearm support (1)	static	At least 3 seconds
Forearm support (2)	static	At least 3 seconds
Extended arm support	static	At least 3 seconds
Reaching from forearm support	dynamic	(1) Perform one time in each record (2) Baby needs to reach upward and forward; toys are not necessary.
Pivoting	dynamic	If the infant change direction (e.g. rotation to right change to left) please stop the annotation and start a new one.
Four-point kneeling (1)	static	At least 3 seconds
Reciprocal crawling	dynamic	(1) If the infant change direction (e.g. crawing to right change to left) please stop the annotation and start a new one. (2) At least two steps or two arm strolls. (needs body foreword or backword displacement)
Four-point kneeling to sitting or half-sitting	dynamic	(1) Perform one time in each record (2) Buttock needs to leave the ground
Reciprocal creeping (1)	dynamic	(1) If the infant change direction (e.g. creeping to right change to left) please stop the annotation and start a new one. (2) At least two steps and two arm strolls (needs body forward displacement).
Reaching from extended arm support	dynamic	(1) Perform one time in each record (2) Baby needs reach upward and forward; toys are not necessary.
Reciprocal creeping (2)	dynamic	(1) If the infant change direction (e.g. creeping to right change to left) please stop the annotation and start a new one (2) At least two steps and two arm strolls. (needs body foreword displacement)

doi:10.6342/NTU202503829

Appendix 2. Labeling instructions for 38 AIMS movements (Con.)

AIMS movements	Type	Demonstration
SUPINE		
Supine lying (3)	static	At least 3 seconds
Supine lying (4)	static	At least 3 seconds
Hands to knees	static	(1) At least 3 seconds (2) The baby can bend the knees to 90 degrees for 3 seconds. (3) Hands are not necessarily touching the legs.
Hands to feet	static	(1) At least 3 seconds (2) The baby can band the knees beyond 90 degrees for 3 seconds. (3) At least on hand touching the foot.
Rolling supine to prone with rotation	dynamic	Perform one time in each record
SITTING		
Sitting with support	static	At least 3 seconds
Sitting with propped arms	static	At least 3 seconds
Pull to sit	dynamic	Perform one time in each record
Sitting with arm support	static	At least 3 seconds
Sitting without arm support (1)	static	At least 20 seconds
Reach with rotation in sitting	dynamic	Perform one time in each record
Sitting to four-point kneeling	dynamic	(1) Perform one time in each record (2) Needs completely transfer to prone position
Sitting without arm support (2)	static	(1) At least 3 seconds (2) Needs to record various leg positions and weight shifting (3) If the infant sits carefully and DOES NOT do other thing, please record the items to <u>sitting without arm support (1)</u>.

doi:10.6342/NTU202503829

Appendix 2. Labeling instructions for 38 AIMS movements (Con.)

AIMS movements	Type	Demonstration
STANDING		
Supported standing (2)	static	At least 3 seconds
Supported standing (3)	static	At least 3 seconds
Pulls to stand with support	dynamic	Perform one time in each record
Pulls to stand	dynamic	Perform one time in each record
Supported standing with rotation	dynamic	Perform one time in each record
Cruising without rotation	dynamic	(1) If the infant change direction (e.g. cruising to right change to left) please stop the annotation and start a new one (2) At least 2 steps in one direction.
Controlled lowering through standing	dynamic	Perform one time in each record
Cruising with rotation	dynamic	(1) If the infant change direction (e.g. cruising to right change to left) please stop the annotation and start a new one (2) At least 2 steps
Stands alone	static	At least 3 seconds
Early stepping	dynamic	(1) If the infant change direction (e.g. walking to right change to left) please stop the annotation and start a new one (2) At least 5 steps, arms high guarding
Standing from modified squat	dynamic	Perform one time in each record
Standing from quadruped position	dynamic	Perform one time in each record
Walks alone	dynamic	(1) If the infant change direction (e.g. walking to right change to left) please stop the annotation and start a new one (2) At least 5 steps, arms down
Squat	static	At least 3 seconds

doi:10.6342/NTU202503829