World Journal of Pediatrics (2025) 21:183–191
https://doi.org/10.1007/s12519-025-00878-7

ORIGINAL ARTICLE

Development and validation of the Parent-Reported Indicator of Developmental Evaluation for Chinese Children (PRIDE) tool

Sai-Shuang Wu[1] · Hao Pan[1] · Radley Christopher Sheldrick[2] · Jie Shao[3] · Xiu-Mei Liu[4] · Shuang-Shuang Zheng[3] · Sergio Miguel Pereira Soares[5] · Lan Zhang[6] · Jin Sun[7] · Ping Xu[8] · Shao-Hong Chen[8] · Tao Sun[9] · Jin-Wen Pang[10] · Ning Wu[11] · Yuan-Cong Feng[11] · Na-Ren Chen[11] · Yun-Ting Zhang[12] · Fan Jiang[13,14,15]

Received: 12 July 2024 / Revised: 29 December 2024 / Accepted: 3 January 2025 / Published online: 23 February 2025
© Children's Hospital, Zhejiang University School of Medicine 2025

Abstract

Background Developmental delay (DD) poses challenges to children's overall development, necessitating early detection and intervention. Existing screening tools in China focus mainly on children with developmental issues in two or more domains, diagnosed as global developmental delay (GDD). However, the recent rise of early childhood development (ECD) concepts has expanded the focus to include not only those with severe brain development impairments but also children who lag in specific domains due to various social-environmental factors, with the aim of promoting positive development through active intervention. To support this approach, corresponding screening tools need to be developed.

Methods The current study used a two-phase design to develop and validate the Parent-Reported Indicator of Developmental Evaluation for Chinese Children (PRIDE) tool. In Phase 1, age-specific milestone forms for PRIDE were created through a survey conducted in urban and rural primary care clinics across four economic regions in China. In Phase 2, PRIDE was validated in a community-based sample. Sensitivity and specificity of both PRIDE and Ages and Stages Questionnaires (ASQ)-3 were estimated using inverse probability weights (IPW) and multiple imputation (MI) to address planned and unplanned missing data.

Results In Phase 1 involving a total of 1160 participants aged 1 to 48 months, 63 items were selected from the initial item pool to create 10 age-specific PRIDE forms. Our Phase 2 study included 777 children within the same age range. PRIDE demonstrated an estimated sensitivity and specificity of 83.3% [95% confidence interval (CI): 56.8%–100.0%] and 84.9% (95% CI: 82.8%–86.9%) in the identification of DD.

Conclusion The findings suggest that PRIDE holds promise as a sensitive tool for detecting DD in community settings.

Keywords Community sample · Developmental delay · Development and validation · Early childhood development · Screening tool

Introduction

With the rise of the Early Childhood Development (ECD) concept and corresponding global initiatives, the development of assessment tools for ECD has become a focus of interest [1]. Besides creating population-level metrics for assessing development from birth to age three [2], individual-level screening tools are widely used in the ECD field. These clinical tools focus on early detection of developmental delay (DD), which is crucial for assessing language, cognitive, motor, and socioemotional development to foster improved educational and well-being outcomes for children [3].

Significant DD is defined as a performance of two or more standard deviations (SD) below the mean on age-appropriate standardized norm-referenced tests [4]. Delays can manifest in a single domain termed isolated developmental delay (IDD), or across multiple domains, known as global developmental delay (GDD) [5]. In clinical settings, GDD is a commonly used diagnosis indicating an underlying developmental disorder or intellectual disability, requiring proactive intervention. However, around 20% of children may not meet GDD criteria but still experience specific learning difficulties [6], as some of these children initially exhibit IDD. With the promotion of the ECD concept, which

Extended author information available on the last page of the article

emphasizes nurturing practices to fully develop children's potential, it has become crucial to develop tools that can screen for both GDD and IDD early and allow for positive intervention through nurturing and caregiving principles [7]. Moreover, existing tools are primarily validated for clinically high-risk populations. From the ECD perspective, it is vital to conduct research based on community samples, especially for children with IDD.

Currently, the most widely used DD screening tools in China are the Denver Developmental Screening Test (DDST) [8, 9] and Ages and Stages Questionnaires (ASQ-3) [10, 11], both adapted from their English versions and validated accordingly. However, development patterns exhibit linguistic and cultural differences [12]. The DDST and ASQ-3 were translated and revised with limited analysis and enhancement of cultural adaptability, particularly in the language domain. For example, mastering subject-verb agreement is a critical milestone in English (e.g., adding "-ed" for past tense verbs), but this does not apply to Chinese due to different syntactic rules. Conversely, Chinese has unique language features that serve as developmental milestones [13]. The first study on Chinese language milestones for children aged 0–5.5 years laid the foundation for developing screening tools based on Chinese language and culture [14]. As mentioned earlier, existing developmental screening tools primarily target clinical GDD populations with no validity and reliability studies conducted on community-based populations.

The current study aimed to: (1) develop and validate a tool suitable for ECD assessment in community settings. This is referred to as the Parent-Reported Indicator of Developmental Evaluation for Chinese Children (PRIDE); and (2) conduct a comparative study between this tool and the widely used ASQ-3, which has demonstrated good validity and reliability in both English and Chinese versions.

Methods

Phase 1: creation of PRIDE age-specific milestone forms

Study design and participants

We conducted a survey in urban and rural primary care clinics representing four different economic regions in China, including 10 sites. Parents of children aged 1 to 48 months were sampled with a stratified strategy based on urban/rural residence, gender, and age group (see Supplementary Methods for further details). Item Response Theory (IRT) models were employed for parameter estimation of survey items. Following previous studies [2, 15, 16], we planned to collect 100 participants for each age group, evenly distributed between males and females. Participants were parents who met the following criteria: (1) primary caregivers of children aged 1 to 48 months receiving preventive vaccinations or regular check-ups at community hospitals, and (2) primary caregivers with basic Chinese reading skills who could complete the questionnaire following standardized instructions. Data collection took place during preventive vaccination/check-up clinics at community hospitals. Doctors asked parents if they were willing to participate in the study. Parents who agreed scanned a QR code to fill out the questionnaire. Data collection for this phase was conducted from October to November 2022.

Measures

The development of PRIDE followed the development process of the Survey of Well-being of Young Children (SWYC) [17], a comprehensive screening tool designed to identify early signs of developmental delays and promote early interventions. Based on existing developmental theories and patterns of children's developmental milestones [18, 19], aligning with the time points of routine well-child visits in China, we created 10 age-specific groups: 1 − < 4 months (2-month form), 4 − < 6 months (4-month form), 6 − < 9 months (6-month form), 9 − < 12 months (9-month form), 12 − < 15 months (12-month form), 15 − < 18 months (15-month form), 18 − < 23 months (18-month form), 23 − < 29 months (24-month form), 29 − < 35 months (30-month form), and 35 − < 47 months (36-month form). After the creation of the item pool, expert panel review and parent cognitive interview, we obtained an initial item pool of 93 developmental milestones (Supplementary Methods). Answers for each item included "not yet", "somewhat", "very much", and "don't know/refuse to answer." An adaptive questionnaire was designed to enhance the alignment between data and models while minimizing the number of items parents need to complete. Adaptive setting logic involves establishing a stop point when parents consecutively answer "not yet" for 10 items, and a start point when they answer "very much" for 10 items. Parents filled out the questionnaires using an online platform (Fig. 3a). Doctors at all sites underwent standardized online training and were instructed not to provide additional explanations for questionnaire items. For example, if parents had questions about the questionnaire's content, doctors were instructed to answer, "Please answer according to your own understanding."

Statistical analysis

First, we cleaned the dataset (Fig. 1a and Supplementary Methods). Item Response Theory (IRT) model [20] was used to estimate difficulty parameters ($\alpha 1$, $\alpha 2$) and discrimination parameters (β) for each item. Additionally, we estimated the developmental age at which there was a 25%, 50%, and 75% probability of passing the item (i.e., choosing "very much"). Item characteristic curves (ICCs) were drawn to describe the probability of each response to each milestone as a function of developmental age. Indicators including item response rates, model fitness, and differential item functioning (DIF), were employed to determine the suitability of each item as a parent-reported monitoring milestone (Supplementary Methods for further explanation). Next, we calculated the item information to determine how much each item contributed to estimating children's developmental age within a specific age band. The top ten items with the highest item information were selected for each age group. These items together formed the questionnaire for that age group sorted based on the estimated 50% passing age for each item. Ultimately, 10 age-specific forms were created each consisting of 10 items. After obtaining parameters for each item and the child's responses to the 10 questions, the IRT model can estimate the child's developmental age. If the developmental age lags behind the chronological age by 15%, then the child is identified as at risk of DD.

Phase 2: validation of PRIDE and comparative accuracy of PRIDE with ASQ-3

Study design and participants

In total, five well-baby check clinics in the Zhejiang and Fujian Provinces were selected as data collection points. Stratified data collection for children aged 1–48 months were conducted based on age groups and gender. Each consenting participant scanned a QR code and filled out both the PRIDE questionnaire and ASQ-3 questionnaire at the clinic. Based on the results of the screening tests, if either of them

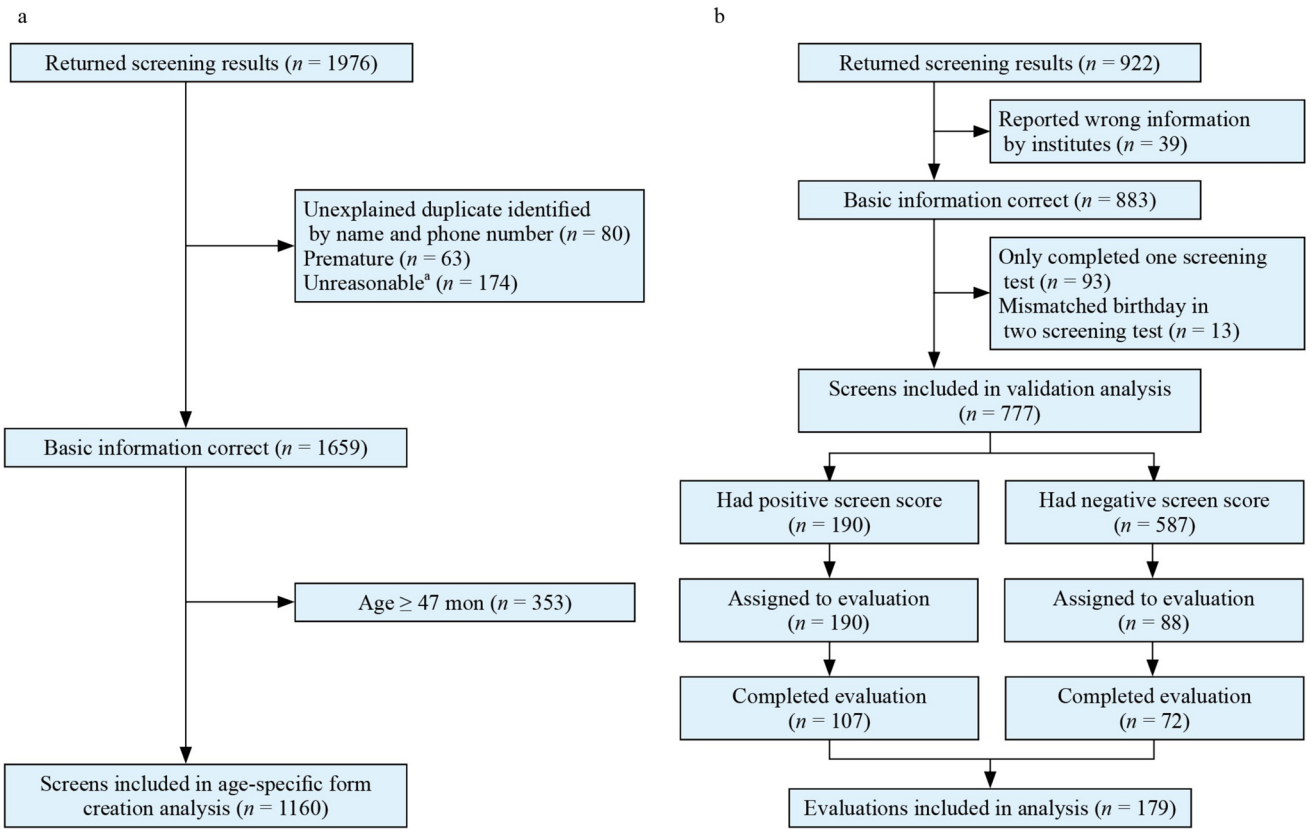

Fig. 1 Enrollment and evaluation flow charts. **a** Flow chart for age-specific form creation stage analysis. [a] Unreasonable data was identified by (1) over 99% of entries filled out with the same answers; (2) data visualization and logical analysis, with the criteria being: any answer of "very much" to any of the last 5 questions for children under 36 months. **b** Flow chart for validation analysis. Screening tests included age-appropriate PRIDE and ASQ-3 questionnaires that assess the risk for developmental delay. Standardized developmental tests (the Griffiths Mental Development Scale-Chinese, GDS-C) were administered during the evaluation. *PRIDE* Parent-Reported Indicator of Developmental Evaluation for Chinese Children, *ASQ-3* Ages and Stages Questionnaires

indicated a positive result (indicating a risk of either IDD or GDD), the child was recommended for further diagnostic assessment. Based on previous studies and accounting for refusal rates [21], we involved the initial random selection of 15% of children for diagnostic assessment when negative results were obtained from both questionnaires to achieve an evaluation percentage surpassing 10% (Fig. 1b). The plan was to collect a total of 60 samples for each age group, with an equal distribution of boys and girls. Additionally, convenience samples were selected for reliability testing, including 60 pairs of parents completing the PRIDE questionnaire simultaneously for inter-rater reliability. In addition, 60 primary caregivers completed the questionnaire again after a 7-day interval for test–retest reliability. The inclusion criteria for participants were the same as in phase 1. Data collection for this phase was completed between May 2023 and July 2023.

Measures

The DD screening questionnaire includes PRIDE and ASQ-3. The developmental diagnostic assessment utilized the Griffiths Development Scale-Chinese (GDS-C) [22], evaluated by qualified pediatricians. The assessment covers five domains: locomotor, personal-social, language, eye-hand coordination, and performance. Since the practical reasoning domain is not assessed for children aged 0–2, the results for this domain are not included in the diagnostic criteria calculation for this study. The interval between screening assessment and diagnostic assessment did not exceed seven days. The diagnostic criteria for DD are defined as any one of the five dimensions of GDS-C scoring $\leq$ 2SD, which includes IDD and GDD. For GDD, the diagnostic criteria are defined as any two or more dimensions of GDS-C scoring $\leq$ 2SD. Additionally, we report diagnostic criteria based on the general quotient (GQ) in the supplementary section. Scores are categorized as typical ($\geq$ 90), mild (80–89), moderate (70–79), or severe ($<$ 70) delays. Other information was collected which included demographic characteristics, four parenting risks and three severe family risks (Supplementary Methods).

Statistical analysis

First, in line with published recommendations [21], inverse probability weights (IPW) were incorporated to address planned missing data caused by the sampling strategy (evaluating children with a positive score on any screening results and randomly selecting children with negative scores) and multiple imputation (MI) with chained equations was employed to handle unplanned missing data (e.g., declining to attend the evaluation). The imputation models included screening results, parents' concerns, and mother's education level to estimate evaluation outcomes. Twenty MI datasets were created on the survey-weighted dataset. Subsequently, generalized estimating equations with logit links were utilized to estimate true and false positive fractions and their 95% confidence intervals. Cronbach's α was calculated to estimate the internal consistency and reliability of PRIDE and the Kappa coefficient was calculated to estimate the test–retest reliability and inter-rater reliability of PRIDE. For sensitivity analyses, we first analyzed the sensitivity and specificity of PRIDE and ASQ-3 based on the actual data. Then, we conducted assumption-based simulations on the unassessed samples according to published literature [23]. Two assumptions were considered: (1) all unassessed children were either true positive or true negative, and (2) the prevalence of DD in unassessed children was the same as in assessed children. Prevalence of DD, sensitivity, and specificity were calculated for both PRIDE and ASQ-3 under the two different assumptions, respectively. Additionally, the diagnostic odds ratio (positive likelihood ratio divided by negative likelihood ratio) was computed as a single indicator of test accuracy.

Analyses were performed using Mplus 8.3 [24] for IRT model construction and item parameter estimation, R 4.1 [25] for item response rates, model fit, and DIF indicator calculations, and Stata/MP 17.0 [26] for all other data analyses.

Results

Phase 1 study included 1160 samples [576 females (49.7%) and 584 males (50.3%)] (Fig. 1a), while the phase 2 study comprised 777 screening samples (337 females, 43.4% and 440 males, 56.6%) and 179 assessment samples (Fig. 1b). In phase 1 initial sample, we identified 174 cases of unreasonable data based on the following criteria: (1) over 99% of entries were filled out with the same answers, indicating potential response bias; and (2) data visualization and logical analysis revealed inconsistencies, specifically any instance of "very much" responses for any of the last five questions concerning children under 36 months. Descriptions of the characteristics of all samples are provided in Table 1. The mean age of the children was 17.4 months (SD = 13.3) in phase 1 and 16.3 years (SD = 10.7) in phase 2. The samples were diverse with respect to socioeconomic status, with 210 (18.1%) participants having completed junior high school education or below; 508 (43.8%) had received high school and junior college education; 438 (37.8%) had achieved undergraduate education or above, and 4 (0.3%) either did not disclose or refused to answer. In phase 2, 59 participants (7.6%) had completed junior high school education or below; 270 (34.8%) had received high school and junior college education; 444 (57.1%) had achieved undergraduate or above education, and 4 (0.5%) either did not disclose or refused to answer.

A total of 63 items were selected to form 10 age-specific forms. ICC of each item is shown in Fig. 2. Supplementary Table 1 presents detailed information on these items, including ICC parameters, ages of attainment, and assignment to forms. A flow chart illustrating the item selection process is shown in Supplementary Fig. 1. After an expert panel review and parent cognitive interview an initial item pool of 93 developmental milestones was generated. Among these items, 26 were excluded because of low response rate (< 99%), evidence of DIF, or evidence of poor item fit; 4 were excluded because of lack of item information. Detailed information on the 30 items excluded during the selection process, estimated 50% passing age of each item and the reasons for exclusion is shown in Supplementary Table 2.

Table 2 shows that when adjusted weights using IPW and MI with the DD diagnostic criterion, the estimated prevalence was 6.7% [95% confidence interval (CI): 3.9% – 9.5%]. PRIDE demonstrated a specificity of 84.9% (95% CI: 82.8% – 86.9%) and a sensitivity of 83.3% (95% CI: 56.8% – 100.0%), whereas ASQ-3 demonstrates 91.8% specificity (95% CI: 90.0%–93.6%) and 54.9% sensitivity (95% CI: 32.6% – 77.3%). With the GDD diagnostic criterion, the estimated prevalence was 3.4% (95% CI: 0.9% – 5.9%). PRIDE demonstrated 81.9% specificity (95% CI: 80.1% – 83.6%) and 64.1% sensitivity (95% CI: 21.0% – 100.0%), while ASQ-3 exhibits a specificity of 90.4% (95% CI: 88.6% – 92.3%) and sensitivity of 60.5% (95% CI: 20.0% – 100.0%).

Table 1 Description of demographic characteristics

Variables	Phase 1 ($n=1160$)	Phase 2 ($n=777$)
Age, mon	17.4 (13.3)	16.3 (10.7)
Age groups (mon)		
1.0–<4.0	132 (11.4)	66 (8.5)
4.0–<6.0	116 (10.0)	65 (8.4)
6.0–<9.0	153 (13.2)	124 (16.0)
9.0–<12.0	123 (10.6)	60 (7.7)
12.0–<15.0	102 (8.8)	84 (10.8)
15.0–<18.0	93 (8.0)	62 (8.0)
18.0–<23.0	95 (8.2)	98 (12.6)
23.0–<29.0	87 (7.5)	77 (9.9)
29.0–<35.0	57 (4.9)	74 (9.5)
35.0–<47.0	202 (17.4)	67 (8.6)
Sex		
Female	576 (49.7)	337 (43.4)
Male	584 (50.3)	440 (56.6)
Mother's education		
Junior high school or below	210 (18.1)	59 (7.6)
High school and junior college	508 (43.8)	270 (34.8)
Undergraduate or above	438 (37.8)	444 (57.1)
Unknown/refused to answer	4 (0.3)	4 (0.5)

Data are mean (standard deviation) or *n* (%)

Frequencies and unadjusted sensitivity and specificity analysis results are shown in Supplementary Table 3 for both PRIDE and ASQ-3. For sensitivity analysis, Supplementary Table 4 and 5 show simulation results with different assumptions for PRIDE and ASQ-3. Under assumption B (i.e., the prevalence of DD is equal in non-evaluated and evaluated children), PRIDE demonstrates a sensitivity of 84.2% and a specificity of 84.9% when testing for DD. For GDD, the sensitivity of PRIDE is 68.0%, and the specificity is 81.9%. The characteristics of unassessed and assessed children are described and compared (Supplementary Table 6 for PRIDE and Supplementary Table 7 for ASQ-3). Results indicate similar characteristics including mother's education level, parent concern, parenting risks and severe family risks between these two populations, supporting assumption B.

When applying diagnostic criteria based on GQ, the diagnosis encompassed mild-to-severe (any) delays (GQ < 90), moderate-to-severe delays (GQ < 80), and severe delays (GQ < 70). Frequencies and unadjusted results are presented in Supplementary Table 8, while adjusted estimates with IPW and MI are provided in Supplementary Table 9. Simulation results under different assumptions are displayed in Supplementary Table 10 and 11, respectively.

Cronbach's α for PRIDE was 0.84, indicating very good internal consistency and reliability. For inter-rater reliability caregivers showed agreement in 78.13% of cases, exceeding the expected random agreement of 63.92%. The calculated Kappa value was 0.39, indicating fair agreement beyond what would be expected by chance ($Z=3.20$, $\text{Prob} > Z = 0.0007$). Similarly, for test–retest reliability, caregivers exhibited agreement in 81.82% of cases, exceeding the expected random agreement of 67.04%. The Kappa value for test–retest reliability was 0.44, which signifies moderate

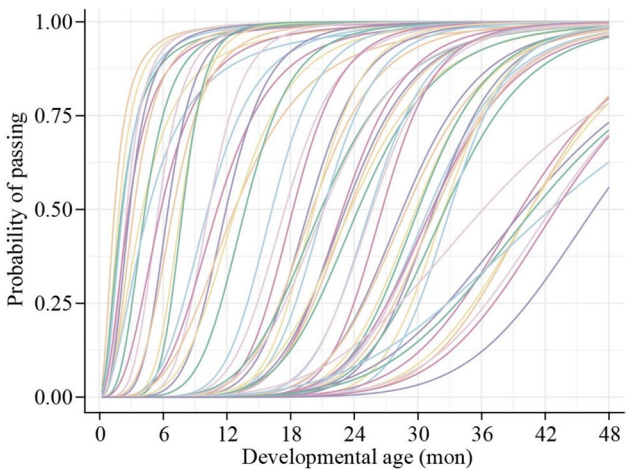

Fig. 2 Item characteristic curves for all 63 items

Table 2 Adjusted estimates of sensitivity and specificity of PRIDE and ASQ-3 with sampling weights[a]

Tools	Specificity	Sensitivity	Prevalence
Developmental delay (any one domain ≤ 2 SD)			
PRIDE milestones	84.9 (82.8–86.9)	83.3 (56.8–100.0)	6.7 (3.9–9.5)
ASQ-3	91.8 (90.0–93.6)	54.9 (32.6–77.3)	
Global developmental delay (any two domains ≤ 2 SD) [b]			
PRIDE milestones	81.9 (80.1–83.6)	64.1 (21.0–100.0)	3.4 (0.9–5.9)
ASQ-3	90.4 (88.6–92.3)	60.5 (20.0–100.0)	

[a] All values presented as percentage (95% confidence interval). Estimates were conducted using inverse probability weights and multiple imputation

[b] Reference test was the Griffiths Development Scales-Chinese (GDS-C)

PRIDE Parent-Reported Indicator of Developmental Evaluation for Chinese Children, *ASQ-3* Ages and Stages Questionnaires, *SD* standard deviation

agreement beyond chance ($Z = 4.0$, $\text{Prob} > Z < 0.0001$). These results reject the hypothesis that determinations are made randomly, further supporting the reliability of PRIDE.

Figure 3a illustrates the process of completing a PRIDE test and the interface of the electronic questionnaire system. When considering future application in clinical settings, particularly during routine well-child visits, we estimated the potential screen and referral burden (Fig. 3b). Based on the screen positive rate, referral rate, and diagnostic rate observed during the validation phase of the current study, for every 1000 children screened, about 244 children will test positive by PRIDE. Among them, 145 will be referred for diagnostic evaluation, ultimately resulting in 41 diagnosed cases of DD.

Discussion

This study describes the development and reliability-validity testing of PRIDE, a 10-item parent-reported Chinese DD screening tool. At the community level, results showed that PRIDE has comparable sensitivity and specificity to ASQ-3 for GDD screening with PRIDE demonstrating higher sensitivity but lower specificity than ASQ. For DD, PRIDE exhibited a sensitivity and specificity of over 80%. The specificity is lower and sensitivity is higher compared to ASQ.

When considering GDD as a primary outcome, results showed that both PRIDE and ASQ-3 exhibit low sensitivity, around 60%. However, when validated in a population with a high risk of GDD, the ASQ-3 demonstrated both sensitivity and specificity above 80% [11]. This finding indicates how sample characteristics influence the results when validating a DD screening tool. There has been a notable increase in the scrutiny of screening tool accuracy within application settings recently [27]. Indeed, the sensitivity of existing screening tools in applications is often lower than initially observed during their validation phase. A meta-analysis revealed that the positive predictive value (PPV) of a screening tool in populations with DD risks was twice as high compared to non-risk populations [27]. Nevertheless, in practice, many tools undergo validation with populations at risk of GDD, such as premature children [11]; this may yield different results when applied to a community sample. Additionally, a considerable number of studies overlook the issue of loss to follow-up in both study design and data analysis. Specifically, neglecting the results of all individuals who undergo screening, particularly those with negative screening results, introduces substantial bias [23]. The current study was conducted using a community-based sample and accounted for the outcomes of unassessed children by adjusting sampling weights and conducting imputation. This explains why both PRIDE and ASQ-3 exhibited low sensitivity for GDD.

By distinguishing DD using the definitions of IDD and GDD, we also found the importance of clarity in defining the target conditions for DD screening tools. As DD screening typically occurs before three years of age, it is challenging to use primary developmental screening tools to identify specific developmental disorders. Our results indicate that PRIDE demonstrates good accuracy in screening for DD, with comparable results to ASQ-3 for GDD and particularly good performance for DD. Unlike GDD, which often suggests intellectual disability, IDD encompasses an elevated risk associated with delays in specific domains, potentially indicative of developmental language disorders, attention deficit hyperactivity disorder, autistic spectrum disorder, and other conditions [5]. By employing DD as the primary target during validation, PRIDE aims to identify children who would benefit most from early intervention.

During the development phase, we ensured that PRIDE possessed both practicality and cultural adaptability. First, through extensive supplementation of language items and expert reviews, we enhanced PRIDE's cultural adaptability for use among Chinese children [14]. Second, we enhanced the reliability of parental reporting by conducting cognitive interviews with parents and employing objective indicators such as response rates to select items. Finally, based on the characteristics of developmental milestones, we utilized an IRT model to select items using indicators such as DIF and item fitness.

During the validation phase, we intended to refer all children with positive screening results and 15% of those with negative screening results for diagnostic evaluation. Consequently, by adjusting the sampling weights, the sample closely mirrors the distribution of the general population [27]. Furthermore, we accounted for the outcomes of

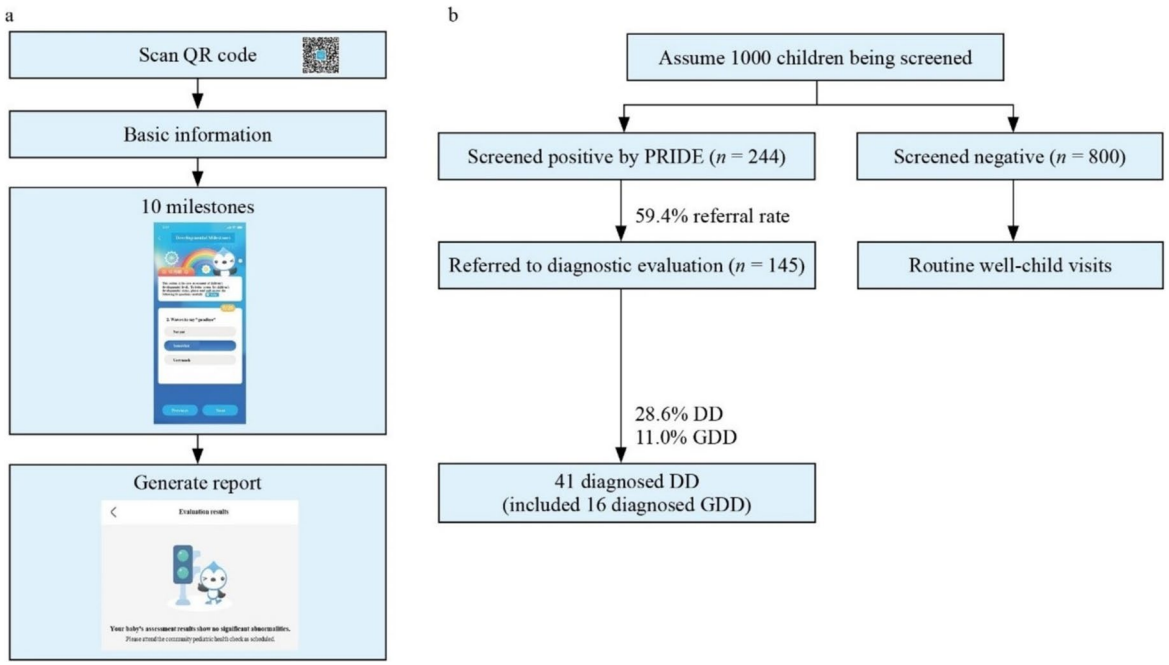

Fig. 3 Clinical applications of PRIDE: flowchart and simulated example. **a** Flow chart of the procedure of completing a PRIDE test. The test is originally in Chinese but presented in English here. **b** Estimation of screening and referral process when screening 1000 children with pride during routine well-child visits. *PRIDE* Parent-Reported Indicator of Developmental Evaluation for Chinese Children, *DD* developmental delay, *GDD* global developmental delay

children who were lost to follow-up and had not undergone diagnostic assessments by utilizing MI. Through these two methods, we addressed both planned and unplanned missing data. We have also conducted several sensitivity analyses, including employing different estimation methods (e.g., simulation with assumptions) and various diagnostic criteria (e.g., GQ). Previous studies have demonstrated similar outcomes affirming that adjustments [21] and simulations [23] are reliable methods for use in practical settings.

There are also some limitations. First, the loss to follow-up rates for both negative and positive screening children were higher than expected, resulting in a smaller number of diagnosed cases and leading to a wide range of 95% CI when estimating sensitivity. Given our study design to include an entire target population, this challenge was difficult to avoid, as the low number of DD cases is constrained by prevalence and sample size. We attempted to address this limitation by conducting hypothesis-based simulations. Further investigations into the reasons for referral failure could be valuable when establishing a screening referral system. Second, this study did not report on the predictive validity of PRIDE. Accumulating longitudinal data in subsequent research and reporting on whether PRIDE accurately predicts the diagnosis of subsequent developmental disorders would provide a more comprehensive assessment of the tool. Third, while we have currently validated the tool in the general population, it is important to explore its applicability in other specific populations. For instance, follow-up studies on neonates and children who experience adverse neurodevelopmental outcomes due to early brain injury would be valuable [28]. Finally, while we provide an estimation of the screening and referral burden when applying PRIDE in routine child well-visits, current information is still insufficient to fully understand the cost-effectiveness of implementing PRIDE for population-wide screening. Referral decisions are often made based on a range of circumstances, with actual referral rates ranging from 10% to 86% [29]. Variations in clinical resources between regions (such as diagnostic and rehabilitation resources and human resources for basic counseling in the community.), as well as parental referral preferences, can influence referral decisions [30]. Therefore, the specific referral criteria may need to be situation-specific to achieve optimal effectiveness of the PRIDE application. Further studies including cost-effectiveness and implementation studies are needed to explore the application of PRIDE in the community population. Since PRIDE can effectively identify GDD and IDD, it can be applied to the ECD intervention projects that the Chinese government is currently promoting at the community level, such as group-based parenting interventions [31], to provide early intervention for these at-risk groups.

In conclusion, the PRIDE tool, developed based on ECD milestone studies conducted in multiple regions of China, is simple and easy to use. Its reliability and validity are comparable with ASQ-3 for GDD screening and it enables IDD screening. In the future, PRIDE can serve as a tool for ECD

in community settings in China. Given the similarities in situations among low- and middle-income countries (LMICs), the development process for this tool holds potential wide applicability.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12519-025-00878-7.

Acknowledgements The authors would like to thank all the workers at the data collection sites for their hard work in this study: Zhejiang Province (Cixi Third People's Hospital Medical Health Group, Cixi Zhouxiang Central Kindergarten, Community Health Service Center of Xiaohe Lake Villa Street, Huamei Weihan Kindergarten, Affiliated Children's Hospital of Zhejiang University School of Medicine), Jiangxi Province (Yudu Second People's Hospital, Yudu Yinkeng Town Central Kindergarten, Yudu Maternal and Child Health Hospital, Yudu Science and Technology Kindergarten), Hainan Province (Yacheng Health Center, Yacheng Central Kindergarten, Sanya Women and Children's Hospital, Sanya First Kindergarten), Sichuan Province (Furong Community Health Service Center, Dongjie Kindergarten, Chengdu Huayang Community Health Service Center, First Kindergarten of Tianfu New Area), Liaoning Province (Tiexi Street Defense Guard Station, Aierfang Tiandu City Kindergarten, Qixianling Community Health Service Center, Wanda Huafu Kindergarten), Fujian Province (Fujian Children's Hospital). Special thanks to Professor Li Sheng, Professor Yi-Wen Zhang, and Professor Jing Zhou for their careful review of the entries and research design and valuable suggestions.

Author contributions WSS and PH contributed equally to this paper. WSS: formal analysis, data curation, writing–original draft. PH, SRC: formal analysis. SJ, LXM, ZSS, ZL, SJ, XP, CSH, ST, WN, FYC, and CNR: investigation, data curation. PSM: writing–review and editing. PJW: data curation, software. ZYT and JF: conceptualization, supervision, writing–review and editing. All authors have read and approved the final manuscript.

Funding National Natural Science Foundation of China (823B2087), Shanghai Municipal Health Commission (GWVI-11.2-XD16), National Natural Science Foundation of China Joint Fund Project (U23A20170), Shanghai Municipal Health Commission (Child Neurological and Psychological Development Screening and Intervention Project), Shanghai Key Laboratory of Child Brain and Development (24dz2260100).

Data availability Data were examined after importing from the online survey system and are held and managed by the research group at Shanghai Children's Medical Center with the format of STATA file (.dta). There is an application process for using the data. After the application is approved by the Publication Committee, deidentified data can be shared with collaborators for research purposes.

Declarations

Conflict of interest No financial or non-financial benefits have been received or will be received from any party related directly or indirectly to the subject of this article. Author Fan Jiang is a member of associate editors of *World Journal of Pediatrics*. The paper was handled by other editors and has undergone a rigorous peer review process. Author JF was not involved in the journal's review of or decisions related to this manuscript.

Ethical approval This study received approval from the Ethics Committee of Shanghai Children's Medical Center, Shanghai Jiao Tong University School of Medicine (SCMCIRB-K2021080-1). All parents provided electronic informed consent and voluntarily participated in the study.

References

1. Black MM, Walker SP, Fernald LCH, Andersen CT, DiGirolamo AM, Lu C, et al. Early childhood development coming of age: science through the life course. Lancet. 2017;389:77–90.
2. Cavallera V, Lancaster G, Gladstone M, Black MM, McCray G, Nizar A, et al. Protocol for validation of the Global Scales for Early Development (GSED) for children under 3 years of age in seven countries. BMJ Open. 2023;13:e062562.
3. Scherzer AL, Chhagan M, Kauchali S, Susser E. Global perspective on early diagnosis and intervention for children with developmental delays and disabilities. Dev Med Child Neurol. 2012;54:1079–84.
4. Bellman M, Byrne O, Sege R. Developmental assessment of children. BMJ. 2013;346:e8687.
5. Choo YY, Agarwal P, How CH, Yeleswarapu SP. Developmental delay: identification and management at primary care level. Singap Med J. 2019;60:119–23.
6. Horowitz S, Rawe J, Whittaker M. The state of learning disabilities: understanding the 1 in 5. New York: National Center for Learning Disabilities. 2017;3:211–9. https://assets.ctfassets.net/p0qf7j048i0q/2Q2TsAzUSM9TXq22LPxQwY/658d54ed0529acddf70f89d367db7915/2017_State_of_LD_-_Executive_Summary_Final_Accessible.pdf. Accessed Dec 28, 2024
7. Britto PR, Lye SJ, Proulx K, Yousafzai AK, Matthews SG, Vaivada T, et al. Nurturing care: promoting early childhood development. The Lancet. 2017;389:91–102.
8. Frankenburg WK, Dodds J, Archer P, Shapiro H, Bresnick B. The Denver II: A major revision and restandardization of the denver developmental screening test. Pediatrics. 1992;89:91–7.
9. Chen J, Wei M, Chen S, He L, Jiang L, Shi J, et al. Standardization of Denver II Developmental Screening Test in Shanghai and its relative factors. Zhongguo Ertong Baojian Zazhi. 2007;5:76-77.
10. Veldhuizen S, Clinton J, Rodriguez C, Wade TJ, Cairney J. Concurrent validity of the ages and stages questionnaires and bayley developmental scales in a general population sample. Acad Pediatr. 2015;15:231–7.
11. Wei M, Bian X, Squires J, Yao G, Wang X, Xie H, et al. Studies of the norm and psychometrical properties of the ages and stages questionnaires, third edition, with a Chinese national sample. Zhonghua Er Ke Za Zhi. 2015;53:913–8.
12. Leonard LB. Specific language impairment across languages. Child Development Perspectives. 2014;8:1–5.
13. Wang D, Zheng L, Lin Y, Zhang Y, Sheng L. Sentence repetition as a clinical marker for mandarin-speaking preschoolers with developmental language disorder. J Speech Lang Hear Res. 2022;65:1543–60.
14. Saishuang W, Jin Z, Haiwa W, Hao P, Yunting Z, Fan J. Language development milestones in Chinese children aged 0 to 5.5 years in China: A cross-sectional survey. Chinese J Evidence-Based Pediatr. 2023;18:325.
15. Dai S, Vo TT, Kehinde OJ, He H, Xue Y, Demir C, et al. Performance of Polytomous IRT Models With Rating Scale Data: An Investigation Over Sample Size, Instrument Length, and Missing Data. Front Education. 2021;6:721963.
16. Jiang S, Wang C, Weiss DJ. Sample Size Requirements for Estimation of Item Parameters in the Multidimensional Graded Response Model. Front Psychol. 2016;7:109.
17. Sheldrick RC, Perrin EC. Evidence-Based Milestones for Surveillance of Cognitive, Language, and Motor Development. Acad Pediatr. 2013;13:577–86.
18. Zubler JM, Wiggins LD, Macias MM, Whitaker TM, Shaw JS, Squires JK, et al. Evidence-Informed Milestones for Developmental Surveillance Tools. Pediatrics. 2022;149:e2021052138.

19. Wallace IF, Berkman ND, Watson LR, Coyne-Beasley T, Wood CT, Cullen K, et al. Screening for speech and language delay in children 5 years old and younger: A systematic review. Pediatrics. 2015;136:e448–62.
20. Drachler MDL, Marshall T, De Carvalho Leite JC. A continuous-scale measure of child development for population-based epidemiological surveys: a preliminary study using Item Response Theory for the Denver Test. Paediatr Perinat Ep. 2007;21:138–53.
21. Sheldrick RC, Marakovitz S, Garfinkel D, Carter AS, Perrin EC. Comparative accuracy of developmental screening questionnaires. JAMA Pediatr. 2020;174:366–74.
22. Tso W, Wong V, Xia X, Faragher B, Li M, Xu X, et al. The Griffiths Development Scales-Chinese (GDS-C): A cross-cultural comparison of developmental trajectories between Chinese and British children. Child: Care, Health Development. 2018;44:378–83.
23. Sheldrick RC, Hooker JL, Carter AS, Feinberg E, Croen LA, Kuhn J, et al. The influence of loss to follow-up in autism screening research: Taking stock and moving forward. J Child Psychol Psychiatry. 2024;65:656–67.
24. Muthén LK, Muthén BO. Mplus. Los Angeles, CA; 2019.
25. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2021.
26. StataCorp. Stata/MP. College Station, TX: StataCorp LLC; 2021.
27. Rah SS, Jung M, Lee K, Kang H, Jang S, Park J, et al. Systematic review and meta-analysis: real-world accuracy of children's developmental screening tests. J Am Acad Child Adolesc Psychiatry. 2023;62:1095–109.
28. Yan K, Tang LK, Xiao FF, Zhang P, Cheng GQ, Wang LS, et al. Brain development in newborns and infants after ECMO. World J Pediatr. 2024;20:556–68.
29. Sheldrick RC, Breuer DJ, Hassan R, Chan K, Polk DE, Benneyan J. A system dynamics model of clinical decision thresholds for the detection of developmental-behavioral disorders. Implement Sci. 2016;11:156.
30. Sheldrick RC, Garfinkel D. Is a positive developmental-behavioral screening score sufficient to justify referral? A review of evidence and theory. Acad Pediatr. 2017;17:464–70.
31. Xu M, Zhang H, Liu A, Zhao C, Huang X, Berman S, et al. Effectiveness and cost-effectiveness of a group-based intervention to improve social-emotional development of young children in poverty-stricken areas: A cluster randomized controlled trial. J Glob Health. 2023;13:04017.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Sai-Shuang Wu[1] · Hao Pan[1] · Radley Christopher Sheldrick[2] · Jie Shao[3] · Xiu-Mei Liu[4] · Shuang-Shuang Zheng[3] · Sergio Miguel Pereira Soares[5] · Lan Zhang[6] · Jin Sun[7] · Ping Xu[8] · Shao-Hong Chen[8] · Tao Sun[9] · Jin-Wen Pang[10] · Ning Wu[11] · Yuan-Cong Feng[11] · Na-Ren Chen[11] · Yun-Ting Zhang[12] · Fan Jiang[13,14,15]

✉ Yun-Ting Zhang
edwinazhang@hotmail.com

✉ Fan Jiang
fanjiang@shsmu.edu.cn

[1] Department of Developmental and Behavioral Pediatrics, National Children's Medical Center, Shanghai Children's Medical Center, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

[2] Radley Christopher Sheldrick, University of Massachusetts, Shrewsbury, MA, USA

[3] Department of Child Health Care, Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, China

[4] Developmental and Behavior Pediatrics Department, Fujian Branch of Shanghai Children's Medical Center), Fujian Children's Hospital, Fuzhou, China

[5] Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

[6] School of Medicine, Chengdu Women's and Children's Central Hospital, University of Electronic Science and Technology of China, Chengdu, China

[7] Dalian Women and Children's Medical Group, Dalian, China

[8] Jiangxi Provincial Maternal and Child Health Hospital, Nanchang, China

[9] Shanghai Children's Medical Center, Hainan Branch, Sanya, China

[10] School of Medicine, Shanghai Jiao Tong University, Shanghai, China

[11] Shenzhen Xinzhi Institute of Psychological Assessment Technology, Shenzhen, China

[12] Child Health Advocacy Institute, National Children's Medical Center, Shanghai Children's Medical Center, School of Medicine, Shanghai Jiao Tong University, 1678 Dongfang Rd, Shanghai 200127, China

[13] Department of Developmental and Behavioral Pediatrics, Pediatric Translational Medicine Institute, Shanghai Key Laboratory of Child Brain and Development, National Children's Medical Center, Shanghai Children's Medical Center, School of Medicine, Shanghai Jiao Tong University, 1678 Dongfang Rd, Shanghai 200127, China

[14] Ministry of Education Shanghai Key Laboratory of Children's Environmental Health, Xinhua Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

[15] Shanghai Center for Brain Science and Brain-Inspired Technology, Shanghai, China