



Development of the Korean Developmental Screening Test for Infants and Children (K-DST)

Hee Jung Chung, MD, PhD¹, Donghwa Yang, MD², Gun-Ha Kim, MD, PhD³, Sung Koo Kim, MD, PhD⁴, Seoung Woo Kim, MD, PhD⁵, Young Key Kim, MD⁶, Young Ah Kim, PhD⁷, Joon Sik Kim, MD, PhD⁸, Jin Kyung Kim, MD, PhD⁹, Cheongtag Kim, PhD¹⁰, In-Kyung Sung, MD¹¹, Son Moon Shin, MD, PhD¹², Kyung Ja Oh, PhD¹³, Hee-Jeong Yoo, MD, PhD¹⁴, Hee Joon Yu, MD, PhD⁴, Seoung-Joon Lim, MD¹⁵, Jeehun Lee, MD, PhD¹⁶, Hae-Ik Jeong, MD¹⁵, Jieun Choi, MD, PhD¹⁷, Jeong-Yi Kwon, MD, PhD¹⁸, Baik-Lin Eun, MD, PhD³

¹Department of Pediatrics, National Health Insurance Service Ilsan Hospital, Goyang, Korea; ²Department of Pediatrics, Severance Children's Hospital, Yonsei University College of Medicine, Seoul, Korea; ³Department of Pediatrics, Korea University College of Medicine, Seoul, Korea; ⁴Department of Pediatrics, Dongtan Sacred Heart Hospital, Hallym University College of Medicine, Hwasung, Korea; ⁵Department of Rehabilitation Medicine, National Health Insurance Service Ilsan Hospital, Goyang, Korea; ⁶Department of Psychiatry, National Health Insurance Service Ilsan Hospital, Goyang, Korea; ⁷Yonsei University Human Behavior Research Institute, and HUNO, Seoul, Korea; ⁸Department of Pediatrics, Keimyung University School of Medicine, Daegu, Korea; ⁹Department of Pediatrics, Catholic University of Daegu School of Medicine, Daegu, Korea; ¹⁰Department of Psychology, Seoul National University, Seoul, Korea; ¹¹Department of Pediatrics, Catholic University College of Medicine, Seoul, Korea; ¹²Department of Pediatrics, Inje University Busan Paik Hospital, Inje University College of Medicine, Busan, Korea; ¹³Department of Psychology, Yonsei University, Seoul, Korea; ¹⁴Department of Psychiatry, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Korea; ¹⁵Korean Pediatric Association, Seoul, Korea; ¹⁶Department of Pediatrics, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea; ¹⁷Department of Pediatrics, Seoul Metropolitan Government - Seoul National University Boramae Medical Center, Seoul, Korea; ¹⁸Department of Physical and Rehabilitation Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

Background: Most developmental screening tools in Korea are adopted from foreign tests. To ensure efficient screening of infants and children in Korea, a nationwide screening tool with high reliability and validity is needed.

Purpose: This study aimed to independently develop, standardize, and validate the Korean Developmental Screening Test for Infants and Children (K-DST) for screening infants and children for neurodevelopmental disorders in Korea.

Methods: The standardization and validation conducted in 2012–2014 of 3,284 subjects (4–71 months of age) resulted in the first edition of the K-DST. The restandardization and revalidation performed in 2015–2016 of 3.06 million attendees of the National Health Screening Program for Infants and Children resulted in the revised K-DST. We analyzed inter-item consistency and test-retest reliability for the reliability analysis. Regarding the validation of K-DST, we examined the construct validity, sensitivity and specificity, receiver operating characteristic curve analysis, and a criterion-related validity analysis.

Results: We ultimately selected 8 questions in 6 developmental domains. For most age groups and each domain, internal consistency was 0.73–0.93 and test-retest reliability was 0.77–0.88. The revised K-DST had high discriminatory ability with a sensitivity of 0.833 and specificity of 0.979. The test supported construct validity by distinguishing between normal and neurodevelopmentally delayed groups. The language and cognition domain of the revised K-DST was highly correlated with the K-Bayley Scales of Infant Development-II's Mental Age Quotient ($r=0.766, 0.739$), while the gross and fine motor

domains were highly correlated with Motor Age Quotient ($r=0.695, 0.668$), respectively. The Verbal Intelligence Quotient of Korean Wechsler Preschool and Primary Scales of Intelligence was highly correlated with the K-DST cognition and language domains ($r=0.701, 0.770$), as was the performance intelligence quotient with the fine motor domain ($r=0.700$).

Conclusion: The K-DST is reliable and valid, suggesting its good potential as an effective screening tool for infants and children with neurodevelopmental disorders in Korea.

Key words: Developmental screening test, Infant and child, Korean Developmental Screening Test for Infants and Children, Standardization, Validation

Key message

Question: Can the Korean Developmental Screening Test for Infants and Children (K-DST) be a useful screening tool for infants and children in Korea?

Finding: The K-DST has high reliability (internal consistency of 0.73–0.93, test-retest reliability of 0.77–0.88) and a high discriminatory ability with a sensitivity of 0.833 and specificity of 0.979.

Meaning: The K-DST is an effective and reliable screening tool for infants and children with neurodevelopmental disorders in Korea.

Introduction

The assessment of growth and development is essential for growing children. Particularly, infants and children are the most important periods for development, and developmental evaluation during this period may reduce the chances of future disorders and prevent secondary sequelae. The most common clinical symptom of neurodevelopmental disorders is the lack of developmental skills consistent with their age. Therefore, if early screening of infants and children who may have future developmental problems is possible through assessment, then a more accurate assessment should be conducted to plan for appropriate treatments and rehabilitation.

National Health Screening Program for Infants and Children (NHSPIC) has been rolled out nationwide since November 2007 to keep track of the growth and development of infants and children and to provide proper education programs to caregivers, with the introduction of a health check-up program suitable for the age of infants and children. NHSPIC is conducted for the first to seventh rounds as protocols divided by the scope of questionnaire and physical examination, physical measurements, development evaluation and consultation, health education, and oral examination. Among these, the developmental evaluation is the major screening item of NHSPIC, with the first 9–12 months, the second in 18–24 months, the third in 30–36 months, the fourth in 42–48 months, the fifth in 54–60 months, and the sixth in 66–71 months of age, all but 4–6 months, the first NHSPIC period.¹⁾

The developmental screening tools used in the NHSPIC should be able to comprehensively assess the developmental domains to sensitively identify children at risk of suffering from neurodevelopmental disorders. As part of the NHSPIC project, Korean Developmental Screening Test for Infants and Children (K-DST) is a new developmental screening tool tailored as per the characteristics of the Korean children.²⁻⁴⁾ It was developed by experts in related fields including the 'The Korean Pediatric Society,' 'Korean Society of Pediatric Rehabilitation and Developmental Medicine,' 'Korean Academy of Child and Adolescent Psychiatry,' and 'Korean Psychological Association.' For pre-school children under 6 years of age (4–71 months), K-DST is a parent-reported screening test that allows parents to directly monitor the development of their children, taking less time and is able to quickly and effectively identify any developmental delays in primary care institutions. The tool was developed to allow the use of not only printed test sheets but also on-line test to increase the accessibility to examinees. This study reviewed the development of K-DST and evaluated the standardization and validation of this tool to see how accurately it could identify the target at risk of neurodevelopmental disorders. Moreover, we investigated the reliability and validity of the revised K-DST, which established new cutoff points by means of restandardization in 2017 based on the 3.06 million cases accumulated since the first edition of K-DST was conducted in 2014. This study was conducted after obtaining approvals from the Insti-

tutional Review Board of Korea University Guro Hospital (2013GR0048, 2016GR0083, 2017GR0223).

Methods

1. Study design

Following the earlier study in 2010, preliminary items were developed, and the standardization and validation of the test were conducted between 2012 and 2014 to test reliability. As a result, the first edition of K-DST was developed. Since then, the restandardization and revalidation of the test were conducted in 2017 involving the 3.06 million infants and children who checked for NHSPIC from 2015 to 2016, and the revised K-DST was subsequently published.

2. Prior study

In the earlier study from August 2010 to April, the content composition, characteristics, and discrimination procedures of Korean and foreign development screening tools for infants and children were reviewed. From the expert meeting, the focused group interview, professional opinions on the assessment domains and questions were reviewed. Based on these data analyses, future test tools were proposed.

3. Standardization and validation of K-DST

1) Preliminary items

The preliminary items were developed for infants and children between 4 months and 83 months of age. NHSPIC was conducted for children between 4 months and 71 months of age. However, groups that extended from 4 months to 83 months of age were selected considering factors such as maximizing the utilization of tools and statistical accuracy as well as further study expansion. The subdomain of the examination consisted of a total of 6 developmental domains, including the gross motor, fine motor, cognition, language, social skills, and self-help, based on the review of the results of the prior studies.

The test was conducted in the form of a parent report and scored on a 4-point scale. Data were collected through the distribution of test papers to the child's parents at either the hospitals or daycare centers. Eventually, 876 cases were recorded, and they filled a questionnaire. On the 4-point scale of the collected data, the 'always possible' was 1 and the rest was converted to zero and then a 2-parameter item response theory (IRT) model was used to estimate the level of difficulty and the discrimination of question, and gauge the ability of each subject.⁵⁾ After obtaining the function of the subject's age and ability by using the Gompertz function, we then used the item characteristic curve (ICC) in combination to search for the appropriate age group to distinguish each item well.⁶⁾

2) Standardization of K-DST

(1) Data collection

Two rounds of data collection were conducted from May

to October 2013. Based on the 2010 Population and Housing Census, stratification sampling was conducted according to age, geographical location, regional size, and gender. The population group was assigned to a distinct age range, and the data were collected with an aim of collecting a total of 3,100 valid samples. The data collection process was aimed at recruiting a total of 1,500 cases by allocating the first large on-line panel based on region, academic background, and gender. After identifying the deficiencies based on the regional target samples based on the data collected from the first round, secondary data collection was conducted by determining the suitability of demographic information in a total of 40 hospitals and daycare centers countrywide. Finally, 8 questions for each domain were selected from the initial questions that consisted of a total of 10 questions through review and prior investigation. Additionally, the self-help domain had the characteristics of developing after acquiring a certain developmental skill, making it difficult to independently evaluate the domain of children under 18 months old. Therefore, the final structure of the examination was also finalized by deciding to measure the self-help domain after 18 months. The final questions selected included 57 items in the gross motor, 56 in the fine motor, 59 in the cognition, 53 in the language, 56 in the social skills, and 44 in the self-help domain.

(2) Data analysis

The reliability analysis of the test assessed the coefficient of internal consistency in the developmental domains with age and test-retest reliability. Within each age group, the coefficient of internal consistency for each domain was calculated using Cronbach alpha, as an indicator of how consistent the multiple questions are in measuring construct concepts. Generally, this value is interpreted as having an acceptable reliability if it has a value of 0.7 or higher.⁷⁾ For the test-retest reliability, 300 children were randomly selected from a population group and re-examined 2–4 weeks later. A total of 106 children completed the re-examination.

Item analysis was conducted through the IRT and the ICC, and investigation done on whether each question had a high discriminability for children of a particular age.^{5,6)}

3) Validation of K-DST

In regards to K-DST validation, we conducted construct validity analysis, sensitivity and specificity analysis, receiver operating characteristic (ROC) curve analysis, and criterion-related validity analysis to determine whether K-DST's items accurately reflect the main domains of development.

To verify the construct validity, confirmatory factor analysis was conducted, where 8 questions were affected in 6 developmental domains for each age group.⁸⁾ Cutoff points were calculated based on the score of the K-DST population group. The cutoff points consist of scores corresponding to the -2 standard deviation (SD), -1 SD of the domain scores for each age group, and less than -2 SD correspond to 'Recommendation for further evaluation, -2 SD and less than -1 SD correspond to 'need for follow-up'.

To determine the sensitivity and specificity of K-DST, 184 cases of subjects diagnosed with neurodevelopmental disorders and 206 cases in the control group were analyzed, according to the diagnosis by the doctors and the cutoff point of K-DST. Groups having scores below the cutoff point in one or more of the 6 domains of K-DST were referred to as 'clinical group,' while those with points above the cutoff point were classified as 'control group' and were subsequently compared with clinical diagnosis. In addition, ROC curve analysis was performed in order to evaluate clinical discrimination as a screening measure of K-DST.

In order to confirm the criterion-related validity of the K-DST, we examined the Korean Bayley Scales of Infant Development-II (K-BSID) for children aged between 4 and 42 months, and the Korean Wechsler Preschool and Primary Scales of Intelligence (K-WPPSI) for children aged between 30 and 71 months.^{9,10)} Finally, the correlation between the subscale values of each test and the total scores of each subdomain of K-DST was analyzed.¹¹⁾

4. Restandardization and revalidation of K-DST

1) Data collection

The restandardization and revalidation was conducted from September 2016 to September 2017 for more than 3.06 million infants and children who were involved in NHSPIC across the country from 2015. The control group for validation analysis administered all the K-DST questionnaires according to age classified as normal in K-DST in a total of 12 research hospitals. Children over the age of 9 months and younger than 6 years undertook K-BSID-II or K-WPPSI-R (or K-WPPSI-IV). The final number of people sampled was 235 for the control group and 413 for the clinical group.

2) Data analysis

Using data from more than 3.06 million subjects from the National Health Insurance Service, new cutoff points were determined after analyzing the average, median, standard deviation, maximum, minimum, and score distribution for subdomain of each age group. During the 1st-test development, the sample size was smaller compared to 3.06 million cases and the distribution of the population was unidentified, thus, the normal distribution was assumed and the cutoff points using standard deviation were established. However, since large-scale data was used for restandardization, the distribution of the population sample was assumed to be asymptotic and the cutoff points were set via percentile scores. Since this test was designed to discriminate infants and children with developmental delay, infants with normal development mostly had high test scores. Therefore, the distribution of the test scores was a one-sided distribution rather than a normal distribution, and when the cutoff point was established using standard deviation from the skewed distribution, the percentage of children from a particular section failed to match the percentage of normal distribution. Therefore, using percentile scores seemed reasonable than arithmetically using -2 SD and -1 SD. In regards to normal and

clinical groups, a validity analysis was performed via ROC curve analysis and investigated the changes in sensitivity, specificity, and accuracy using both the previous cutoff points and the new cutoff points. K-DST scores were categorized into 2: control groups (-1 SD or higher), and clinical groups (less than -2 SD) in order to analyze whether they are properly discriminated between normal and clinical groups. To analyze criterion-related validity, we investigated the relevance of K-BSID-II, K-WPPSI test in comparison with the revised K-DST.

Results

1. Standardization and validation of the first edition of K-DST

1) Basic statistical analysis

The number of population group was 3,284, the results based on age, region, regional size, and gender were presented in Supplementary Table 1. The distribution by region was consistent with the 2010 Population and Housing Census, with the highest rates being in the metropolitan area (n=1,022, 31.1%). The distribution of the population groups was evenly distributed between 100–200 cases for each age group, with 1,602 cases being men (48.8%) and 1,682 cases (51.2%) being women (Supplementary Table 2). For each age group, the mean and standard deviation for each area of the examination paper were presented in Supplementary Table 3.

Table 1. Internal consistency of developmental domains of the K-DST 1st edition by monthly age group

Age group (mo)	Cronbach alpha					
	Gross motor	Fine motor	Cognition	Language	Social skills	Self-help
4-5	0.77	0.84	0.77	0.80	0.78	-
6-7	0.76	0.81	0.77	0.80	0.74	-
8-9	0.88	0.73	0.77	0.83	0.78	-
10-11	0.66	0.76	0.77	0.80	0.77	-
12-13	0.89	0.84	0.71	0.81	0.83	-
14-15	0.80	0.74	0.69	0.85	0.76	-
16-17	0.89	0.87	0.84	0.87	0.85	-
18-19	0.81	0.87	0.90	0.89	0.90	0.84
20-21	0.82	0.82	0.84	0.90	0.84	0.83
22-23	0.89	0.88	0.87	0.91	0.89	0.90
24-26	0.82	0.76	0.86	0.90	0.90	0.83
27-29	0.84	0.74	0.84	0.90	0.89	0.87
30-32	0.79	0.79	0.89	0.87	0.89	0.82
33-35	0.84	0.84	0.86	0.91	0.87	0.83
36-41	0.76	0.86	0.85	0.93	0.90	0.88
42-47	0.76	0.88	0.85	0.88	0.86	0.85
48-53	0.80	0.90	0.84	0.90	0.90	0.86
54-59	0.81	0.87	0.78	0.84	0.83	0.82
60-65	0.87	0.88	0.82	0.84	0.83	0.78
66-71	0.84	0.87	0.88	0.87	0.87	0.84

K-DST, Korean Developmental Screening Test for Infants and Children.

2) Reliability analysis

There were medium to high levels of internal consistency (0.71–0.93) within all age groups and subdevelopmental domains, except for the gross motor domain (10–11 months), which had a slightly lower internal consistency of about 0.66 (Table 1). The coefficients of test-retest reliability were as high as 0.77–0.88 (0.86 for gross motor, 0.83 for fine motor, 0.86 for cognition, 0.88 for language, 0.83 for social skills, and 0.77 for self-help).

3) Item analysis

The discrimination ability and difficulty level were calculated using the IRT and are shown in Supplementary Table 4. As a result of the analysis, the discrimination ability for most of the questions was 2.3 or higher. The difficulty level of the question was observed to increase with an increase in the number of question, confirming that the difficulty level was fitting.

4) Validity analysis

Confirmatory factors analysis for determining the construct validity showed that the root mean square error approximation (RMSEA) value was within an acceptable range of 0.052–0.097 and the factor loading for each item was also estimated to be high (Table 2, Supplementary Table 5).

Regarding the sensitivity and specificity analysis, 10 of the 206 cases were classified as developmental abnormalities for children in the control group, and 163 of the 184 cases in the clinical group were classified as developmental abnormalities (sensitivity, 0.886; specificity, 0.951; accuracy, 0.921). The false-positive value was 4.9%, and the false-negative value was 11.4%.

As a result of ROC curve analysis, most of the domains were well classified and exhibited larger area under the curve (AUC) values (0.763–0.999). Cerebral palsy was screened by the gross and fine motor domains with high accuracy (AUC=0.971, 0.936), and developmental language disorder was screened by language and social domains with high accuracy (AUC=0.924, 0.957). Autism spectrum disorder was screened by the social

Table 2. Result of model conformity of the K-DST 1st edition by monthly age group

Age group (mo)	Number	CFI	RMSEA	Age group (mo)	Number	CFI	RMSEA
4-5	121	0.646	0.097	24-26	142	0.733	0.075
6-7	121	0.661	0.089	27-29	142	0.776	0.065
8-9	120	0.696	0.090	30-32	137	0.700	0.082
10-11	143	0.671	0.080	33-35	131	0.792	0.063
12-13	145	0.687	0.088	36-41	195	0.846	0.059
14-15	127	0.665	0.081	42-47	203	0.831	0.056
16-17	122	0.730	0.092	48-53	181	0.856	0.056
18-19	123	0.737	0.085	54-59	211	0.829	0.052
20-21	120	0.692	0.085	60-65	168	0.760	0.069
22-23	129	0.828	0.074	66-71	183	0.794	0.064

K-DST, Korean Developmental Screening Test for Infants and Children; CFI, comparative fit index; RMSEA, root mean square error of approximation.

skills domain with the highest accuracy (AUC=0.999), while intellectual disability was also screened by cognition, language, and social skills domains with high accuracy (AUC=0.951, 0.959, 0.964) (Table 3).

In the case of K-BSID-II test, MDI and PDI results showed no detailed scores for less than 50, thus the correlation was performed by converting them into the Mental Age Quotient and

Motor Age Quotient. In the criterion-related validity analysis, the correlation between the first edition K-DST and K-BSID-II was relatively high ($r=0.38-0.68$). Particularly, the correlation between Bayley's mental age quotient (Mental Q) and K-DST's language and social skills domains were high ($r=0.68, 0.67$), and

Table 3. ROC curve analysis results of K-DST 1st edition by neurodevelopmental disorder

Variable	Area under the curve			
	Cerebral palsy	Developmental language disorder	Autism spectrum disorder	Intellectual disability
Gross motor	0.971	0.763	0.887	0.936
Fine motor	0.936	0.921	0.984	0.952
Cognition	0.860	0.893	0.953	0.951
Language	0.874	0.924	0.952	0.959
Social skills	0.874	0.957	0.999	0.964
Self-help	0.932	0.930	0.954	0.890

ROC, receiver operating characteristic; K-DST, Korean Developmental Screening Test for Infants and Children

Table 4. Results of correlation analysis of K-DST 1st edition, K-BSID-II, and K-WPPSI

Variable	Gross motor	Fine motor	Cognition	Language	Social skills	Self-help
K-BSID-II						
Mental Q	0.38	0.55	0.60	0.68	0.67	0.53
Motor Q	0.57	0.61	0.59	0.61	0.66	0.62
K-WPPSI						
FSIQ	0.64	0.63	0.62	0.65	0.61	0.52
VIQ	0.62	0.62	0.72	0.74	0.61	0.45
PIQ	0.69	0.66	0.64	0.63	0.58	0.48

K-DST, Korean Developmental Screening Test for Infants and Children; K-BSID-II, Korean Bayley Scales of Infant Development-II; Mental Q, mental age quotient; Motor Q, Motor age quotient; K-WPPSI, Korean Wechsler Preschool and Primary Scales of Intelligence; FSIQ, full scale intelligence quotient; VIQ, verbal intelligence quotient; PIQ, performance intelligence quotient.

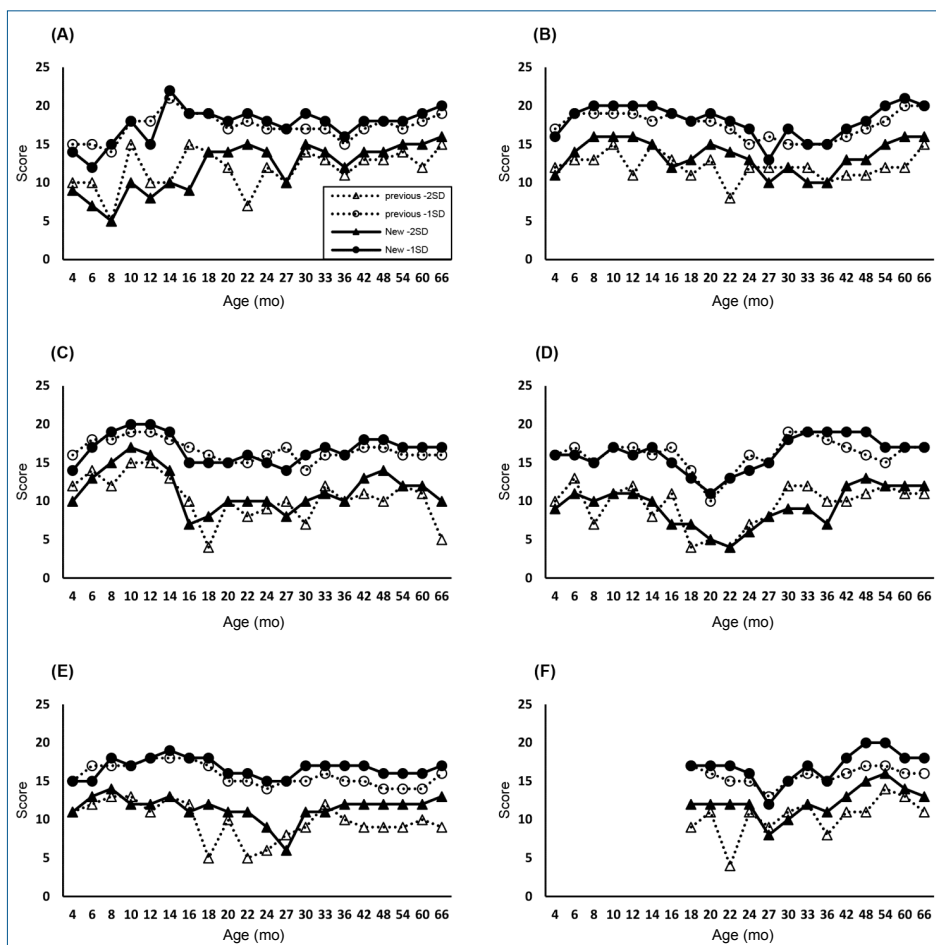


Fig. 1. Change in cutoff points based on monthly age group and subdomains. (A) Gross motor domain; (B) fine motor domain; (C) cognition domain; (D) language domain; (E) social skills domain; (F) self-help domain. SD, standard deviation. Dotted lines represent arithmetically calculated standard deviations; solid lines represent percentage scores.

motor age quotient (Motor Q) and the fine motor domain was also high ($r=0.61$). The correlations between the K-WPPSI's full scale intelligence quotient (FSIQ) and K-DST were 0.52–0.65, which was also showed high correlation. For instance, the verbal intelligence quotient (VIQ) of K-WPPSI was highly correlated with the language domain of K-DST ($r=0.74$), and the performance intelligence quotient (PIQ) with the fine motor domain ($r=0.66$) (Table 4).

2. Restandardization and revalidation of K-DST

1) Basic statistics and new cutoff points

The distribution by age group of population groups for restandardization was shown as in Supplementary Table 6. Approximately 3.06 million people were examined, and most of the subjects were evenly distributed based on their age. However, there were very few cases in certain age groups that did not have NHSPIC (6–7 months, 27–29 months). The change of the cutoff point based on the restandardization is illustrated by the area-by-area diagram (Fig. 1). The cutoff points set using percentiles were relatively lower compared to those set using arithmetic standard deviations.

2) Validity analysis

Sensitivity and specificity analysis revealed that 11 of 235 subjects were classified as having developmental problems in normal children and 350 of 413 patients in the clinical group. Based on previous cutoff points, the sensitivity was 0.847 and the specificity was 0.953, and when the cutoff points were subjected to restandardization, the sensitivity was 0.833 and the specificity

was 0.979. The accuracy was about 0.886, indicating a high predictive accuracy as a screening tool (Table 5).

ROC curve analysis of neurodevelopmental disorders revealed that the AUC values in the gross and fine motor domains for cerebral palsy (AUC: 0.969, 0.907), the language and social skills domains for developmental language disorders (AUC: 0.980, 0.998), the language and social skills domains for autism spectrum disorder (AUC: 0.974, 0.943), and the cognition, language, and social skills domains for intellectual disabilities (AUC: 0.960, 0.974, 0.979), were high, and correlated very well (Table 6).

Based on the criterion-related validity analysis, the 6 subdomains of revised K-DST showed a high correlation with the K-BSID-II test scores, and also showed high correlations with the K-WPPSI-R and K-WPPSI-IV tests. Mental Age Quotient showed high correlations with the language, cognition, and social skills domains of the revised K-DST ($r=0.766, 0.739, 0.810$) and the Motor Age Quotient was highly correlated with K-DST's gross and fine motor domains ($r=0.695, 0.668$). The K-WPPSI-R and K-WPPSI-IV tests also exhibited a high correlation with the revised K-DST. For instance, VIQ of Wechsler's tests showed the highest correlation with the language domain ($r=0.770$ for K-WPPSI-R, 0.701 for K-WPPSI-IV), while PIQ also showed the highest correlation with the fine motor domain ($r=0.681$ for K-WPPSI-R, 0.700 for K-WPPSI-IV) (Table 7).

Discussion

Since the 2000s, the low birth rate coupled with an aging society has led to the widespread perception that social responsibility should be enhanced to create an environment favorable for childbirth and child nurturing. NHSPIC is in line with the basic direction of the nation's welfare policy, which aims to progress

Table 5. Sensitivity and specificity analysis of restandardization (before and after setting of new cutoff points)

Variable	Previous cutoff points		New cutoff points	
	Control group	Clinical group	Control group	Clinical group
Normal	224	63	230	69
Abnormal	11	350	5	344
Total	235	413	235	413
Sensitivity	0.847		0.833	
Specificity	0.953		0.979	
Accuracy	0.886		0.886	

Table 6. ROC curve analysis results of K-DST revised edition by neurodevelopmental disorder

Variable	Area under the curve			
	Cerebral palsy	Developmental language disorder	Autism spectrum disorder	Intellectual disability
Gross motor	0.969	0.881	0.769	0.858
Fine motor	0.907	0.944	0.894	0.943
Cognition	0.836	0.940	0.914	0.960
Language	0.849	0.980	0.974	0.974
Social skills	0.881	0.998	0.943	0.979
Self-help	0.894	0.893	0.860	0.889

ROC, receiver operating characteristic; K-DST, Korean Developmental Screening Test for Infants and Children.

Table 7. Results of correlation analysis of K-DST revised edition, K-BSID-II, and K-WPPSI

Variable	Gross motor	Fine motor	Cognition	Language	Social skills	Self-help
K-BSID-II						
Mental Q	0.475	0.690	0.739	0.766	0.810	0.508
Motor Q	0.695	0.668	0.656	0.620	0.706	0.552
K-WPPSI-R						
FSIQ	0.613	0.743	0.648	0.761	0.780	0.548
VIQ	0.577	0.713	0.633	0.770	0.765	0.577
PIQ	0.564	0.681	0.605	0.667	0.703	0.441
K-WPPSI-IV						
FSIQ2	0.583	0.737	0.697	0.691	0.714	0.498
VIQ2	0.527	0.65	0.651	0.701	0.729	0.487
PIQ2	0.513	0.700	0.616	0.592	0.606	0.421

K-DST, Korean Developmental Screening Test for Infants and Children; K-BSID-II, Korean Bayley Scales of Infant Development-II; Mental Q, mental age quotient; Motor Q, Motor age quotient; K-WPPSI, Korean Wechsler Preschool and Primary Scales of Intelligence; FSIQ, full scale intelligence quotient; VIQ, verbal intelligence quotient; PIQ, performance intelligence quotient.

towards a better welfare state through investing efficiently in the health of the people.

One of the effective ways to assess development is to adopt standardized developmental tools. Developmental screening test is a test developed to screen children who do not lie within the normal range of development and determines whether professional diagnosis is required.¹²⁾ The development of infants and children should be considered as a continuous series of functions that are variable. Therefore, developmental screening test should be conducted more than once, and it is essential to repeat the test in time series in order to monitor the developmental changes of infants and children with age.^{13,14)}

The developmental testing tools used in Korea are either developed in-house or in foreign countries. The latter was used after some amendments and standardization by Korean standards or re-standardizing foreign tests. Some of the developmental tests developed in foreign countries include Denver Development Assessment Testing, Bayley Scales of Infant Development (BSID), Early Screening Inventory (ESI), and Developmental Indicators for Assessment of Learning (DIAL).^{9,15-17)} The developmental testing tools that were developed in foreign countries and standardized in Korea include Korean Denver-II, Korean DIAL-3, Korean Ages and Stages Questionnaire (K-ASQ), Korean Bayley Scales of Infant Development (K-BSID), and ESI-Revisited (ESI-R).^{9,15-18)} However, most test tools used in Korea are limited to children living in cities and thus not reflective of the national samples based on the standardization process. Also, some tests did not report the characteristics of samples such as areas or demographic information.

The NHSPIC, which has been implemented in Korea since 2007, utilizes developmental screening tools to evaluate development, but initially used either K-ASQ or Denver-II. However, Denver-II is difficult to apply to all children in the Korean medical environment, as it is time-consuming since the examiner must perform the test in person. Therefore, K-ASQ was favored as parents would fill out the questionnaire themselves in the NHSPIC. K-ASQ is a Korean version of Ages & Stage Questionnaires, the second edition developed in the United States and standardized for Korea.^{19,20)} However, K-ASQ is a test developed in the United States, which makes it unsuitable for Korean infants since children grow in culturally different environments.

To verify the standardization and validity of K-DST, the reliability analysis was conducted by evaluating the coefficient of internal consistency, construct validity analysis, sensitivity and specificity analysis, ROC curve analysis, and criterion-related validity analysis.

The results obtained by the analyses were as follows. In the item analysis identified by the IRT, the variability of most questions was higher than 1.7, which was very high by the criterion of Baker.²¹⁾ The internal consistency coefficient for each domain for reliability analysis showed good Cronbach alpha values of about 0.73–0.93 in most domains.

Compared with K-ASQ, showing that the RMSEA value was significantly lower in the model's fit at 0.096–0.118 over 24

months, K-DST's Confirmatory Factor Analysis confirmed that the RMSEA value was within the acceptable range of 0.052–0.097 for all the age groups.²²⁾ Therefore, the structure of K-DST was verified.

The sensitivity of the K-ASQ was 0.75 and the specificity was 0.86 and that level was reportedly good.¹⁸⁾ The sensitivity and specificity of the K-ASQ in the 30 and 36 months groups were 0.88–0.96 and the accuracy was 0.92 and 0.89, respectively, with good discrimination power in general. In contrast, in the 60 months group, the specificity was high at 0.95, but the sensitivity was relatively low at about 0.65.²³⁾ In comparison, the results of the sensitivity and specificity analysis of the first edition of K-DST showed that the sensitivity was 0.886 and the specificity was 0.951, which is generally higher compared to that of K-ASQ.

In addition, The K-ASQ had the ability to select clinical groups such as intellectual disability and autism spectrum disorders as a risk group for neurodevelopmental disorders, but failed to select the groups with delayed language development such as developmental language disorders.²³⁾ In contrast, the ROC curve analysis results of the first edition K-DST showed an AUC value of 0.9 or higher for each disease. In particular, in the case of developmental language disorders, AUC values of 0.924 for language and 0.957 for social skills showed high correlation with the subdomains associated with neurodevelopmental disorders.

In the criterion-related validity analysis, K-ASQ showed statistically significant correlation with criterion variables but not generally high. There were also areas where there were no significant correlations related to other measures. For example, K-ASQ's gross motor domain in the 30-month age group had no significant correlation with K-BSID-II's motor quotient score.²³⁾ This suggests that it may be insufficient to predict the actual motor functions for that particular month. In comparison, the correlations between mental quotient of K-BSID-II and the cognition, language, and social skill domains areas of the first edition of K-DST showed high correlation at 0.60, 0.68, and 0.67, respectively, and the correlations between motor quotient of K-BSID-II and the gross motor and fine motor domains of the first edition K-DST were also high at about 0.57 and 0.61. Between K-WPPSI and first edition K-DST, the correlation between VIQ and the cognition and language domains of the first edition K-DST was higher than compared to that of the other developmental domains. In addition, the correlation between PIQ and the gross and fine motor domain was relatively high at 0.69 and 0.66, respectively. These results suggest that although K-DST is a developmental screening test based on parental reports, it has a high correlation with K-BSID-II and K-WPPSI, which are the most widely used test tools for confirming neurodevelopmental disorder. This suggests that K-DST is a highly reliable developmental screening tool for Korean infants and children.

In the case of revised K-DST, the sensitivity slightly decreased (from 0.847 to 0.833) in the newly adjusted cutoff point analysis compared to the existing cutoff point, while the specificity increased slightly (from 0.953 to 0.979). The false-negative values

increased slightly from a value of 15.3% for the previous cutoff points to 16.7% using the new cutoff points, while the false-positive values decreased from 4.7% for previous cutoff points to 2.1% for the new cutoff points, reducing the probability of judging a normal child as a child with developmental problems

According to the ROC curve analysis of the revised K-DST, the highest AUC values were found in the gross motor domain for cerebral palsy, the language domain for developmental language disorder, the social skills and language domains for autism spectrum disorder, and the cognition domain for intellectual disability. Compared with the first edition of K-DST, the revised K-DST showed increasing AUC values of the language and social domains (AUC=0.980, 0.998) in developmental language disorder, and also higher AUC values of the cognition, language, and social domains (AUC=0.960, 0.974, 0.979) in intellectual disability, showing better discrimination power for the revised K-DST compared to that of the first edition K-DST. In the correlation between K-BSID-II and revised K-DST, comparing with the first edition K-DST, the correlation between Mental Age Quotient and revised K-DST increased from 0.60 to 0.739 in the cognition domain, and from 0.68 to 0.766 in the language domain. The correlation between Motor Age Quotient and revised K-DST, also increased from 0.57 to 0.695 in the gross motor domain, and from 0.61 to 0.668 in the fine motor domain, signifying better criterion-related validity than that of the first edition K-DST. In the case of K-WPPSI-R, K-WPPSI-IV and revised K-DST, also showed an increase or slight decrease in the correlation between VIQ and the language domain from 0.74 to 0.701 in K-WPPSI-R, and from 0.74 to 0.770 in K-WPPSI-IV. Increased correlation between PIQ and the fine motor domain was also noted, increasing from 0.66 to 0.681 in K-WPPSI-R and from 0.66 to 0.700 in K-WPPSI-IV. Therefore, we verified that the revised K-DST has a higher criterion-related validity compared with that of the first edition of K-DST. However, the correlation between the gross motor domain of the revised K-DST and K-WPPSI-R and K-WPPSI-IV was somewhat low ($r=0.513, 0.564$). This is because the minimum age of the Wechsler intelligence test should be above 30 months, which is considered to be the main development period of the fine motor domain as opposed to that of the gross motor domain.

The illustration of the results derived from this study is as follows. First, the first edition K-DST showed high sensitivity and specificity compared to K-ASQ, which is known as a suitable screening test, which distinguishes well various neurodevelopmental diseases disorder. In particular, K-DST is an independently developed tool that suits the characteristics of infants in Korea, not a modified or standardized tool for existing foreign tests, and the reliability and validity of the tests were verified through various analyses. Compared to the K-BSID-II and K-WPPSI, the correlations for reliability and validity were high, therefore the validity as a screening test is confirmed. Second, the revised K-DST has become a more powerful discriminating tool because the AUC values are higher in the ROC curve analysis, and the correlation coefficient is higher in the criterion-related

validity analysis as compared to that of the first edition of K-DST. Since the revised K-DST had the advantage of using 3.06 million NHSPIC's big data and the distribution of the population sample is assumed to be asymptotic and the cutoff points were set using percentile scores, the revised K-DST exhibits higher reliability and validity as compared to that of 1st edition K-DST. However, in the case of revised K-DST, the number of certain age group, not included in the NHSPIC schedule were small (27–29 months). As a result, the cutoff point for the 27–29 months group is lower than that of the other age group, which limits the interpretation of this particular section.

In conclusion, K-DST is an independently developed tool that has been developed to suit the characteristics of Korean infants and children, in addition, it is a screening test tool that checks the reliability and validity through various procedures, and is restandardized using large-scale data. In particular, the revised K-DST has a higher sensitivity and specificity compared to K-ASQ, which is considered as a good screening test tool, and shows better discrimination ability compared to the first edition of K-DST through the ROC curve analysis and criterion-related validity analysis. The K-DST shortens the test time and enhances accessibility to the examinee by allowing the test to be conducted on-line as well as in paper-and-pencil test. K-DST can be utilized as a useful developmental screening tool for NHSPIC, as well as a developmental assessment tool for developmental surveillance, screening, and posttreatment changes of normal infants. In the future, if the standardization is conducted periodically using the accumulated data from NHSPIC, it can be applied as a better tool through the application of changes based on the socio-cultural development period. As more clinical data accumulate, further research may be required to determine how the classification accuracy of K-DST varies in a more diverse group of neurodevelopmental disorders that were not included in the validation study at the time of development.

Conflicts of interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

This work was supported by Research Program funded by the Korea Centers for Disease Control and Prevention (fund code 2010-E33033-00, 2012-E33016-00, 2013-E33018-00, 2016-E33011-00).

Supplementary materials

Supplementary Tables 1-6 can be found via <https://doi.org/10.3345/cep.2020.00640>.

References

1. Eun BL, Kim SW, Kim YK, Kim JW, Moon JS, Park SK, et al. Overview of the national health screening program for infant and children. Korean J

- Pediatr 2008;51:225-32.
2. Eun BL, Chung HJ. A study on the development and validation of Korean developmental screening test for infant and children (1st year). Cheongju (Korea): Korea Centers for Disease Control and Prevention, 2013.
 3. Eun BL, Chung HJ. A study on the development and validation of Korean developmental screening test for infant and children (2nd year). Cheongju (Korea): Korea Centers for Disease Control and Prevention, 2014.
 4. Eun BL. Standardization and validity reevaluation of the Korean developmental screening test for infants & children. Cheongju (Korea): Korea Centers for Disease Control and Prevention, 2017.
 5. Hambleton RK, Swaminathan H, Rogers HJ. Fundamentals of item response theory. London: Sage Publications, 1991.
 6. Ramsay JO. Kernel smoothing approaches to nonparametric item characteristic curve estimation. *Psychometrika* 1991;56:611-30.
 7. Santos JR. Cronbach's alpha: a tool for assessing the reliability of scales. *J Ext* 1999;37:1-5.
 8. Cole DA. Utility of confirmatory factor analysis in test validation research. *J Consult Clin Psychol* 1987;55:584-94.
 9. Park HW, Cho BH. Korean Bayley scales of infant development: interpretation manual. 2nd ed. Seoul (Korea): KIDSPOP Publishing Co., 2006.
 10. Park HW, Kwak KJ, Park KB. The development of Korean version of WPPSI: the standardization study (1). *Korean J Dev Psychol* 1996;9:60-70.
 11. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. New York (NY): Lawrence Erlbaum Associates, 1988.
 12. Meisels SJ, Shonkoff JP. Handbook of early childhood intervention. Cambridge: Cambridge University Press, 1990.
 13. Mardell-Czudnowski C, Goldenberg D. DIAL-3: developmental indicators for the assessment of learning. Circle Pines (MN): American Guidance Service, 1998.
 14. Bagnato SJ, Neisworth JT, Pretti-Frontczak K. LINKing authentic assessment and early childhood intervention: best measures for best practice. Baltimore (MD): Paul H Brookes Publishing, 2010.
 15. Frankenburg WK, Dodds J, Archer P. Denver II: screening manual. Denver (CO): Denver Developmental Materials Inc., 1990.
 16. Bayley N. Bayley scales of infant development. 2nd ed. San Antonio (TX): Psychological Co., 1993.
 17. Meisels SJ, Marsden DB, Wiske MS, Henderson LW. ESI-R: early screening inventory-revised. examiner's manual. Ann Arbor (MI): Rebus Incorp., 1997.
 18. Heo KH, Squires J, Yovanoff P. Cross-cultural adaptation of a pre-school screening instrument: comparison of Korean and US populations. *J Intellect Disabil Res* 2008;52:195-206.
 19. Bricker D, Squires J. The ages and stages questionnaires (ASQ): a parent-completed, child monitoring system. 2nd ed. Baltimore (MD): Paul H Brookes Publishing Co., 1999.
 20. Squires J, Bricker D, Potter L. Revision of a parent-completed development screening tool: ages and stages questionnaires. *J Pediatr Psychol* 1997;22:313-28.
 21. Baker FB, Kim SH. Item response theory: parameter estimation techniques. New York (NY): CRC Press, 2004.
 22. Eun BL, Chung HJ, Cho S, Kim JK, Shin SM, Lee JH, et al. The appropriateness of the items of Korean ages and stages questionnaires (K-ASQ) developmental screening test in Korean infants and children. *J Korean Child Neurol Soc* 2014;22:29-41.
 23. Chung HJ, Eun BL, Kim HS, Kim JK, Shin SM, Lee JH, et al. The validity of Korean ages and stages questionnaires (K-ASQ) in Korean infants and children. *J Korean Child Neurol Soc* 2014;22:1-11.

How to cite this article: Chung HJ, Yang D, Kim GH, Kim SK, Kim SW, Kim YK, et al. Development of the Korean Developmental Screening Test for Infants and Children (K-DST). *Clin Exp Pediatr* 2020;63:438-46. <https://doi.org/10.3345/cep.2020.00640>