

**Development of a new screening tool for neuromotor development in children aged two - the Neuromotor 5 minute Exam 2-year-old version (N5E2)**

Sayaka AOKI <sup>1</sup>, Keiji HASHIMOTO <sup>2</sup>; Hidetoshi MEZAWA <sup>2</sup>; Yuhei HATAKENAKA <sup>3,4</sup>; Kahoko YASUMITSU-LOVELL <sup>3,4</sup>; Narufumi SUGANUMA <sup>4</sup>; Yukihiro OHYA <sup>5</sup>; Philip WILSON <sup>6</sup>; Elisabeth FERNELL<sup>4</sup>; Yoko KAMIO<sup>7</sup>; Christopher GILLBERG <sup>4</sup>

Authors' institutions

<sup>1</sup>Center for Environmental Health Sciences, National Institute for Environmental Studies, Japan

<sup>2</sup>Division of Rehabilitation Medicine and Developmental Evaluation Center, National Center for Child Health and Development, Japan

<sup>3</sup> Kochi University Medical School

<sup>4</sup> Gillberg Neuropsychiatry Centre Sahlgrenska Academy, University of Gothenburg

<sup>5</sup> Division of Allergy, National Center for Child Health and Development, Japan

<sup>6</sup> Centre for Rural Health, University of Aberdon, Scotland

<sup>7</sup> Department of Child and Adolescent Mental Health, National Center of Neurology and Psychiatry, Japan

**Corresponding author:**

Keiji Hashimoto, MD, PhD.

Division of Rehabilitation Medicine and Developmental Evaluation Center, National Center for Child Health and Development

2-10-1 Okura Setagaya-ku Tokyo, 157-8535 Japan

Phone: +81-3-3416-0181

Fax: +81-3-2416-2222-

E-mail: hashimoto-k@ncchd.go.jp

## **Abstract:**

### **Objective**

As a new screening tool for neuromotor development in children aged two, we developed the Neuromotor 5 minute Exam 2-year-old version (N5E2), which can be easily administered by pediatricians or primary care physicians. In this study, as an initial attempt to examine the utility of the N5E2, the inter-rater reliability on scoring for the individual items in this scale was assessed.

### **Methods**

The participants of the study were 29 children (aged 1-5 years, mean age = 2.79) diagnosed with a variety of neuromotor/developmental disorder/high-risk conditions. Inter-rater reliability was examined on the following 11 items in the N5E2: (1) Retrieving a rolling ball, (2) Gait, (3) Toe-walking, (4) Asymmetries of posture and/or movement, (5) Age at unsupported walking, (6) Speaking in two-word understandable sentences, (7) Hypotonus, (8) Hypertonus, (9) Eye movement, (10) Vision problem, (11) Hearing problem. The items were administered to children by two pediatricians with different expertise and clinical experience, separately.

### **Results**

The results showed that among the eleven items in the N5E2 examined, a high level of agreement ( $\kappa \geq .60$ ) was found on 4 items, and a moderate level of agreement ( $.40 \leq \kappa < .60$ ) was found on 5 items. The level of agreement somewhat improved after the dichotomization of the score; using this format, a high level of rater agreement ( $\kappa \geq .60$ ) was found on 6 out of 11 items. The analyses also revealed high inter-rater reliability on the sum score of the 11 items ( $r = .84$ ).

### **Conclusions**

The results suggest the possibility that this brief screening tool could be feasible in settings where clinicians' experience varies, based on its inter-rater reliability on individual items between the clinicians with different expertise and amount of clinical experiences.

### **Key-words:**

**Neuromotor development, Developmental disabilities, Screening, Inter-rater reliability, Japan Environment and Children's Study (JECS), Young children**

## Background

For children with developmental disability, early identification and intervention are important, which is now recognized by many clinical professional organizations. For example, in 2006, the American Academy of Pediatrics (AAP) distributed a revised policy statement that recommends all pediatricians to routinely administer a developmental screening test at 9, 18, 24 and 30 month visits [1], providing a list of various general developmental screening tools with reasonable specificity and sensitivity [2, 3, 4]. Despite this, according to the Periodic Survey conducted in 2009, only 43 % of pediatricians (US AAP members) reportedly used recommended formal tests for developmental screening with patients younger than 36 months, and many still depended on informal checklists that are not standardized [5]. The authors of the report inferred that the many screening instruments might not be practicable for use due to their administration time and cost. Therefore, it is desirable to develop a reliable and valid developmental screening tool that can be easily administered by clinicians in their daily work.

Meanwhile, experts also recently pointed out that development and validation of screening and surveillance tools for neuromotor development has lagged behind those for social and language development [6]. According to Health for All Children, the authors mentioned that there is no reliable, valid, and useful screening tool of neuromotor development in early childhood [7]. However, neuromotor difficulty is not uncommon among young children. For example, according to the Centers for Disease Control and Prevention, estimated world-wide prevalence rate of cerebral palsy was 1.5 to more than 4 per 1000[8]. Developmental coordination disorder (DCD) is also a common disorder causing motor difficulty; 5-6% of children aged 5-11 are affected [9]. Furthermore, researchers found that neuromotor difficulty/abnormality tends to be experienced by children with developmental disabilities other than DCD, such as autism [10, 11], attention-deficit/ hyperactivity disorder [12] [13], and learning disability[14]. Taking account of the paucity of screening tools for neuromotor development and of the prevalence of neuromotor difficulties among young children, specifically, those with developmental disability, it is desirable to develop a new screening tool for motor problems.

In response to this need, our research group developed a screening tool that can be utilized by clinicians to identify neuromotor abnormality among young children. The new screening tool was named Neuromotor 5-minute Exam (N5E) as the screening can be completed within 5 minutes. We first developed a 2-year-old version of the N5E (N5E2), which we were planning to administer in a large-scale birth cohort survey, the Japan Environment and Children's Study (JECS), when the participants would become two year old. [To develop items in the N5E2, we referred to Noritz et al.' s proposal published in 2013 \[6\]](#)

while reflecting clinical and research experiences of the authors who consist of neurologists, pediatricians, psychiatrists, psychologists, and epidemiologists. Among the items coming up in our mind, we selected the items that meet the following criteria: (1) The item indicates neuromotor abnormality of children aged 2, (2) The tool can be administered after receiving minimal training, and (3) Scoring criteria for the item can be so clear that the examinees are scored in the same manner regardless of the examiner's expertise and clinical experiences. As a result, the selected items were (1) Retrieving a rolling ball, (2) Gait, (3) Toe-walking, (4) Asymmetries of posture and/or movement, (5) Age at unsupported walking, (6) Building a block tower, (7) Pointing out body parts, (8) Speaking in two-word understandable sentences, (9) Hypotonus, (10) Hypertonus, (11) Head circumference (12) Weight, (13) Height, (14) Eye movement, (15) Vision problems, and (16) Hearing problems. The selected items not only cover motor problems, tone abnormality, and physical characteristics, but also more aspects of development, such as language, cognition, and perception, because examination of these areas are recommended by American Academy of Pediatrics to identify children with motor delay [6]. Moreover, as abnormality of such areas can be an early sign of neurodevelopmental difficulties [15], inclusion of these items would increase the clinical utility of the N5E2.

As an initial attempt to examine the utility of the N5E2, we assessed the inter-rater reliability for the items in this scale. Since the primary purpose of this examination at this point was to validate the N5E2 before using the tool in the JECS, data were collected only for the items which would actually be administered in the JECS. As a result, we did not collect data for five items (building a block tower, pointing out body parts, head circumference, weight, and height) because for the JECS, these variables would be measured as a part of other tests/exams. Therefore, we investigated inter-rater reliability between pediatricians for the remaining 11 items. The more extended psychometric property of this scale will be examined in other studies with more participants.

## **Methods**

### ***Participants***

The participants of this study were 29 children (18 boys and 11 girls), who visited the orthotic and prosthetic outpatient service and developmental evaluation center in National Center for Child Health and Development (NCCHD) in Tokyo, Japan. The reason for collecting data from this group of children was that the N5E2 is a developmental screening tool and so planned to be administered to at-risk individuals who are expected to fail a few of them due to their symptoms, not only to typically developing children who are more likely to pass all the items. The children were diagnosed with a variety of neuromotor/developmental disorder/ high-risk conditions; eight suffered from low birth weight (< 2,500g), four had documented genetic

abnormalities, three had developmental delay, three spina bifida, two blood disease, one cerebral palsy, one arachnoid cyst, one craniosynostosis, one traumatic brain injury, one tuberous sclerosis, one patent ductus arteries, one tumor, one achondroplasia, and one pes planus. Age of the children ranged from one to five years (mean = 2.79, SD = 1.35). The data were collected from only those whose parents agreed to have his/her child participate in this study, and those who were alert and calm enough to follow the directions of the pediatricians.

### ***Examiners/raters***

Two pediatricians participated in this study as examiners of the N5E2. As N5E2 was developed as a screening tool which can be utilized by pediatricians with diverse types of background experiences, we decided to recruit one generalist and one specialist as raters. The generalist was a pediatrician with a few years' clinical experience with general training in pediatrics, but did not have specific expertise in neuromotor development. The specialist had worked as a clinician in pediatrics and rehabilitation facilities for more than 10 years, with high expertise in neuromotor development. [Neither of the raters was provided any information about the participating children, except for the fact that they visited the prosthetic outpatient service.](#)

### ***Measure***

*N5E2*. The exam was developed to identify neuromotor abnormality of young children. As described previously, the exam is consisted of 16 items. For this study, we examined inter-rater reliability of the 11 items mentioned in the introduction: (1) Retrieving a rolling ball, (2) Gait, (3) Toe-walking, (4) Asymmetries of posture and/or movement, (5) Age at unsupported walking, (6) Speaking in two-word understandable sentences, (7) Hypotonus, (8) Hypertonus, (9) Eye movement, (10) Vision problem, (11) Hearing problem. For each item, the examiner was asked to rate using a 3 point scale. When abnormality is clearly observed, a score of 2 was assigned for that item. When no abnormality was observed, the score was 0. When the pediatrician could not judge abnormality with confidence (possible abnormality), a score of 1 was given. The pediatrician was asked to assign 0 or 2 points as often as possible, and use the score of 1 only when he/she was truly unsure. The detailed administration procedure and scoring criteria for each item are described in the Appendix 1.

### ***Procedure***

Parents of the participants were asked to let their children participate in the study during their visit on the outpatient service. After obtaining an informed consent from the participant's parent, one pediatrician administered N5E2, followed by the other pediatrician in a separate room. All the data were collected by the same pair of pediatricians. The order of administration was randomized. Based on the collected data, inter-rater reliability was calculated for each item, as well as total score. To estimate inter-rater reliability on the

scoring for each item, Cohen's  $\kappa$  was calculated. The first data collection period was between November 2014 and February 2015. During this time, we collected the data from 20 participants and analyzed them, before the JECS started to collect the data using the N5E2 in April 2015. We continued to collect the data until November 2015, adding the data from the other 9 participants, and reanalyzed the data. The results of the analyses are presented in this paper. It should be noted the study was approved by the Institutional Review Board of the National Center for Child Health and Development (NCCHD).

## Results

The distribution of scores for each item is presented in Table 1. Except for age of walking and 2-word sentences, for all the items, more than 65 % of the participants were given scores of 0, indicating skewness of the score distribution.

The results of the analyses are shown in Table 2, the column "original rating." Cohen's  $\kappa$  for each item ranged from -.05 to 1.00, and its average was .55. Based on the interpretation suggested by Altman [16], good agreement ( $\kappa \geq .60$ ) was reached for four of the 11 items (Gait, Age of walking, 2-word sentence, and Hearing), and moderate agreement ( $\kappa \geq .40$ ) was reached for five items (Retrieving a ball, Hypotonus, Hypertonus, Eye movement, and Vision). The agreement on the item Asymmetry was relatively low ( $\kappa = .31$ ), and the agreement on the item Toe-walking was very poor ( $\kappa = -.04$ ).

As N5E2 was developed for screening, it is important for the scale to distinguish those without abnormality from those with some abnormality. Therefore, it would be meaningful to calculate inter-rater reliability, after the score was dichotomized into 0 (no abnormality) versus 1 and 2 (at least some abnormality). The results are presented in the column, "when score 1 and 2 are combined," in Table 2. For "Retrieving a ball," and "Hypotonus," a great increase in agreement was found. A smaller increase was observed for "Hypertonus" and "Vision." Agreement did not change much for "Gait," "Toe-walking," "Asymmetry," "Age of walking," "2-word sentence," "Eye movement," and "Hearing." In consequence, the conversion of the score led to following levels of agreement on each item. Good agreement was found for "Retrieving a ball," "Age of walking," "2-word sentence," "Hypotonus," "Vision," and "Hearing." Moderate level of agreement was found for "Gate," "Hypertonus," and "Eye movement." The level of agreement for the "Asymmetry" remained low. For toe-walking, the agreement was poor even after dichotomization of the score.

In addition to inter-rater reliability on the scoring for each item, we also calculated how much N5E2 total scores are correlated between those derived from ratings of two different raters. The correlation was  $r=.87$ , indicating that the raters generally reached a good agreement on overall neuromotor development of the examinees.

## Discussion

This study aimed to examine inter-rater reliability on the scoring of a new developmental screening test, Neuromotor 5-minute exam. The results showed that on all but one item (toe-walking) at least a moderate level of agreement was found. The level of agreement improved when the scores 1 and 2 were combined; on this format, good inter-rater agreement ( $\kappa \geq .60$ ) was found on 6 out of 11 items. The analyses also revealed that the overall inter-rater reliability for the N5E2 was high.

Among all the items, agreement tends to be high on “2-word sentence,” “age of walking,” and “hearing.” It is reasonable since for these items, scoring relied only on answers from the parents of the examinees; scoring should be the same as long as the parent responded in the same manners to the two raters. Interestingly, a few parents seemed to change their answers in the middle of exam for “age of walking” and “hearing,” judging from the fact that agreement was imperfect for these items.

For the items requiring the raters to observe the tonus of the children (“hypotonus” and “hypertonus,”) inter-rater reliability fell in the moderate range, if original scoring system was used, but improved when score 1 and 2 were combined. The improvement of agreement seems to indicate that clinicians can commonly notice “subtle difference from normal functioning” if they see a patient with some muscle problem, but the level of confidence in judgment varies depending on clinicians.

For the items assessing children’s gross motor characteristics through observation, such as “Retrieving a ball,” “Gait,” “Asymmetry,” inter-rater reliability was likely to fall in to the moderate to high range. The exception was the item “Toe-walking,” for which the inter-rater reliability was found to be extremely low. One possible reason is skewness of the score distribution of this item; Only a few participants were rated as 1 or 2 points. Another possible reason is that toe-walking is not a behavior that can be observed all the time, even for children who reportedly perform toe-walking, according to our clinical experience. All the 4 participants who were rated as 1 or 2 on this item by one of the raters showed toe-walking only once or twice during the administration of the N5E2; they walked more normally at the rest of the time. Meanwhile, all of their parents mentioned that they frequently engaged in toe-walking in daily lives. The result suggests that rating of the item “Toe-walking” should take into account of parents’ report, rather than only relying on direct observation.

In consequence, this study showed high inter-rater reliability of N5E2 on its total score and moderate to high inter-rater reliability on its item scores for most of the items. This was a very important step for developing the new screening tool for neuromotor development of children. Specifically, taking account of the fact that a sufficient level of inter-rater reliability

was found based on the ratings of two pediatricians with difference in their subspecialty and amount of clinical experience, N5E2 could be a useful tool in real clinical settings where the clinicians' background knowledge and experiences are quite diverse.

As the next step of development of N5E2, further studies are essential. First, *as this study was conducted with children aged one to five years who visited prosthetic outpatient service because of our purpose of administering the tool to at-risk individuals, the sample did not include many healthy children aged 2 years. The sample size of this study was also small though it was almost sufficient according to Gwet (2010)'s estimation of sample size necessary to calculate inter-rater reliability [17]. Therefore, to confirm the utility of this tool for 2-year-old children for whom this version of N5E2 is originally developed, it is crucial to collect more data from children aged 2 years. Second, In addition, to examine the reliability of this scale for a broader range of populations, it is important to replicate the results of this study with healthy children; this study did not include any children without a medical diagnosis. Second, Third, to understand the psychometric property of the N5E2 further, and determine the cut-off score indicating future risk, it is necessary to investigate its criterion-related validity, such as concurrent validity and predictive validity, as well as specificity and sensitivity on several different developmental/neuromotor disorders. Third, Forth, to broaden the utility of this screening tool, it is vital to test whether the results of rating would change if the tool is administered by professionals other than pediatricians, such as nurses and psychologists.*

## **Conclusion**

The results of the study showed moderate to high inter-rater reliability on individual items in the N5E2, and high reliability on the sum score of the 11 rated items. Taking account of the fact that this level of agreement was reached among the raters with different level of clinical experiences, N5E2 can potentially be an efficient and useful tool to assess neuromotor development in young children in real clinical settings where clinicians experience varies.

## **Acknowledgement**

This study is a part of the project related to the Japan Environment and Children's Study (JECS) that is conducted and funded by the Ministry of Environment of Japan.

## **Potential conflict of interest report**

The authors have no financial or personal relations that could pose a conflict of interest.



## References

- [1] Bright Futures Steering Committee, and Medical Home Initiatives for Children With Special Needs Project Advisory Committee. Identifying infants and young children with developmental disorders in the medical home: An algorithm for developmental surveillance and screening. *Pediatrics* 2006; 118: 405-20.
- [2] Frankenburg WK, Dodds J, Archers P, Shapiro H, Bresnick B. Denver Developmental Screening Test II. Denver: Denver Developmental Materials; 1990.
- [3] Newborg J. Battelle Developmental Inventory, Second Edition (BDI-2). Itasca: Riverside; 2004.
- [4] Squires J, Bricker D. Ages and Stages Questionnaires, Third Edition (ASQ-3). Baltimore: Brookes Publishing; 2009.
- [5] Radecki L, Sand-Loud N, O'Connor KG, Sharp S, Olson LM. Trends in the use of standardized tools for developmental screening in early childhood: 2002–2009. *Pediatrics* 2011; 128: 14-9.
- [6] Noritz GH, Murphy NA, Neuromotor Screening Expert Panel. Motor delays: Early identification and evaluation. *Pediatrics* 2013; 131: e2016-27.
- [7] Hall DMB, Elliman D. *Health for all children*, revised 4<sup>th</sup> Edition. Oxford: Oxford University Press; 2003
- [8] Centers for Disease Control and Prevention. Data and statistics for cerebral palsy: prevalence and characteristic. 2017. (<https://www.cdc.gov/ncbddd/cp/data.html>. Accessed on 15 May 2017.)
- [9] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* (5<sup>th</sup> ed.). Washington, DC: Author; 2013.
- [10] Bhat AN, Galloway JC, Landa RJ.. Relation between early motor delay and later communication delay in infants at risk for autism. *Infant Behav Dev* 2012; 35: 838-46.
- [11] Teitelbaum P, Teitelbaum O, Nye J, Fryman J, Maurer RG. Movement analysis in infancy may be useful for early diagnosis of autism. *Proc Natl Acad Sci U S A*. 1998; 95: 13982-7.
- [12] Brossard-Racine M, Shevell M, Snider L, Bélanger SA, Majnemer A. Motor skills of children newly diagnosed with Attention Deficit Hyperactivity Disorder prior to and following treatment with stimulant medication. *Res Dev Disabil* 2012; 33: 2080-7.
- [13] Kaiser ML, Schoemaker MM, Albaret JM, Geuze RH. What is the evidence of impaired motor skills and motor control among children with attention deficit hyperactivity disorder (ADHD)? Systematic review of the literature. *Res Dev Disabil* 2014; 36C: 338-57.
- [14] Vuijk PJ, Hartman E, Mombarg R, Scherder E, Visscher C. Associations between academic and motor performance in a heterogeneous sample of children with learning disabilities. *J Learn Disabil* 2011; 44: 276-82.

- [15] Gillberg C. The ESSENCE in child psychiatry: Early symptomatic syndromes eliciting neurodevelopmental clinical examinations. *Res Dev Disabil* 2010; 31: 1543-51.
- [16] Altman DG. *Practical statistics for medical research*. London: Chapman and Hall; 1991.
- [17] Kilem L. Gwet, Ph.D. *Inter-Rater Reliability Discussion Corner*. 2010.  
([http://agreestat.com/blog\\_irr/sample\\_size\\_determination.html](http://agreestat.com/blog_irr/sample_size_determination.html). Accessed on 17 Dec 2017.)

Appendix 1: Administration procedure and scoring criteria for each examined items  
in the Neuromotor 5-minutes Exam 2-year old version

- (1) *Retrieving a rolling ball by a pediatrician.* The pediatrician rolls a tennis ball toward the child and observes whether he/she retrieves the ball with his/her hands. If he/she does, a 2 point is assigned.
- (2) *Gait.* Throughout the session, the child is observed while walking. If his/her gait is clearly abnormal (e.g., stamping, dragging one leg, etc.) from the examiner's perspective, a 2 point is assigned. If the child cannot walk without support, a 2 point is assigned.
- (3) *Toe-walking.* The pediatrician observes how the child walks and see whether the child walks without touching his/her heel on the ground. If so, a 2 point is assigned. If the child cannot walk, the item should be skipped without being scored.
- (4) *Asymmetry.* The child is observed while he is engaged in various activities (e.g., walking, sitting, retrieving a ball, etc.). If the child's posture appears to be clearly asymmetric, a 2 point is assigned.
- (5) *Age at unsupported walking.* The pediatrician asks the child's parent when the child started walking without support. The answer should be scored as following: at 18 month or after = 2 point, between 15 and 17 month = 1 point, at 14 month or before = 0 point. If the child was born prematurely, his/her corrected age should be used for scoring.
- (6) *Speaking in two-word understandable sentence.* The pediatrician observes a conversation between the child and his/her parent. If the child was not observed to speak a two-word understandable sentence, then the pediatrician should ask the parent whether the child can speak a two-word sentence. 2 point is assigned when the child is not observed/reported to speak a two/word understandable sentence.
- (7) *Hypotonus.* The child is asked to sit down on a chair without leaning on the backrest. If he could not maintain his posture (e.g., collapsing onto the chair or rocking), a 2 point is assigned.
- (8) *Hypertonus.* The pediatrician bends the child's ankles and observes their range of dorsiflexion. If the pediatrician finds limited range of motion or resistance, a 2 point is assigned.
- (9) *Eye movement.* The pediatrician presents a toy in front of the child, asking him/her to follow its movement. Then, the pediatrician moves the toy up, down, right, and left, and observes the child's eye movement. A 2 point is assigned when obvious eye movement problem, (e.g., squint), was observed.
- (10) *Vision.* The pediatrician asks the child's parent whether they he/she has any concerns about the child's vision. If the parent has some concerns, a 2 point is assigned. A

2 point is also assigned if the child has already received medical treatment for his/her vision (e.g., glasses were prescribed).

- (11) *Hearing.* The pediatrician asks the child's parent whether they he/she has any concerns about the child's hearing ability. If the parent has some concerns, a 2 point is assigned. A 2 point is also assigned if the child has already received medical treatment for his/her vision (e.g., A hearing aid was prescribed).

## Neuromotor 5-minutes exam 2 year old version

Examinee's information

ID : \_\_\_\_\_

Age : \_\_\_\_ years \_\_\_\_ months

Gender : \_\_\_\_\_

Examiner's name: \_\_\_\_\_ Examiner's occupation: \_\_\_\_\_

Examination date

Please mark the most appropriate option for each item.

Gross motor characteristics			
1. Retrieving a ball	2pt: Did not	1pt : Could not judge	0pt : Did
2. Gate	2pt: Abnormal	1pt : Could not judge	0pt : Normal
3. Toe-walking	2pt: Did	1pt : Could not judge	0pt : Did not
4. Asymmetry	2pt: Clearly so	1pt : Could not judge	0pt : No
5. Age at unsupported walking	2pt: 18 month or after	1pt : 15-17 month	0pt : 14 month or before
Language			
1. 2-word sentence (parent's report)	2pt: Do not speak	1pt : Could not judge	0pt : Speaks
Neurological function			
1. Hypotonus	2pt : Could not maintain posture	1pt : Could not judge	0pt : Could maintain posture
2. Hypertonus (dorsiflexion)	2pt : Had limited range of motion/resistance	1pt : Could not judge	0pt : No limited range of motion
3. Eye movement	2pt : Found problem	1pt : Could not judge	0pt : No problem
Perception			
1. Vision (parent's report)	2pt : Has some concerns	1pt : Could not judge	0pt : No concern
2. Hearing (parent's report)	2pt : Has some concerns	1pt : Could not judge	0pt : No concern

Total score : \_\_\_\_\_

Special notes : \_\_\_\_\_