




RESEARCH ARTICLE

REVISED Assessment of a neuro-developmental screening tool in children in Bhutan [version 2; peer review: 3 approved, 1 approved with reservations]

Brian Wong¹, Sara Grundy², Lhab Tshering³, Kinley Tshering⁴, Farrah J. Mateen ^{2,5}

¹Department of Pediatric Neurology, Children's Hospital of Los Angeles, Los Angeles, CA, USA

²Department of Neurology, Massachusetts General Hospital, Boston, MA, 02114, USA

³Department of Psychiatry, Jigme Dorji Wangchuck National Referral Hospital, Thimphu, Bhutan

⁴Department of Pediatrics, Jigme Dorji Wangchuck National Referral Hospital, Thimphu, Bhutan

⁵Harvard Medical School, Boston, USA

V2 First published: 20 Jun 2019, 3:1496
<https://doi.org/10.12688/gatesopenres.13037.1>
 Latest published: 11 Sep 2019, 3:1496
<https://doi.org/10.12688/gatesopenres.13037.2>

Abstract







Background: Developmental screening tools are designed to fit the cultural context in which they are utilized, yet often find a wider international audience. This study evaluates the efficacy of one such tool, the Parental Evaluation of Developmental Status: Developmental Milestones (PEDS:DM), developed in the United States and tested in the lower income Asian country of Bhutan. We aimed to test the PEDS:DM instrument to measure neurodevelopmental delay in children in Bhutan.





Methods: In total, 96 community-dwelling Bhutanese children (3-7 years old) without diagnosed neurocognitive conditions were recruited from ambulatory clinics in urban Bhutan in 2016 as part of a larger study on retinal imaging and cognitive and growth parameters. Scoring was based on neurocognitive domains (gross and fine motor, receptive and expressive speech, self-help, social-emotional). Rates of failure (meant to indicate delay) within domains were calculated.

Results: Modifications of some standard questions were deemed necessary by the study staff to suit the cultural context, such as replacing kickball with football in a question regarding games played with rules to maintain local relevance. In a modified PEDS:DM test with these improvised modifications, the mean percentage of age-appropriate domains failed was 58.8% and the mean percent delay was 12.3% (range 0-41.4%, available in n=83). The highest prevalence of failures was 59.4% for receptive language and 76.3% for expressive language, much higher than the lowest rate of failure seen in self-help (5.4%).

Open Peer Review

Approval Status    

	1	2	3	4
version 2 (revision) 11 Sep 2019	 view			
				
version 1 20 Jun 2019	 view	 view	 view	 view

1. **Philip Wilson** , University of Aberdeen, Inverness, UK
2. **Charles R. Newton** , Kenya Medical Research Programme-Wellcome Trust Collaborative Programme, Kilifi, Kenya
3. **Naila Z. Khan** , Bangladesh Protibondhi Foundation, Dhaka, Bangladesh
4. **Issarapa Chunsuwan** , Thammasat University (TU), Pathum Thani, Thailand

Any reports and responses or comments on the article can be found at the end of the article.

Conclusions: The PEDS:DM requires further modifications and validation studies before it can be reliably implemented to assess developmental delay in children in Bhutan. In this pilot study, the rate of delay as reported by the PEDS:DM would be scored as markedly elevated, especially when compared to available epidemiologic studies in the region.

Keywords

children, cognition, developmental delay, neurodevelopment, epidemiology

Corresponding author: Farrah J. Mateen (fmateen@mgh.harvard.edu)

Author roles: **Wong B:** Writing – Original Draft Preparation; **Grundy S:** Data Curation, Writing – Review & Editing; **Tshering L:** Data Curation; **Tshering K:** Supervision; **Mateen FJ:** Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the Bill and Melinda Gates Foundation [OPP1116337].

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2019 Wong B *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Wong B, Grundy S, Tshering L *et al.* **Assessment of a neuro-developmental screening tool in children in Bhutan [version 2; peer review: 3 approved, 1 approved with reservations]** Gates Open Research 2019, 3:1496 <https://doi.org/10.12688/gatesopenres.13037.2>

First published: 20 Jun 2019, 3:1496 <https://doi.org/10.12688/gatesopenres.13037.1>

REVISED Amendments from Version 1

The reviews from the international group of expert reviewers led to several important revisions and clarifications in our manuscript. The discussion was expanded to report on issues of bilingualism in children, additional sources of potential bias in our convenience sample, and a comment that our female preponderance in this study was likely a spurious finding. One reviewer remarked that "surveillance" is a more appropriate term than "screening" and changes to the terminology were made where appropriate. An additional reviewer made several suggestions on clarifying the tables (2 and 3 in particular). As a result, a new table 4 was created, terms and clarifications were made on the tables, and details were changed such as subheadings, footnotes, and units of analysis. Where helpful, additional changes to improve the readability of the manuscript were also made. Multiple references to the literature, upon the suggestion of the reviewers, that are pertinent to our study were added.

Any further responses from the reviewers can be found at the end of the article

Introduction

Neurodevelopmental delay is a common finding in United States pediatric practices, with the prevalence measured at 12–16% of all children^{1,2}. For the purpose of referring children to appropriate early intervention services, various surveillance tools have been designed to help clinicians identify developmental delay along different cognitive domains. While in Western countries there exist robust protocols for surveilling children to identify developmental delay, there are few available data regarding the utility of such protocols in lower income countries. Nonetheless, it is estimated that the majority of children with developmental disabilities (up to 80%) reside in low- or middle-income countries³.

A commonly utilized developmental screening tool in the U.S. is the [Parents' Evaluation of Developmental Status: Developmental Milestones](#) (PEDS:DM)⁴. The survey is designed to assess developmental delay in children from birth to age 7–11 years across several cognitive domains: fine motor, gross motor, receptive language, expressive language, self-help, and social-emotional. While the PEDS:DM has been well validated in the United States where the instrument was constructed^{5,6}, there are few studies confirming its validity across cultural contexts⁷.

We aimed to assess the validity of the PEDS:DM survey in the lower-income country of Bhutan, a Himalayan country with no pediatric neurologists. Bhutan is typical of several lower-income countries where English is a language of school instruction and pediatric neurodevelopmental surveillance could be valuable, particularly if used by non-physician health care workers in the field.

Methods

Ethics approvals

The study was approved by the Research Ethics Board of Health, convened by the Bhutanese Ministry of Health (PO-2015-011) and the Institutional Review Boards at Massachusetts General Hospital (2015P000159) and Children's Hospital of Los Angeles (CHLA-15-00539). Neurocognitive screening was

implemented as part of a larger study to correlate cognitive functioning and retinal and other growth parameters⁸. Informed consent (written, or via thumb print if parent had low literacy) was obtained from each subject's parent or next of kin over the age of the majority.

Location

The Kingdom of Bhutan is bordered by China and India. In 2014, the total expenditure on healthcare per capita was 89 USD, compared to 9402 USD in the USA⁹. Subjects were recruited from the Jigme Dorji Wangchuck National Referral Hospital (JDWNRH) located in the capital city, Thimphu (population 138,736 in 2017)¹⁰. Specialist care is available, but there are no pediatric neurologists or subspecialized developmental pediatricians. Patients presenting to the JDWNRH Department of Ophthalmology outpatient clinic for routine vision screening were recruited as subjects so as to recruit community-dwelling children who were not seeking care for neurodevelopmental or neurological issues directly. Dzongkha is the primary spoken language in Bhutan. English is taught in schools and is the primary written language¹¹.

Enrollment

Recruitment took place in 2016. Inclusion criteria included all children aged three to seven years old. Children who did not have a parent present and who were unable to complete a retinal scan as part of a concurrent study⁸ were excluded from the cognitive testing. Subjects were provided with reimbursement of travel expenses amounting to 500 Bhutanese Ngultrums (equivalent to approximately 12 USD).

Anthropometric characteristics

At the time of enrollment, each child's height, weight, and head circumference were measured and recorded according to the methods of the World Health Organization¹². Z scores were calculated for the height, weight, and head circumference of each child by an independent scorer using normative data from the World Health Organization.

Neurocognitive testing

The PEDS:DM survey was conducted on all eligible children by a U.S.-based pediatrician (BW) who was a fellow of pediatric neurology and native Bhutanese research staff (LT, KT). The survey assesses for cognitive development up to a maximum age limit, varying by task category. The domains and their corresponding upper age limits are: fine motor (6 years, 11 months), receptive language (7 years, 11 months), expressive language (6 years, 11 months), gross motor (4 years, 5 months), self-help (6 years, 11 months), and social-emotional (5 years, 5 months). Each domain includes questions of increasing difficulty correlated with specific ages. For each domain, each child was asked the question for their age, and then if incorrect, was asked the question for progressively younger age groups until he or she provided two consecutive answers correctly. Several questions were modified to provide greater cultural relevancy (see Extended data¹³).

The reading and math sections of the screen were not used because those questions were deemed irrelevant to the population.

Examples include a math question which requires recognition of a penny. Bhutanese currency does not include coins. We also excluded a reading question that involved recognition of American “Exit” and “Caution” signs. To control for potential differences in parental literacy levels, surveys were conducted via interview by the pediatrician rather than having parents fill out a written survey. Subjects were only assessed in domains for which they were not over the maximum age limit for that domain. No other questions were excluded from the testing.

Although Dzongkha is the primary spoken language in Bhutan, the evaluations were conducted in English due to high levels of English proficiency in the majority of participants and parents¹¹. Forward and backward translation to Dzongkha was provided by a Bhutanese research assistant, who was available at every participant encounter, when requested.

With the assistance of the Bhutanese research assistant, some questions were mildly modified to be culturally relevant. For example, the social-emotional domain question, “can your child play games with rules, like board or card games, kickball, or hopscotch,” was altered to the example of “football” which is a more common childhood game in Bhutan.

Scoring

Responses to questions were scored either based on parental reports of the child’s experience, or direct subject demonstration of the required task, as appropriate. For example, on a question in the fine motor domain, “can your child write any of the letters of the alphabet?” parents were asked to answer this question from their past experience of their child, and the subject was asked to perform the task. If there was a discrepancy between parental report and subject performance, then the score was determined by subject performance.

The PEDS:DM responses were scored as both an age-dependent *screen* (pass or fail for the age appropriate task) and *assessment* (percentage delay). For example, if the child was 6 years 5 months old and the highest level of a correctly answered question

was for the level of 6 years 0 months, then she was scored as having failed the age-dependent *screen*. She would be considered 7.7% delayed $((1-(72 \text{ months}/78 \text{ months}))*100)$ in the *assessment*.

In addition to determining the number of children who fail (screen) and the percentage delay (assessment) within individual domains, a *screen composite score* (the number of domains failed divided by the number of age appropriate domains assessed) and *assessment composite score* (mean percent delay) were calculated across all domains per subject. This provided an overall assessment of delay for each child combining the domains assessed. To assess the patterns of failure across the study population, the rates of failure for each domain were assessed for all children as well as for each age group individually. Descriptions of each scoring metric are available in [Table 1](#).

Statistical analysis

Basic data analysis was descriptive and included means, proportions, and percentages for the variables of interest. Where sample sizes were reduced due to incomplete data, the analyzable subcohort sample size is provided. Scoring of the PEDS:DM test was performed using the provided testing manual by a study staff member who did not perform the cognitive or anthropometric testing of the subjects directly (SG). All calculations were performed using [Stata 12](#) (StataCorp. 2011. *Stata Statistical Software: Release 12*. College Station, TX: StataCorp LP).

Results

Subject demographics and anthropometric characteristics

In total, 96 children (37 male) aged 3 years, 1 month to 7 years, 11 months old were enrolled. There were no refusals from parents to have their child participate. The mean age was 6 years 0 months old. 26 subjects were not yet enrolled in school ([Table 2](#) and [Underlying data¹⁴](#)). Head circumference and height measurements followed a normal distribution. Weight and body mass index measurements skewed towards underweight (65.2% of subjects scored were under average weight) ([Figure 1](#)).

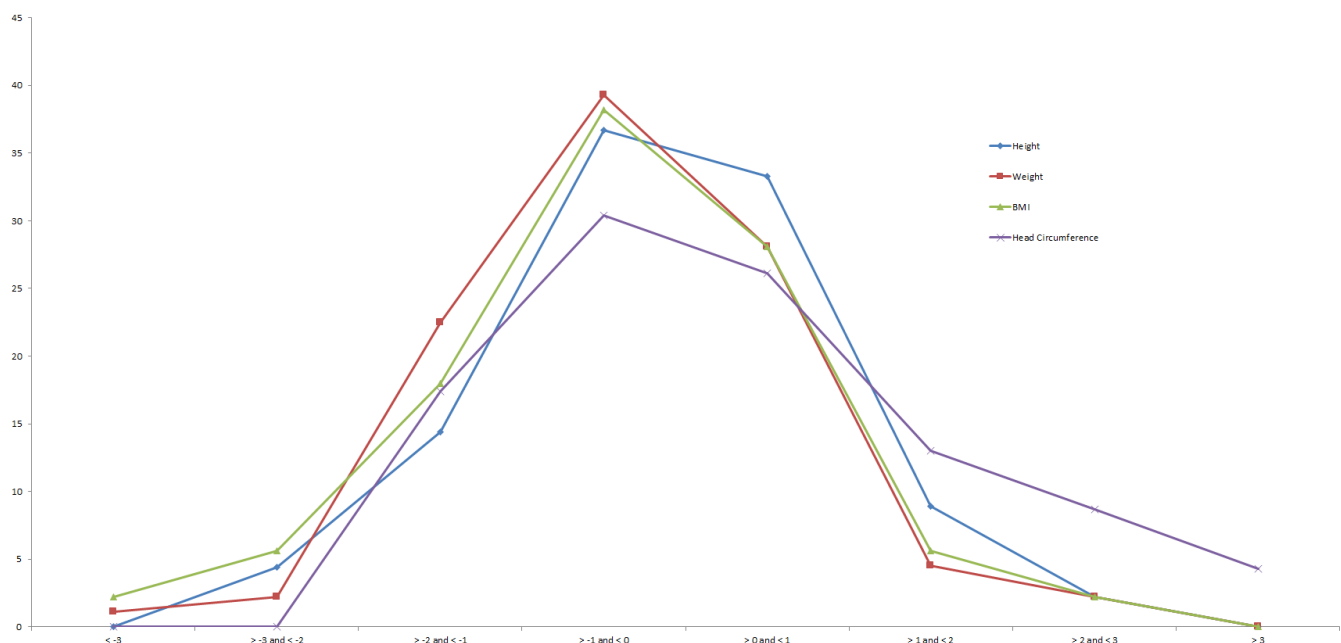
Table 1. Definitions of tool metrics on the PEDS:DM.

Metric	Subject(s) Assessed	Number of Developmental Domain(s) Assessed	Calculation
Domain specific screen	One child	Single	Pass or fail determination on the age-appropriate question for the domain
Screen composite score	One child	Multiple	Number domains failed/Number domains assessed
Domain specific assessment	One child	Single	$(1 - [\text{Age level of the highest correctly answered question}] / [\text{child's age} * 100])$
Assessment composite score	One child	Multiple	Mean percentage delay across all domains assessed
Domain screen failure	All children	Single	Number of children who failed the domain/ Number of children assessed for the domain
Mean assessment percent delay	All children	Single	Mean percent delay for all children across a single domain

Table 2. Participant characteristics at time of enrollment: age, education, medical comorbidities. (n=96 children).

	Male	Female	Total
Total Participants*	37	54	
Mean Age (months) \pm SD	74 \pm 14	70 \pm 14	72 \pm 14
Range (months)	37–95	37–94	37–95
Duration of Schooling			
No school	9 (24%)	15 (28%)	26 (28%)
>0 and <6 months	3 (8%)	1 (2%)	4 (4%)
>6 and <12 months	4 (11%)	7 (13%)	12 (13%)
>12 and <24 months	8 (22%)	15 (28%)	24 (26%)
>24 months	13 (35%)	14 (26%)	27 (29%)
Medical Complications	(n=36)	(n=52)	
Pregnancy complications (mother) (n=93)	1 (3%)	5 (10%)	6 (6%)
Pregnancy hospitalizations (mother) (n=87)	4 (12%)	8 (15%)	12 (13%)
Child medical problems (n=88)	5 (14%)	7 (14%)	12 (13%)

*Sex not recorded for 5 participants.

**Figure 1. Distribution of weight, height, and body mass index for age, according to World Health Organization standardized growth curves.** Y-axis is number of children (n=96). X-axis is z-score.

PEDS:DM Scores

PEDS:DM screen composite scores were generated for 92 of 96 subjects, with possible scores ranging from 0 to 100. For the screen composite score, only an answer to the age appropriate question was required. Four children were missing data on the age appropriate question for their age. Of the remaining 92 children who could be scored, the average score was 41.2 points.

The *overall assessment score* determines the percentage of delay for a child across all domains. It required both the age-appropriate question and younger age groups' questions to be answered correctly. Two questions in a row answered correctly were required to give a score. Of the 96 subjects, 13 were missing data that made the percent delay impossible to score. Of the remaining 83 children, the average score was 12.3 (range 0–41.4).

Domain screen failure rates were calculated for individual neurocognitive domains. The language domains contained the largest rates of failure on the PEDS:DM: 59.4% for receptive language, 76.3% for expressive language. The lowest rate of failure was seen in self-help (5.4%).

The domain screen failure rates produced non-uniform results across childhood age groups. Percentage failure of the fine motor domain by age group ranged from 33.3–44.4%, with the striking exception of 0% for six-year-olds. Five-year-old children had the highest failure rate in receptive language (94.1%) compared to other ages, yet had the lowest failure rate in expressive language (64.7%). For the social-emotional domain, failure rate appeared to increase with age (Table 3).

The assessment composite score (percent delay) measured the degree of delay within particular domains. The score

across all domains was 12.3% (range 0–41.4%, n=83). Similar to the screen composite scores, the language domains had the highest assessment composite scores. Within the language domains, the greatest delays were in the five- and six-year-old age groups (receptive language 24.1–29.3%, expressive language 22.2–28.4%). Relatively smaller total percent delays were observed in the motor (both fine and gross), social-emotional, and self-help domains, in a manner that roughly aligns with the failure rates for these domains in the domain specific screen.

Discussion

The validation of culturally relevant screening tools provides a means for more accurate assessment of developmental delay in children in lower income countries as well as prevalence data on the extent of developmental delay in these locations. Here we evaluated the applicability of the PEDS:DM in Bhutan to determine its utility in this non-Western setting.

Table 3. Parental Evaluation of Developmental Status: Developmental Milestones (PEDS:DM) results. “Total positively screened” refers to the number of children by age group who screened positive in any single domain. The average screen composite score represents the number of domains failed over the number of domains assessed. The mean assessment is a measure of the mean percent delay across a single domain, while the average assessment composite score measures the mean percent delay across all of the domains assessed.

Age years [months]	3 [36-47]	4 [48-59]	5 [60-71]	6 [72-83]	7 [84-95]	Total
Number of subjects assessed	6	18	17	35	20	96
Total positively screened by domain						
Fine Motor (<95mo)	2 (33.3%)	8 (44.4%)	7 (41.2%)	0 (0.0%)	8 (40.0%)	25 (26.0%)
Receptive Language (<95mo)	1 (16.7%)	10 (55.6%)	16 (94.1%)	20 (57.1%)	10 (50.0%)	57 (59.4%)
Expressive Language (<83mo)	4 (66.7%)	15 (83.3%)	11 (64.7%)	28 (80.0%)	-	78 (76.3%)
Self-help (<83mo)	0 (0.0%)*	1 (5.6%)	1 (6.2%)*	2 (5.7%)	-	26 (5.4%)
Social-emotional (<65mo)	0 (0.0%)	6 (35.3%)*	6 (66.7%)**	-	-	76 (37.5%)
Gross Motor (<53mo)	1 (16.7%)	1 (16.7%)**	-	-	-	86 (16.7%)
Average screen composite score by age group	26.6 ± 8.8*	41.8 ± 21.8*	51.6 ± 21.2*	36.0 ± 17.6*	45.0 ± 39.4	41.2 ± 25.3*
	*n=5	*n=17, **n=6	*n=16, **n=9	*n=34		*n=92
Mean assessment percent delay (% delay ± SD)						
Fine Motor (<95mo)	4.5 ± 7.1	13.0 ± 13.0	8.6 ± 14.0	0 ± 0	2.9 ± 4.4	4.8 ± 9.7
Receptive Language (<95mo)	1.9 ± 4.7	18.8 ± 18.9	24.1 ± 12.0*	29.3 ± 12.7*	21.7 ± 22.1*	23.0 ± 17.1
Expressive Language (<83mo)	4.7 ± 4.2	19.6 ± 15.5	22.2 ± 9.1**	28.4 ± 12.0**	-	22.8 ± 13.5
Self-help (<83mo)	0 ± 0*	0.3 ± 1.3	0.3 ± 1.2**	1.3 ± 5.7	-	0.7 ± 4.0
Social-emotional (<65mo)	0 ± 0	5.7 ± 11.0*	6.1 ± 5.5***	-	-	4.7 ± 8.5
Gross Motor (<53mo)	2.9 ± 7.1	10.1 ± 16.0**	-	-	-	6.5 ± 12.4
Average assessment composite score by age group	2.3 ± 1.6	11.1 ± 10.0*	12.9 ± 6.3*	14.7 ± 5.8*	12.1 ± 11.5*	12.3 ± 8.6
	*n=5	*n=15, **n=6	*n=15, **n=16, ***n=9	*n=33, **n=32	*n=19	*n=83

Cultural differences leading to suboptimal performance

The exceedingly high rates of calculated delay in expressive and receptive language limit the ability to accurately interpret the other domains of the PEDS:DM test. A similar conclusion was reached in a comparative study conducted using the PEDS versus the PEDS:DM in a population of Thai children¹⁵. That study reported a greater proportion of the subjects being classified with medium risk for developmental delay when compared to prior standardization studies in the United States. The authors postulated that local culture influenced certain responses of parental concern about their child's development (such as expressing worry about a child being left-handed), resulting in an overestimation in the identification of actual delay. The high rates in the early school-year ages when Bhutanese children are increasingly exposed to the English language may also contribute to a perceived language delay, when truly there is none. Such a phenomenon has been reported consistently in bilingual children during early developmental stages of language, who are initially perceived to have expressive language delay in both languages, but eventually mature and have no language delay in either language¹⁶.

The PEDS:DM was originally designed and intended for implementation in a Western patient population. Bhutanese cultural differences appeared to consistently invalidate certain questions of the test. For instance, for the self-help question, "Can your child get dressed by himself or herself?" almost all of the parents answered "no." Bhutanese children are dressed in traditional outfits that are more complicated to put on compared to Western-style clothing, and thus children are usually unable to independently dress until an age much later than what was tested on the PEDS:DM. Children were often unable to correctly answer a question regarding irregular plural nouns (i.e. "tooth" and "teeth"), suggesting such irregular English grammar structures are not introduced in the early school years, when most Bhutanese children are taught English. This may also account for the peak for delays in the assessment for expressive and receptive language occurring in five- and six-year old children, ages when subjects are first enrolled in formal schooling.

In school-age children, there was striking consistency in the manner that incorrect answers were provided. This may suggest a difference in academic curricula between the United States and Bhutan that led to suboptimal performance of Bhutanese children on an American-designed test. One such example may explain the high failure rate in receptive language among five-year-old children. In order to pass this domain, subjects had to identify at least three of the following body parts: shoulders, elbows, heels, ankles. Nearly all five-year-old children could correctly identify shoulders and elbows and very few could identify ankles and heels.

For six-year-old children to earn a "pass" on the fine motor domain, they had to correctly write three or more letters of the alphabet and draw a triangle with all lines connecting correctly. The abnormally high pass rate (100%) for this category amongst six-year-old children may be due to such skills being highly emphasized in school curriculum or at home by this age.

Our study gave credit to subjects who were able to demonstrate tasks, rather than accepting parents' report on subjects' ability. For example, subjects on one question were asked to recite the alphabet and would only be scored as "correct" if subjects could do so during the study visit. Subjects for the most part provided attempts at most questions. Valuing parental reporting over subject performance may have been a more representative assessment of the child's abilities, as parents would more likely know the child's skill performance in a natural environment.

Comparison to the Two-Stage Child Disability Study Findings

The Two-Stage Child Disability Study was conducted by UNICEF and the National Statistics Bureau of Bhutan and was the largest recent country-wide screen of disability in Bhutan, including 3491 subjects¹⁷. The study utilized a first-stage screen with the Ten Questions questionnaire (TQ)¹⁸. Those who screened positive were then screened with the second stage: the Rapid Neuro-developmental Assessment (RNDA) for two to five year old children and the Rapid Functional Assessment (RFA) for five to nine year old children¹⁹. The TQ, RNDA, and RFA were designed for use in low-income settings and focus on general function abilities, avoiding culture-specific skills (i.e. tying shoelaces). These tests were initially developed in large epidemiologic studies in Bangladesh, conducted by the Bangladesh Protibondhi Foundation, who provided these tests to the Two-Stage Child Disability Study¹⁷. The tools have since been validated in other lower income settings on a country-wide scale in Jamaica and Pakistan^{20,21}.

The Two Stage Child Disability Study provides a worthwhile comparison for our results. It reached a wider group of children geographically and with a much larger and population-based sample. While the authors of that study noted lower sensitivity and specificity of the first-stage TQ methodology when compared to similar studies conducted in other countries, they believed the final determination of *disability* from the second-stage screen to be accurate. Similar to our study, the Two-Stage Child Disability Study uses assessment tools that were created and validated elsewhere (Bangladesh), however the authors were careful to choose these assessments for their lack of testing of culture-specific skills.

As a rough point of comparison, when our data are compared to that of the Two-Stage Child Disability Study, we see that in all domains for which there was a direct comparison available, the percentage of disability measured by PEDS:DM exceeded that seen in the Two-Stage Study (Table 4). However, with only one available comparable study, normative data in this population remain unclear.

Limitations and strengths

One limitation of our study is sampling from a more developed region of Western Bhutan. Our study does not generalize to the entire Bhutanese population which is more rural and remote, such as in the Eastern gewogs (Dzongkha term for a group of villages) where we did not enroll subjects. We suspect children suffering from significant disability in a lower income

Table 4. Comparison of PEDS:DM results to Two-Stage Child Disability Study in Bhutan.

Screen (failure rates)	PEDS Total	Two-Stage Study [% with disability]
	n=96	n=3491
Fine Motor (<95mo)	25 (26.0%)	199 (5.5%)
Receptive Language (<95mo)	57 (59.4%)	-
Expressive Language (<83mo)	78 (76.3%)	102 (2.3%)
Self-help (<83mo)	26 (5.4%)	-
Social-emotional (<65mo)	76 (37.5%)	-
Gross Motor (<53mo)	86 (16.7%)	86 (2.0%)

country could have accessibility issues, going to a hospital less often to accompany their parents for routine eye care or other appointments, and thus these children would not be well represented here. There is also the potential for selection bias, as subjects who could not complete the retinal scan as part of the larger study were not selected to complete the PEDS:DM. The retinal scan required children to sit still for several seconds; children with severe developmental delay would have had a more difficult time participating and would not be reflected in the PEDS:DM screening here. Furthermore, our subjects were chosen from children receiving routine vision examinations. This may bias our sample towards children whose parents have higher educational levels and socio-economic status, since these parents may be more likely to seek out such care. Parents seeking routine vision screening may also be more likely to have concerns regarding their child's vision; if present, visual problems could be comorbid with developmental conditions.

While Dzongkha instructions were provided verbally, having written materials may also improve response accuracy during testing. Additionally, the PEDS:DM evaluation is designed as a screening test and therefore is intentionally sensitive in comparison to a confirmatory test such as was used in the second-stage of the Two Stage Child Disability study. This could be one of the reasons for the inflated rates of developmental delay measured in this study. It is also possible that subject performance anxiety increased the number of responses categorized as "incorrect." If children are not accustomed to directly interacting with a healthcare provider during outpatient clinic visits, then requesting cognitive tasks may be overwhelming, leading to worse test outcomes. Finally, our small sample size overall and within each age group may contribute to the wide variations displayed between groups.

The female preponderance in our sample is considered a spurious finding. At the time of this study, the population ages 5 to 9 years old in Bhutan was 50.9% male and 49.1% female²². To our knowledge there are no gender disadvantages at the pre-school or early school years for Bhutanese children. Nor are there known cultural considerations that may favor school age girls to receive routine medical care over boys.

Our study strengths include assessment of a community-dwelling pediatric population that is not well represented in the neurodevelopmental research literature. We are unaware of other studies of the PEDS:DM in Bhutan or neighboring regions. Children were evaluated by a senior fellow of pediatric neurology where none are otherwise available. We also provide a detailed assessment of how and likely why children may be erroneously *labeled as delayed* using a well-established surveillance tool in the U.S.A.

We demonstrate that the PEDS:DM may not be an appropriate developmental surveillance tool for use in children in Bhutan due to cultural differences in questions and required developmental testing tasks. When compared to the Bhutan Two-Stage Child Disability Study which recruited 3,491 Bhutanese children, our study of just less than 100 community-dwelling Bhutanese children using PEDS:DM categorized significantly more subjects with disability in one or more neurocognitive domains, at times more than half of the study population (Table 4).

Diagnostic assessments that are sensitive to small differences in developmental delay are an important tool for early discovery and allow for early interventions in these cases. Low- and middle- income countries have the highest burden of developmental delay yet lack these important tools to recognize changes early when the largest differences on children's outcomes can be made through early life intervention. The presumed lack of cross-cultural applicability of the PEDS:DM to a population in Bhutan shows the importance of developing pediatric neuro-developmental tools to better surveil children and determine accurate prevalence data in non-Western, non-English speaking countries. Locally-constructed developmental surveillance instruments such as the Angkor Hospital for Children Milestone Assessment Tool²³ utilized in Cambodia, the South Africa Road to Health Booklet²⁴ utilized in South Africa, and the Malawi Developmental Assessment Tool²⁵ utilized in Malawi, represent the promise of more valid, culture-specific tools on the horizon in other locations. While we do not suggest that tools such as PEDS:DM have no utility in this setting, we would advise culture-specific revisions of existing assessments before widespread implementation studies.

Data availability

Underlying data

Figshare: PEDS:DM Testing in Bhutan. <https://doi.org/10.6084/m9.figshare.8226644>¹⁴

This project contains the following underlying data:

Bhutan_fullTable_4June2019.xlsx (This is raw field data from 96 Bhutanese children undergoing the PEDS:DM evaluation. Cells where "N/A" or "not applicable" is entered reflect that the screening instrument is staged by the age of subject. Several subjects either were too young or failed to progress through the assessment, making more difficult questions no longer applicable to their testing.)

Extended data

Figshare: Extended Data: Field Notes and Modifications to the PEDS:DM Test for Bhutanese Children. <https://doi.org/10.6084/m9.figshare.8246282>¹³

This project contains the following extended data:

- Modifications to PEDSDM_8Jun2019_10PM_Fig-Share.pdf (These are the field notes of the investigators

on the modifications to the PEDS:DM in Bhutan. They are provided in text format and itemized by cognitive testing domain.)

Acknowledgements

Dechen Wangmo, MD of the Department of Ophthalmology, Jigme Dorji Wangchuck National Referral Hospital, Thimphu, Kingdom of Bhutan.

References

- Boyle CA, Decoufle P, Yeargin-Allsopp M: **Prevalence and health impact of developmental disabilities in US children.** *Pediatrics*. 1994; **93**(3): 399–403.
[PubMed Abstract](#)
- Glascoe FP, Dworkin PH: **The role of parents in the detection of developmental and behavioral problems.** *Pediatrics*. 1995; **95**(6): 829–36.
[PubMed Abstract](#)
- Kolucki B: **Global trends.** Rehabilitation International Publications. 1999; **20**: 3–9.
- Glascoe FP, Robertshaw NS: **PEDS: Developmental Milestones.** Nashville, TN: Ellsworth & Vandermeer Press; 2007. Last accessed March 1, 2019.
[Reference Source](#)
- Glascoe FP: **Summary of PEDS Research from Glascoe FP, Collaborating with Parents.** 2nd edition. Nolensville, TN; 2013.
[Reference Source](#)
- Glascoe FP: **Collaborating with Parents: Using Parents' Evaluations of Developmental Status (PEDS) to Detect and Address Developmental and Behavioral Problems.** Ellsworth and Vandermeer. Nashville, TN; 1998.
[Reference Source](#)
- Kijing JS, Low PS, Chan YH, *et al.*: **Interpreting parents' concerns about their children's development with the Parents Evaluation of Developmental Status: culture matters.** *J Dev Behav Pediatr*. 2012; **33**(2): 179–83.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Grundy SJ, Tshering L, Wanjala SW, *et al.*: **Retinal Parameters as Compared with Head Circumference, Height, Weight, and Body Mass Index in Children in Kenya and Bhutan.** *Am J Trop Hyg*. 2018; **99**(2): 482–88.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- World Bank Data: **The World Bank Group.** Last accessed May 25, 2019.
[Reference Source](#)
- National Statistics Bureau of Bhutan: **Launch of 2017 population and housing census of Bhutan report.** 2018. Last accessed online May 7, 2019.
[Reference Source](#)
- Dorjee K: **Linguistic landscape of Bhutan: An overview of number of languages, language policy, language education, and language use in Bhutan.** *Bhutan Journal of Research & Development*. 2014; **3**: 87–102.
[Reference Source](#)
- WHO Multicentre Growth Reference Study Group: **WHO Child Growth Standards: Length/ height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development.** Geneva, World Health Organization, 2006.
[Reference Source](#)
- Wong B, Mateen F, Grundy S, *et al.*: **Extended Data: Field Notes and Modifications to the PEDS:DM Test for Bhutanese Children.** *figshare*. Journal contribution. 2019.
<http://www.doi.org/10.6084/m9.figshare.8246282>
- Mateen F, Tshering L, Grundy S, *et al.*: **PEDS:DM Testing in Bhutan.** *figshare.com*. Dataset. 2019.
<http://www.doi.org/10.6084/m9.figshare.8226644>
- Chunsuwan I, Hansakunachai T, Pomsamrit S: **Parent Evaluation of Developmental Status (PEDS) in screening: The Thai experience.** *Pediatr Int*. 2016; **58**(12): 1277–1283.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Uljarević M, Katsos N, Hudry K, *et al.*: **Practitioner Review: Multilingualism and neurodevelopmental disorders - an overview of recent research and discussion of clinical implications.** *J Child Psychol Psychiatry*. 2016; **57**(11): 1205–1217.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Two-Stage child disability study among children 2 to 9 years: Bhutan 2010–2011.** National Statistics Bureau, Government of Bhutan. 2012.
[Reference Source](#)
- Stein ZA, Durkin MS, Davidson LL, *et al.*: **Guidelines for identifying children with mental retardation in community settings.** In: World Health Organization. *Assessment of people with mental retardation*. Geneva: World Health Organization, 1992; 12–41.
- Khan NZ, Muslima H, Shilpi AB, *et al.*: **Validation of rapid neurodevelopmental assessment for 2- to 5-year-old children in Bangladesh.** *Pediatrics*. 2013; **131**(2): e486–94.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Thorburn M, Desai P, Paul T, *et al.*: **Identification of childhood disability in Jamaica: the ten question screen.** *Int J Rehabil Res*. 1992; **15**(2): 115–127.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Durkin MS, Davidson LL, Desai P, *et al.*: **Validity of the ten questions screened for childhood disability: results from population-based studies in Bangladesh, Jamaica, and Pakistan.** *Epidemiology*. 1994; **5**(3): 283–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
- 2017 Population & Housing Census of Bhutan.** National Statistics Bureau, Government of Bhutan. 2018.
[Reference Source](#)
- Ngoun C, Stoeys LS, van't Ende K, *et al.*: **Creating a Cambodia-specific developmental milestone screening tool - a pilot study.** *Early Hum Dev*. 2012; **88**(6): 379–85.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Francis D: **South African National Department of Health. Road to Health Booklet.** 2011. Last accessed July 18, 2018.
- Gladstone M, Lancaster GA, Umar E, *et al.*: **The Malawi Developmental Assessment Tool (MDAT): the creation, validation, and reliability of a tool to assess child development in rural African settings.** *PLoS Med*. 2010; **7**(5): e1000273.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:    

Version 2

Reviewer Report 12 September 2019

<https://doi.org/10.21956/gatesopenres.14192.r27825>

© 2019 Wilson P. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Philip Wilson 

Centre for Rural Health, University of Aberdeen, Inverness, UK

Thank you for responding to the reviews so comprehensively. The article now reads very well and I am happy to give my approval.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Developmental assessment of children, primary care, epidemiology.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 27 August 2019

<https://doi.org/10.21956/gatesopenres.14150.r27618>

© 2019 Chunsuwan I. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Issarapa Chunsuwan 

Department of Pediatrics, Faculty of Medicine, Developmental and Behavioral Pediatrics, Thammasat University (TU), Pathum Thani, Thailand

Thank you for giving me a chance to review this article. Research on the developmental screening in eastern countries and developing countries is interesting. Because of differences in context and

culture. The use of screening tools which developed from the basis of western countries may have different results from from using that tool in the country of origin. This research shows that the use of PEDS: DM screening tools in Bhutan revealed different results from the United States.

I have two main ideas about this research and some small suggestions.

First, for the interpretation of the screening results using the PEDS: DM tools, it should be interpreted in the same way as the original, as this research aim to compare the use of this tool in Bhutan compare with original and other country. In table 1 which showed the definition of screening tool interpretation is quite different. The interpretation methods of the screening results by PEDS: DM tool, showed in table1, were different from the original. In the original, pass screening means that a child passes every item in a specific age range. Creating a new interpretation method made a comparison between the use of screening tools in Bhutan with in other countries difficult and confusing. The assessment composite score, which wants to show how slowly the child is behind the benchmark, seem not be the proper method for calculating the percentage of delay. The using of PEDS: DM assessment level screening should be more appropriate for evaluation. This version provides both age-equivalent scores, which can be used to determine percentage of delay¹.

Second, In table 2, the author wanted to show the screening results that how many children in each age group have delays on each domain. Reporting the number of children with at least 1 delay in each age group may be more useful than showing the average screen composite score of age. Since the first reporting method can get values that are compared with other research.

For minor suggestion:

- Table 3, demographic data should presented before table 2
- Table 3, some numbers is confused, such as PEDs total number of gross motor domain was 86, while the total number of children age 3 + 4 who were screened in this domain was only 24
- The 2 staged study should be explained before showing results in table 3
- In the discussion about screening in Thailand, in that study, we use PEDS for screening (asked concern of parents about their child's delay in each domain) and compared the results with PEDS: DM assessment level (assess children's ability by researcher). We didn't use the PEDS: DM for screening.

I agree with the author that the screening tools developed from one place should be adjusted to be used in another area, especially in the field of language. Due to the different language formats, the tools used to assess language development based on English ability cannot be used in other languages. Interestingly, that Bhutan uses English as the primary language in school, however development of the secondary language is differ from the native speaker, and the modification of language screening part should be done.

References

1. Chunsuwan I, Hansakunachai T, Pornsamrit S: Parent Evaluation of Developmental Status (PEDS) in screening: The Thai experience.*Pediatr Int.* 2016; **58** (12): 1277-1283 [PubMed Abstract](#) | [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Developmental Pediatrics, PEDS screening

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 12 August 2019

<https://doi.org/10.21956/gatesopenres.14150.r27664>

© 2019 Khan N. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Naila Z. Khan

Clinical Neurosciences Center, Bangladesh Protibondhi Foundation, Dhaka, Bangladesh

The paper reads well. Reading the results, *without seeing the tables*, was easier. The tables are confusing in the discrepancies between the language of the column headings and the data provided. I have listed the confusions below.

Table 2:

1. The language of the first column headings and row numbers, as presented, don't match. For example, in the second row it says 'Screen (failure rates)', while all the numbers entered within brackets (why brackets?) in that row, by age group, probably indicate number of children who failed. The first column should have read 'number of screen positives' and not 'rates'. If rates are to be calculated then for each of the age groups (3,4,5,6,7) a 'total number screened' should be mentioned, and a percentage calculated. The next 8 rows calculate percentages of the 2nd row 'number of children'. Therefore, second row first

column heading needs to be changed.

2. Last column: Two Stage Survey being used for comparison: Conversely, the second row number is the 'Total number of children screened'. If we go by the 2nd row heading of 'Screen (failure rates)', it does not make sense.
3. First column row 10, under Average Score Composite Score by Age Group, is confusing, as this row does not have a texted heading. How does this row differ from row 19, i.e. last row of the table, which also has number of children in brackets with asterisks? Please clarify, or redesign this table. I understand that each item in the first column has a different set of children seen. Still, too many asterisks above (row 10) and below (row 19) to follow.

Table 3:

Rows and column heading are confusing. Repeats the same pattern as table 2.

1. Column one heading Age, years (96) simply repeats the Total Participants seen listed in row 2. Why not just say Mean Age in Years \pm SD (total=96) which is listed in row 4, and delete this row?
2. Years in School: This would mean to the reader 'years of schooling'. However only total male/female numbers are given in the row. The next 5 rows under Years in School are even more confusing, as the age range goes from 0-24 months, i.e., below the pre-school age group. Otherwise, the reasons for their inclusion has to be made clear.
3. Sex not recorded asterisk below the table for 5 children: The asterisk should be in the 2nd row after total number of participants, i.e., 96, as the numbers of male/female in that row don't match the total.

My conclusions: The authors should re-design their tables, bearing in mind that each table should be self-explanatory. That is, they should not depend upon the readers reading the text to be able to understand. Headings, sub-headings, and sub sub-headings might be helpful. Just a suggestion.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Child Neurology, Neurodevelopment, Neuroepidemiology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 22 Aug 2019

Farrah Mateen, Massachusetts General Hospital, Boston, USA

"The paper reads well. Reading the results, *without seeing the tables*, was easier. The tables are confusing in the discrepancies between the language of the column headings and the data provided. I have listed the confusions below."

Response: Thank you to the reviewer. We address the point by point comments in the revised text and in point form below.

Table 2:

"1. The language of the first column headings and row numbers, as presented, don't match. For example, in the second row it says 'Screen (failure rates)', while all the numbers entered within brackets (why brackets?) in that row, by age group, probably indicate number of children who failed. The first column should have read 'number of screen positives' and not 'rates'. If rates are to be calculated then for each of the age groups (3,4,5,6,7) a 'total number screened' should be mentioned, and a percentage calculated. The next 8 rows calculate percentages of the 2nd row 'number of children'. Therefore, second row first column heading needs to be changed."

Response: Adjusted first column heading to read "Total Positively Screened". Added row entitled "number of subjects assessed" to better clarify how many subjects were assessed and how many subjects had screened positive.

"2. Last column: Two Stage Survey being used for comparison: Conversely, the second row number is the 'Total number of children screened'. If we go by the 2nd row heading of 'Screen (failure rates)', it does not make sense."

Response: This column has been removed and placed into its own table (Table 4) for clarity.

"3. First column row 10, under Average Score Composite Score by Age Group, is confusing, as this row does not have a texted heading. How does this row differ from row 19, i.e. last

row of the table, which also has number of children in brackets with asterisks? Please clarify, or redesign this table. I understand that each item in the first column has a different set of children seen. Still, too many asterisks above (row 10) and below (row 19) to follow."

Response: We now provide a heading that details out the various instruments used.

Table 3:

"Rows and column heading are confusing. Repeats the same pattern as table 2.

1. Column one heading Age, years (96) simply repeats the Total Participants seen listed in row 2. Why not just say Mean Age in Years \pm SD (total=96) which is listed in row 4, and delete this row?"

Response: The mentioned row was deleted. We changed from measuring in years to months, as now seen in Table 2.

"2. Years in School: This would mean to the reader 'years of schooling'. However only total male/female numbers are given in the row. The next 5 rows under Years in School are even more confusing, as the age range goes from 0-24 months, i.e., below the pre-school age group. Otherwise, the reasons for their inclusion has to be made clear."

Response: This may have been a source of confusion. We simply meant to say "duration of school attendance to date."

"3. Sex not recorded for the asterisk below the table for 5 children: The asterisk should be in the 2nd row after total number of participants, i.e., 96, as the numbers of male/female in that row don't match the total."

Response: Asterisk position changed accordingly.

"My conclusions: The authors should re-design their tables, bearing in mind that each table should be self-explanatory. That is, they should not depend upon the readers reading the text to be able to understand. Headings, sub-headings, and sub sub-headings might be helpful. Just a suggestion."

Response: Thank you. We have also reordered and added a new table to clarify and allow the tables to stand separately from the paper. We thank the reviewer for the helpful comments and focus on the tables.

Competing Interests: None

Reviewer Report 07 August 2019

<https://doi.org/10.21956/gatesopenres.14150.r27420>

© 2019 Newton C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Charles R. Newton 

Kenya Medical Research Programme-Wellcome Trust Collaborative Programme, Kilifi, Kenya

This well-written paper describes the adaption and testing of the Parental Evaluation of the Developmental Status: Developmental Milestones (PEDS: DM) in 96 children (aged 3-8 years) who attended ambulatory clinics in urban Bhutan. The investigators found a high prevalence of developmental delay (59%), particularly in receptive language (59%) and expressive language (76%). They conclude (correctly in my view) that the PEDS: DM may not be culturally applicable to assess children in Bhutan.

It is not clear why the investigators choose to examine this tool in this setting when there are other screening tools e.g. Rapid Neurodevelopmental Assessment, Malawi Developmental Assessment Tool. A justification needs to be made more explicit.

There is little information provided on the adaption of the tool to the circumstances in Bhutan. There is a well-established process of adapting tools to different cultures and circumstances (van de Vijver and Tanzer, 2004¹).

The children were recruited from a clinic in a national referral hospital, and thus it is unclear how representative they are of the population. Most evaluations were conducted in English, suggesting that the population sampled were the more educated population. No data is provided on the socio-economic status of the parents or the maternal education.

The influence of sex does not appear to have been taken into consideration during the analysis. There was no information on the psychometric properties of the adapted PEDS: DM in Bhutan.

References

1. van de Vijver F, Tanzer N: Bias and equivalence in cross-cultural assessment: an overview. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*. 2004; **54** (2): 119-135 [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Neurodevelopmental disorders; epidemiology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 15 Aug 2019

Farrah Mateen, Massachusetts General Hospital, Boston, USA

"This well-written paper describes the adaption and testing of the Parental Evaluation of the Developmental Status: Developmental Milestones (PEDS: DM) in 96 children (aged 3-8 years) who attended ambulatory clinics in urban Bhutan. The investigators found a high prevalence of developmental delay (59%), particularly in receptive language (59%) and expressive language (76%). They conclude (correctly in my view) that the PEDS: DM may not be culturally applicable to assess children in Bhutan.

It is not clear why the investigators choose to examine this tool in this setting when there are other screening tools e.g. Rapid Neurodevelopmental Assessment, Malawi Developmental Assessment Tool. A justification needs to be made more explicit."

Thank you for making us aware of the other options available. We were not aware of such developmental screening tools at the time of this study. We now provide some examples of other relevant tools at the end of discussion.

"There is little information provided on the adaption of the tool to the circumstances in Bhutan. There is a well-established process of adapting tools to different cultures and circumstances (van de Vijver and Tanzer, 2004¹).

The children were recruited from an outpatient clinic in the national referral hospital. They represent a convenience sample only. It is unclear how representative they are of the full pediatric population. Most evaluations were conducted in English, suggesting that the population sampled were a more educated population. No data were queried on the socio-economic status of the parents or on the level of maternal education."

We plan to include this citation in the revised version.

We now discuss the potential for selection bias in our sample. You are correct that the sample is limited to a population living within travelling distance of the major urban center of the country. Since there is only one national hospital, it is possible that many children traveled long distances but more likely, these children have a marginally higher socio-economic status when compared to the rest of the country, which is primarily rural with less access to healthcare. Children were specifically selected as receiving routine vision screening, and maternal education likely play a part in which parents decide to seek such care. However, the entire health system is public, so all children would be seen free of direct costs.

“The influence of sex does not appear to have been taken into consideration during the analysis. There was no information on the psychometric properties of the adapted PEDS: DM in Bhutan.”

The following text will be included in results regarding gender distribution:

The female preponderance in our sample is thought to be a spurious finding. The percentage of population of males to females ages 5-9 years at time of study was 50.9% male to 49.1% female. To our knowledge there are no gender disadvantages at the pre-school or early school years. Nor are there known cultural considerations that may favor school age females for receiving routine medical care over males.

Competing Interests: None

Reviewer Report 02 August 2019

<https://doi.org/10.21956/gatesopenres.14150.r27617>

© 2019 Wilson P. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Philip Wilson

Centre for Rural Health, University of Aberdeen, Inverness, UK

Thank you for asking for an opinion on this interesting paper. It is a useful contribution to the world literature on developmental assessments and raises some important issues about the transfer of tools across cultures. The key finding is that, when using the PEDS: DM, a far higher proportion of children in Bhutan aged 3-7 years fail to meet expected milestones compared to children raised in the USA. The reasons for many such ‘failures’ lie in societal child-rearing practices described by the authors and are relatively easy to understand. Most would be relatively easy to remedy and the authors give valuable insights into how this might be achieved.

I have two significant comments to make about the paper, and a few minor points which are largely stylistic.

1. The first major point is that it is not clear how representative of the general population of Bhutan (or indeed of the locality from which the sample was drawn) the 96 children are. Is the female preponderance a chance finding? Is the ophthalmological screening programme from which the children were drawn free of charge? Is it well attended, apart from by children with severe disabilities? Are parents with worries about visual abilities of their children more likely to attend? These sorts of concerns are known to be more common among parents of children with autism, for example. There is some discussion of these matters on page 7 in the Limitations and Strengths section, but a much more critical analysis of the potential for selection bias and its possible contribution to the findings is required. It is well understood that a completely representative sample is extremely difficult to obtain in lower income countries but readers will want to have some sort of idea about the proportion and nature of the potentially eligible population who would not have attended the clinic.
2. My second comment is about the use of the word screening. It is commonplace to refer to developmental screening, but no neurodevelopmental assessments have to date met standard WHO screening criteria - largely because there is little evidence that early intervention is any better than late intervention for most problems in most of the tested domains. Screening is most appropriately used for one-off assessments conducted in the context of an organised programme: 'surveillance' more closely reflects what is being done here. The authors might be interested in the discussion of this matter in a recent paper comparing child developmental assessments in five countries (Wilson *et al.*, 2018¹).

More minor points:

- I found Table 2 a little difficult to follow. The use of parentheses is confusing and the final column on the two-stage study does not make sense until the reader reaches the Discussion section.
- Ages would be best expressed in Months in both Table 2 and Table 3, and probably elsewhere in the text.
- The legend to Figure 1 requires more information including total n.
- The discussion about language assessment on page 6 should perhaps reference the literature on bilingualism (I assume that at least some of the children fall into this category). Young children reared in bilingual households commonly appear to have delayed language acquisition in each language but eventually achieve fluency in both.
- There are some minor typographical issues, such as in the first sentence of the Results section of the Abstract and the second sentence of the Strengths and Limitations section.

References

1. Wilson P, Wood R, Lykke K, Hauskov Graungaard A, et al.: International variation in programmes for assessment of children's neurodevelopment in the community: Understanding disparate approaches to evaluation of motor, social, emotional, behavioural and cognitive function. *Scand J Public Health*. 2018; **46** (8): 805-816 [PubMed Abstract](#) | [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Developmental assessment of children, primary care, epidemiology.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 15 Aug 2019

Farrah Mateen, Massachusetts General Hospital, Boston, USA

“Thank you for asking for an opinion on this interesting paper. It is a useful contribution to the world literature on developmental assessments and raises some important issues about the transfer of tools across cultures. The key finding is that, when using the PEDS: DM, a far higher proportion of children in Bhutan aged 3-7 years fail to meet expected milestones compared to children raised in the USA. The reasons for many such ‘failures’ lie in societal child-rearing practices described by the authors and are relatively easy to understand. Most would be relatively easy to remedy and the authors give valuable insights into how this might be achieved.

I have two significant comments to make about the paper, and a few minor points which are largely stylistic.

1. The first major point is that it is not clear how representative of the general population of Bhutan (or indeed of the locality from which the sample was drawn) the 96 children are. Is the female preponderance a chance finding? Is the ophthalmological screening programme from which the children were drawn free of charge? Is it well attended, apart from by children with severe disabilities? Are parents with worries about visual abilities of their children more likely to attend? These sorts of concerns are known to be more common

among parents of children with autism, for example. There is some discussion of these matters on page 7 in the Limitations and Strengths section, but a much more critical analysis of the potential for selection bias and its possible contribution to the findings is required. It is well understood that a completely representative sample is extremely difficult to obtain in lower income countries but readers will want to have some sort of idea about the proportion and nature of the potentially eligible population who would not have attended the clinic."

We will include the following text in the discussion on limitations of study:

Our study includes a convenience sample only. The referral hospital from which subjects were recruited was in the major urban area of the country. Subjects were chosen from those receiving routine vision screening, which likely may skew towards parents of a higher education level and socio-economic status, who would be assumed to be mostly likely to seek such form of child care. Those parents seeking routine vision screening may also be more likely to have pre-existing concerns about their child's vision, which in turn may be associated with comorbid developmental conditions.

Since the Bhutanese health care system is public and free for all Bhutanese citizens, the wealth gradient is not anticipated to be a major impact except for out of pocket costs to reach the hospital. These journeys to the main hospital can be costly in a mountainous country with difficult terrain but are culturally normal there.

"2. My second comment is about the use of the word screening. It is commonplace to refer to developmental screening, but no neurodevelopmental assessments have to date met standard WHO screening criteria - largely because there is little evidence that early intervention is any better than late intervention for most problems in most of the tested domains. Screening is most appropriately used for one-off assessments conducted in the context of an organised programme: 'surveillance' more closely reflects what is being done here. The authors might be interested in the discussion of this matter in a recent paper comparing child developmental assessments in five countries (Wilson *et al.*, 2018¹)."

Thank you for this observation. Will change screening to surveillance where appropriate.

More minor points:

- I found Table 2 a little difficult to follow. The use of parentheses is confusing and the final column on the two-stage study does not make sense until the reader reaches the Discussion section. *Removed some extraneous parentheses around sample sizes.*
- Ages would be best expressed in Months in both Table 2 and Table 3, and probably elsewhere in the text. *Changed ages to months in table 2 and 3.*
- The legend to Figure 1 requires more information including total n. *Added total n to y-axis.*
- The discussion about language assessment on page 6 should perhaps reference the literature on bilingualism (I assume that at least some of the children fall into this category). Young children reared in bilingual households commonly appear to have

delayed language acquisition in each language but eventually achieve fluency in both.

Thank you for this helpful observation. Added text on bilingualism in children to beginning of discussion.

- There are some minor typographical issues, such as in the first sentence of the Results section of the Abstract and the second sentence of the Strengths and Limitations section.

These have been fixed. Thank you.

Competing Interests: None.