Comparison of the ASQ and PEDS in Screening for Developmental Delay in Children Presenting for Primary Care

Marjolaine M. Limbos, PhD,* David P. Joyce, MD†

ABSTRACT: *Objectives:* This study investigated the sensitivity and specificity of two brief, parent-completed developmental screening measures—the Ages and Stages Questionnaire (ASQ) and the Parents' Evaluation of Developmental Status (PEDS)—in children presenting to their primary care providers. *Method:* A sample of 334 children aged 12 to 60 months was recruited. Parents completed the PEDS and the ASQ in their home or the primary care clinic of one of the investigators. The presence of ≥1 predictive concerns or abnormal domains was considered a positive screen. All children underwent evaluation (administered by a psychologist) with the following criterion measures: the Bayley Scales of Infant Development—Third Edition or the Wechsler Pre-school and Primary Scale of Intelligence—Third Edition, the Preschool Language Scale—Fourth Edition, and the Vineland Adaptive Behavior Scales—Second Edition. *Results:* The mean age of children was 32.3 months. Developmental delay was identified in 34 children (10%). The PEDS had moderate sensitivity (74%) but low specificity (64%); comparatively, the ASQ had significantly higher sensitivity (82%) and specificity (78%). The ASQ had moderate sensitivity and specificity across age subgroups, whereas the PEDS had either low sensitivity or specificity in each of the age subgroups, except for the ≤30 month group, where there was moderate sensitivity (78%) and specificity (75%). Using ≥2 predictive concerns on the PEDS or ≥2 abnormal domains on the ASQ significantly improved specificity of both tests (89% and 94%, respectively) but resulted in very low sensitivity (41% and 47%, respectively). *Conclusions:* These findings support the guidelines of the American Academy of Pediatrics, demonstrating that both the ASQ and, to a lesser extent, the PEDS have reasonable test characteristics for developmental screening in primary care settings. Although the ASQ seems to have higher sensitivity and specificity across a variety of age groups, the choice of which measure to use should be determined by the practice setting, population served, and preference of the physician.

(J Dev Behav Pediatr 32:000–000, 2011) **Index terms:** developmental delay, screening, primary care, well-child visit.

Developmental delays are common, affecting up to 15% of children.[1,2] Recognizing developmental delay is difficult, especially during early childhood when there is a large variation in the ages at which children achieve developmental milestones. During the first few years of life, neural pathways required for optimal development are being formed, and brain development in these early years sets up the basis for future learning, behavior, and health.[3] The brain's capacity for higher level functions builds on this platform, starting with basic processes and moving to more complex ones.[4] Many critical periods of development conclude by the age of 6 years; if problems are not identified within these initial years of development, opportunities for intervention may be missed.

Early recognition of developmental delay facilitates the implementation of prevention and intervention pro-

grams and results in improvements in cognitive, behavioral, academic, and adaptive functioning.[5-7] It has been shown that children with developmental delay are more likely than their peers to have emotional, behavioral, and health problems in later life, although the extent to which this is related to earlier undiagnosed problems in early childhood has yet to be determined.[3,8] Furthermore, it has been shown that such individuals incur higher future costs related to development-related treatment, as well as health problems, use of the welfare system, and crime.[9]

Studies of early intervention programs for children identified with developmental problems have demonstrated success at alleviating much of the morbidity associated with developmental disorders.[5-7,10-12] For children with physical handicaps or marked cognitive impairments, preschool intervention programs have led to improvements in social, communication, and self-help skills and have provided parents with support and guidance.[13-16] Likewise, for children who are at high risk of school failure because of environmental risk factors such as poverty, preschool programs have also been shown to be particularly beneficial.[14,17-21] Children identified early with developmental problems who have participated in such programs are subsequently less likely to repeat

grades, require special services, and drop out of school early. Early identification can also improve a child's outcome by enabling the family to develop the strategies and obtain the resources for successful family functioning.[22]

Although it is known that early childhood intervention improves outcomes, there may be challenges in linking families to the necessary resources, such as early childhood settings and specialized services. However, most families with preschool children routinely access primary care providers, such as family physicians or pediatricians for common child health services (e.g., vaccinations). This situation provides a unique opportunity for perhaps the only contact that many families will have with a health professional. Regular contact with a primary care clinician provides an opportunity to engage parents in ongoing discussions about child development and services available in the community, to assess for developmental delays, and to help identify and initiate early treatment for children with developmental delay.[8,23,24]

In response to evidence linking development delay to adverse long-term outcomes, there has been a renewed focus on the importance of developmental screening in primary care settings, particularly for preschool children.[3,5,8,9,23] This ideology is reflected in recent guidelines developed by the American Academy of Pediatrics (AAP) and other authorities, who recommend that physicians institute a program of regular developmental screening for young children using standardized screening tools at the ages of 9, 18, and 30 months.[8,23,25] Despite these recommendations, there remains a significant disparity between guidelines and actual practice. Indeed, surveys have demonstrated that a minority of physicians perform routine screening using standardized tools, with the result that many children with developmental problems are not identified until they reach school age.[26-32] Although there are many possible contributing factors to the issue of lack of early identification, lack of screening by physicians may result from several factors including inadequate time or remuneration for screening processes, conflicting reports on the accuracy of available screening tools, and the paucity of research that has been conducted in primary care settings.[30,32-35]

The Parents' Evaluation of Developmental Status (PEDS) and Ages and Stages Questionnaire (ASQ) are parent-completed developmental screening tools that are increasingly being recommended for use in clinical practice.[25,30,33,35-38] The popularity of these tools is due to several favorable qualities that make them suitable for incorporation into busy practices, including being parent-completed, requiring little physician time, ease of administration and interpretation, and low cost.[39,40] Although both screening tools have been studied separately in standardization samples, to date there has been little research directly comparing the accuracies of developmental screening tools in primary care samples. Rydz et al[35] studied a sample of children presenting to their pediatrician for an 18-month well-child visit and assigned children to be screened with either the ASQ or

the Child Development Inventory (CDI). A single developmental inventory (Battelle Developmental Inventory) was used as the criterion measure and was only administered to those who tested positive on the screening test. Their findings suggested that both measures had low sensitivity (67% and 50%, respectively) and that only the CDI had acceptable specificity (86% vs 39% for ASQ). The interpretation of these findings is significantly limited by failure to administer the criterion measure to all children (i.e., making determination sensitivity and specificity problematic) and the significant lag time between the administration of the screening test and criterion measure.

Recently, Sices et al[33] compared the degree of agreement between the PEDS and the ASQ in a sample of preschool children. Results indicated that the PEDS and ASQ had only fair overall agreement, with discordant classification of children occurring one-third of the time. Although the tests were compared directly in the same cohort of children, there was no criterion measure included for developmental delay. Although informative in describing differences between the 2 tests, the study by Sices et al was limited by the absence of a criterion measure, making it impossible to determine the utility of either measure in definitively classifying children as developmentally delayed.

Although there is now substantial evidence demonstrating the accuracy of both the ASQ and PEDS separately in research settings, generalization of the findings to primary care settings is limited by several factors including (1) a scarcity of studies in primary care settings,[41-45] (2) conflicting results for those studies that have used primary care samples,[33,35] (3) failure to administer an adequate criterion measure or failure to administer a criterion measure to all children,[33,35] (4) small sample sizes, and (5) failure to compare both screening tools within the same sample.[46,47] Furthermore, as the PEDS and ASQ are increasingly incorporated into day-to-day clinical practice, many other questions are emerging, including which cutoff point to use on the screening tests, whether prescreening with the PEDS is advantageous, and whether these tools accurately classify children with developmental delay in the age groups for which routine screening has been recommended.[33,35,40,43,48]

The current study set out to compare the sensitivity and specificity of the ASQ and the PEDS in preschool children presenting to their primary care providers for routine care. The goal was to examine the performance of these tools in the primary care setting—for which they are being recommended for use—using a large unselected sample and administering both screening tools and a clinically relevant criterion measure to all children.

PATIENTS AND METHODS
Study Procedures

The current research is part of a larger study examining the accuracy of a number of developmental screening mea-

sures, risk factors for developmental delay, and physician factors associated with screening in primary care.

A convenience sample of children, aged 12 to 60 months, who presented to their primary care provider for routine care were considered eligible for participation. Children were recruited using several methods: (1) direct recruitment by a research assistant from waiting rooms of participating clinics; (2) direct recruitment by receptionists who were trained and provided with a written script for recruitment of children at the time that they checked in for a routine visit; (3) self-referral by parents who had read a poster announcing the study in waiting rooms. Participating parents were given an appointment to meet in their homes or at the primary care clinic of one of the investigators (D.P.J.). Recruitment began in December 2007 and was completed in September 2008. The study was approved by the Research Ethics Board of Laurentian University, and all parents gave written informed consent prior to participation in the study.

On the assessment day, a registered child psychologist administered the criterion measures to the child, while parents completed the Ages and Stages Questionnaire (ASQ) and the Parents' Evaluation of Developmental Status (PEDS) in a separate room. Parents and the psychologist were blinded to the results of each others' testing. The instruction manuals for the ASQ and PEDS were used for administration and scoring.[41,42,44,45] Order of administration of the screening tools was alternated in the research packages at the onset of the study. Following the psychological assessment, parents were given time with the child to complete activities on the ASQ if they were unsure of their response. If necessary, they were provided with items (e.g., toys and scissors) required to complete items. The testing process concluded with the psychologist engaging the parent in a structured Vineland Adaptive Behavior Scales—Second Edition (VABS-II) interview.

The PEDS and ASQ were scored by a family physician (D.P.J.), and the psychologist independently scored all of the criterion measures. Each was blinded to the results of the other's scoring. A summary of the results of the testing was mailed to the child's primary care provider, with instructions given for appropriate follow-up.

Information on medical and developmental history was obtained by a research assistant who reviewed the child's medical record. Parents also completed a non-standardized demographic questionnaire that included their self-reported ethnic, language, and socioeconomic background, and a summary of the child's medical, developmental, and family history. The child's age, gender, and language of testing were recorded by the psychologist on the date of the assessment.

Participants
Children and Parents
A total of 462 children aged 12 to 60 months were approached to participate in the study. Children were excluded if chart review or the demographic questionnaire revealed a history of (1) developmental delay or disability, (2) participation in an intervention program for developmental delay, (3) psychiatric disorder, or (4) birth <36 weeks gestation. The screening tools and demographic questionnaire were completed by the child's primary caregiver, usually the child's biological mother. Other primary caregivers who completed the tools included father (n = 4), maternal grandmother (n = 3), and adoptive mother (n = 11). Testing was conducted in the participant's home for all except 11 children, who underwent testing in the primary care office of one of the investigators (D.P.J.). The study was conducted in communities in northern Ontario, Canada, approximately 4 hours drive from the city of Toronto. This area of Canada has a significant proportion of individuals who are bilingual in both of Canada's official languages (English and French). Although some parents in this study reported speaking both French and English in the home, all participating parents reported being fluent in written and spoken English and further demonstrated this in their interaction with the research team during the recruitment and consent process. Participants were said to come from an urban community if they resided in the cities of Sudbury (population: 157,857) or North Bay (population: 503,966), Ontario. Residents in smaller towns, within a 1.5-hour drive of these cities, were classified as being nonurban.

Primary Care Providers
Children were recruited from the offices of 80 community-based, primary care providers. Seventy providers were family physicians, 7 were nurse practitioners, and 3 were pediatricians. Of these primary care providers, 22 were actively involved in a larger study on developmental screening, while the remaining providers simply allowed recruitment of children from their offices.

Measures
Developmental Screening Measures
Ages and Stages Questionnaire—Second Edition
The ASQ[42,45] is a brief (15 min) measure, in which parents rate their child's current skills and development. Parents answer 30 questions covering 5 domains of development including communication, gross motor, fine motor, problem-solving, and adaptive skills. Parents are instructed to try activities with their child to facilitate assessment. Materials required to complete certain sections of the ASQ (e.g., ball and toy) were made available to parents, in the event that the items were not readily found in the home. A pass/fail score was assigned for each area of development. The form closest to the child's chronological age was used. For each domain, a child who scored within the shaded portion of the scoring bar graph (which represents <2 SDs below the mean) would be classified as having failed the domain. The presence of any failed domain of the ASQ was considered a positive screen. The ASQ has been validated in large, standardized samples of children from

diverse ethnic and socioeconomic backgrounds and has a moderate sensitivity (70–90%) and specificity (76–91%).[33,39,42,45] The English version of the test was administered in all instances.

Parents' Evaluation of Developmental Status

The PEDS[41,44] is a brief, validated, 1-page developmental screening tool to evaluate children ranging in age from birth to 8 years of age. The tool elicits parent concerns on 10 items across 9 domains of development or behavior and takes 2 to 5 minutes to complete. Response options include *yes*, *no*, or *a little*. Unlike the ASQ, this measure does not inquire about a child's ability to perform specific tasks, and parents are not required to engage their child in any activities. The presence of ≥1 predictive/significant concern(s) was considered a positive screen for developmental delay. The PEDS has a reported moderate sensitivity (74–79%) and specificity (70–80%).[25,33,40,41,49] The English version of the test was administered in all instances.

Criterion Measures for Identification of Developmental Delay

The criterion measures for developmental delay in children consisted of a battery of psychological tests of adaptive, cognitive, motor, developmental, and language functioning. Although no battery of tests can prove unequivocally that a child has a developmental delay, current practice involves the use of standardized instruments administered by a psychologist. The battery of measures used in the current study were previously widely used in clinical practice and in prior research on developmental delay screening, and it is considered to have excellent reliability and validity.[50,51] All children underwent testing with 3 criterion measures: (1) a measure of adaptive functioning—the VABS-II, (2) a measure of speech/language functioning—the Preschool Language Scale (PLS-IV), and (3) a measure of cognitive functioning— either the (a) Bayley Scales of Infant Development—Third Edition (BSID-III) (for children younger than 30 mo) or (b) Wechsler Preschool and Primary Scale of Intelligence—Third Edition (WPPSI-III) (for children 30 mo and older). The testing psychologist was fluent in both English and French.

Bayley Scales of Infant Development—Third Edition

The BSID-III[52] is a measure of development, for children aged 1 to 42 months. There are 5 subscales: cognitive, language, motor, social-emotional, and adaptive behavior. To avoid overlap with the other tests (see below), the language and adaptive behavior scales were not administered. The BSID-III has excellent reliability and validity and is considered by many to be one of the best measures of infant development.[50-53] The BSID was used as the primary measure of cognitive and motor functioning in children younger than 30 months, as well as for those children older than 30 months who did not reach basal levels on the WPPSI-III. The BSID-III is not available in the French language and as such the language subscale was not used in this study. The majority of the items on the cognitive and motor subscales of the

BSID-III are nonverbal and involve observation of the child in response to a stimulus. The psychologist administering the test was fluent in French and was able to communicate adequately with the child when presenting the stimuli and observing responses.

Wechsler Preschool and Primary Scale of Intelligence—Third Edition

The WPPSI-III[54] is one of the most widely used tests of cognitive functioning for children 30 months of age and older. It is composed of 15 tests that are divided into 4 core tests and 1 supplemental test. Additional core subtests are administered to children older than 4 years. The WPPSI-III provides composite scores of specific domains of intellectual functioning: Verbal Intelligence Quotient (IQ) and Performance IQ, as well as a Full Scale IQ, which represents a child's overall intellectual ability. There are several decades of research to support the clinical utility of the scale, and its correlation with a large number of other scales of intelligence and development supports its validity.[54-57] The French language version of the WPPSI-III was used when French was the language most often spoken at home by the child.

Vineland Adaptive Behavior Scales—Second Edition

The VABS-II[58] is a 297-item, 30-minute, caregiver interview that assesses the adaptive functioning of children from birth to the age of 19 years. Adaptive behavior is measured in 4 domains: communication, daily living skills, socialization, and motor skills. An overall Adaptive Behavior Composite is calculated. The VABS-II was standardized using 3000 individuals, has demonstrated reliability and validity, and has been used repeatedly to validate developmental screening measures.[57-59] The English version of the VABS-II was used in all instances.

Preschool Language Scale—Fourth Edition

The PLS-IV[60,61] is a diagnostic instrument for receptive and expressive language disorders in children aged 2 weeks through 6 years. It consists of 8 receptive/expressive language tasks for each 6-month interval for children between the ages of 2 weeks and 4 years 11 months and 8 tasks for each 12-month interval for children aged 5 and 6 years. Administration of the test takes ~15 to 40 minutes. The scale has been standardized and validated, has demonstrated internal consistency and interrater reliability, and is used extensively in clinical and research settings to examine language development in preschool children.[60,61] The French language version of the PLS-IV was used when French was the language most often spoken at home by the child.

Definition of Developmental Delay

For the purpose of this study, a child was classified as having a developmental delay if he/she scored below the 10th percentile on any of the criterion measures (i.e., Cognitive, Motor, or Social-Emotional Scale Composite Scores of the BSID-III; Verbal IQ, Performance IQ, or Full Scale IQ scores on the WPPSI-III; or Auditory Comprehension, Expressive Communication, or Total Language on the PLS-IV) and had a concurrent score below the

10th percentile on the Adaptive Behavior Composite score of the VABS-II. This criterion is consistent with standards used in research and clinical practice for the identification of developmental delay and intellectual disability.[50,51,56,59,62,63] Because the definition of developmental delay or disability varies, particularly with respect to defining those children who can access services, secondary analyses using more stringent cutoffs for developmental delay were performed. These included children with a score of ≤1.5 and ≤2.0 SDs below the mean, respectively, on any one of the criterion measures (as outlined above) and concurrently on the VABS-II.

Data Analysis

A sample size calculation was performed using a variation of the log odds ratio test.[64] The sample size required to demonstrate a difference of 10% in sensitivity or specificity was 210 participants. Differences in characteristics between groups were examined using t tests for continuous variables and χ^2 tests for categorical variables. To compare sensitivity and specificity of various screening methods, the procedures outlined by Hawass[65] were followed. Matched sample tables were prepared comparing the PEDS and ASQ results, for children with (for comparison of sensitivity) and without (for comparison of specificity) developmental delay. McNemar's χ^2 test, with Yates correction, was then calculated for each matched sample.[65] The DAG_Stat spreadsheet and SPSS statistical software were used for statistical analyses.[66] Cohen's kappa was used to determine the level of agreement between screening tests. Sensitivity, specificity, positive and negative predictive values, and likelihood ratios were calculated, expressed as percentages with 95% confidence intervals. For sensitivity and specificity, values <70% were classified as low, 70% to 85% as moderate, and >85% as high.[67] To examine the tests' performance at different ages, children were divided into subgroups according to the ages recommended for screening by the American Academy of Pediatrics (AAP), namely, 18 months and 30 months. Sensitivity and specificity were calculated for the following age subgroups: 18 months and younger, 19 months and older, 30 months and younger, and 31 months and older. Secondary calculations of sensitivity and specificity using the presence of ≥2 predictive concerns on the PEDS, or ≥2 abnormal domains on the ASQ, as the cutoff for positive screening, were also performed. All results were considered significant at $p < .05$.

RESULTS

A total of 334 children, aged 12 to 60 months, participated in the study. This represented 72% of children who were initially approached to participate. Six children had incomplete data on either the Parents' Evaluation of Developmental Status (PEDS) or Ages and Stages Questionnaire (ASQ), and data for those tests were not included in the analysis. In total, 331 children completed the ASQ, 331 completed the PEDS, and 328 completed both the ASQ and the PEDS. There were 4 reasons for exclusion in the study: (1) parents not interested in the study (n = 58), (2) inability to attend assessment (n = 32), (3) outside of age range for inclusion (n = 24), or (4) prior diagnosis of developmental delay (n = 14).

The mean age of the children was 32.3 ± 16.3 months, and there was no significant difference between the age of those with (33.1 ± 16.1 mo) and without developmental delay (32.2 ± 16.5 mo, $p = .74$). Table 1 summarizes the remainder of the demographic variables. Male children from lower income families were more likely to be identified as having a developmental delay. There was no difference in the mean age of mothers of children with (29.9 ± 7.2 mo) and without (30.5 ± 6.3 mo) developmental delay ($p = .63$).

Children with Developmental Delay

Using the criterion measures, 34 children (10%) were identified as having a developmental delay. The primary domain of developmental delay was isolated cognitive delay in 15 children, speech language delay in 12 children, and motor delay in 2 children. Three children had mixed cognitive, motor, and speech language delays; 1 child had mixed cognitive and speech language delays; and 1 child had mixed motor and speech language delays.

Agreement Between ASQ and PEDS

To understand the test characteristics of the PEDS and ASQ and to calculate differences in sensitivity and specificity of the 2 groups using McNemar's test, the degree of agreement in classifying children with and without developmental delay was examined. Table 2 shows 2 × 2 matched sample tables for the ASQ and PEDS for children with and without developmental delay. Overall, there was agreement in the classification of 206 children (63%). More specifically, there was fair agreement in the classification of 25 (74%) children with developmental delay ($\kappa = 0.38$). However, there was little agreement between the PEDS and ASQ in classifying children without developmental delay (n = 181, 62%; $\kappa = 0.02$).

Sensitivity and Specificity of the ASQ and PEDS

Table 3 summarizes the results on the accuracy of the ASQ and PEDS. The ASQ had moderate sensitivity (82%) and specificity (78%) in screening for developmental delay. The PEDS had moderate sensitivity (74%) but low specificity (64%) in screening for developmental delay. There was a significant difference in both the sensitivity and specificity of the ASQ and PEDS. Significant differences were found in the sensitivity or specificity of the following test pairs: PEDS ≥1 predictive concern and ASQ ≥1 failed domain ($\chi^2 = 4.3$, $p = .04$, for sensitivity and $\chi^2 = 4.0$, $p = .04$, for specificity), PEDS ≥1 predictive concern and PEDS ≥2 predictive concerns ($\chi^2 = 8.1$, $p = .00$, for sensitivity and $\chi^2 = 75$, $p = .00$, for specificity), and ASQ ≥1 failed domain and ASQ ≥2 failed domains ($\chi^2 = 10.1$, $p = .00$, for sensitivity and $\chi^2 = 47.0$, $p = .00$, for specificity).

 5

Table 1. Characteristics of Participating Children Identified With and Without Developmental Delay[a]

	All Children, n (%)	Developmental Delay, n (%)	No Developmental Delay, n (%)	Comparison,[b] χ^2, p Value
Child age (mo)				
12–18	90 (27)	8 (24)	82 (27)	.63
18–24	47 (14)	3 (9)	44 (15)	
24–36	75 (22)	10 (29)	65 (22)	
>36	122 (37)	13 (38)	109 (36)	
Gender				
Male	187 (56)	25 (73)	162 (54)	.03
Female	147 (44)	9 (27)	138 (46)	
Mother's race/ethnicity				
White	284 (85)	29 (85)	255 (85)	.79
Black	1 (1)	0	1 (1)	
Aboriginal[c]	42 (12)	5 (15)	37 (12)	
Other	7 (2)	0	7 (2)	
Father's race/ethnicity				
White	270 (81)	25 (74)	245 (82)	.15
Black	9 (3)	0	9 (3)	
Aboriginal	41 (12)	8 (23)	33 (11)	
Other	14 (4)	1 (3)	13 (4)	
Language of child testing				
English	254 (76)	23 (68)	231 (77)	.23
French	80 (24)	11 (32)	69 (23)	
Maternal age (y)				
<20	27 (8)	0	27 (9)	.05
21–29	147 (44)	20 (59)	127 (42)	
30–39	138 (41)	10 (29)	128 (43)	
≥40	22 (7)	4 (12)	18 (6)	
Maternal education				
Less than high school	60 (18)	9 (27)	51 (17)	.15
High school	59 (18)	7 (21)	52 (17)	
More than high school	196 (59)	14 (41)	182 (61)	
Unknown	19 (5)	4 (11)	15 (5)	
Annual household income[d]				
<15,000	42 (13)	10 (29)	32 (11)	.00
15,000–30,000	54 (16)	10 (29)	44 (15)	
30,000–60,000	73 (22)	5 (15)	68 (23)	
>60,000	131 (39)	4 (12)	127 (42)	
Refused	34 (10)	5 (15)	29 (9)	
Community type[e]				
Urban	198 (59)	25 (74)	173 (58)	.07
Nonurban	136 (41)	9 (26)	127 (42)	

[a]Developmental delay refers to children with a score of <10th percentile on adaptive function and one other criterion measure. [b]Comparisons were made using the Pearson χ^2 test. [c]Aboriginal refers to Canadian descendants of the original inhabitants of North America. [d]Household income is expressed in Canadian dollars. [e]Urban location refers to cities in northern Ontario. Nonurban locations were areas within 1.5 hours of the urban locations.

To address concerns regarding administration of the English version of the ASQ and PEDS to bilingual parents, we recalculated sensitivity and specificity for caregivers who reported English as the only language spoken at home. The analyses found little evidence to suggest that the screening tests performed better in English-only families: the sensitivity (83%) and specificity (76%) of the ASQ were virtually identical in this subgroup compared with the total

Table 2. Two × Two Matched Sample Tables for PEDS and ASQ for Children With and Without Developmental Delay[a]

ASQ	No Developmental Delay (N = 294)[b] PEDS		Developmental Delay (N = 34) PEDS	
	≥1 Predictive Concerns[c]	No Concern	≥1 Predictive Concerns	No Concern
≥1 Failed domains[c]	21	45	20	8
No concern	68	160	1	5
	$\kappa = 0.02$		$\kappa = 0.38$	
	McNemar's $\chi^2 = 4.0$, $p = .04$		McNemar's $\chi^2 = 4.3$, $p = .04$	
	≥2 Predictive Concerns	No Concern	≥2 Predictive Concerns	No Concern
≥2 Failed domains	1	16	6	10
No concern	11	266	5	13
	$\kappa = 0.02$		$\kappa = 0.10$	
	McNemar's $\chi^2 = 0.6$, $p = .44$[d]		McNemar's $\chi^2 = 1.1$, $p = .30$	

PEDS, Parents' Evaluation of Developmental Status; ASQ, Ages and Stages Questionnaire. [a]Developmental delay refers to children with a score of <10th percentile on adaptive function and one other criterion measure. [b]Three hundred twenty-eight (294 without developmental delay and 34 with developmental delay) children were administered both the PEDS and ASQ. [c]The presence of ≥1 predictive concern on the PEDS or ≥1 failed domain on the ASQ was used as the cut-off for an abnormal screen. [d]Refers to the McNemar's test with Yates correction, using 1 degree of freedom. Matched pair tables for children with no developmental delay were used to determine difference between specificities of the tests, whereas matched pair tables for those with developmental delay were used to determine difference between sensitivities of the tests.[65]

group, and there were only minor differences in the sensitivity (70%) and specificity (68%) of the PEDS.

Sensitivity and Specificity According to Age Subgroup

Figure 1 summarizes subgroup analysis of sensitivity and specificity of the ASQ and PEDS, according to the following age subgroups: (1) ≤18 months (n = 89), (2) >18 months (n = 242), (3) ≤30 months (n = 181 for ASQ, n = 182 for PEDS), and (4) >30 months (n = 150 for ASQ, n = 149 for PEDS). The ASQ had a consistent, moderate sensitivity (75%, 85%, 89%, and 75%, respectively) and specificity (78%, 78%, 74%, and 83%, respectively) across age groups.

Table 3. Two × Two Contingency Tables for PEDS and ASQ Using Various Screening Cutoff Points

Screening Test Used	Screening Result	Criterion Measure		Sensitivity/ Specificity	95% CI
		DD[a] Absent	DD[a] Present		
PEDS	≥1 Predictive concern	106	25	Sens: 0.74	0.56–0.87
	No concern	191	9	Spec: 0.64	0.59–0.70
				PPV: 0.19	0.13–0.27
				NPV: 0.96	0.92–0.98
				LR+: 2.06	1.60–2.65
	≥2 Predictive concerns	32	14	Sens: 0.41	0.25–0.59
	No concern	265	20	Spec: 0.89	0.85–0.93
				PPV: 0.30	0.18–0.46
				NPV: 0.93	0.89–0.96
				LR+: 3.82	2.28–6.42
ASQ	≥1 Failed domains	66	28	Sens: 0.82	0.65–0.93
	No concern	231	6	Spec: 0.78	0.73–0.83
				PPV: 0.30	0.21–0.40
				NPV: 0.97	0.95–0.99
				LR+: 3.70	2.85–4.82
	≥2 Failed domains	17	16	Sens: 0.47	0.30–0.65
	No concern	280	18	Spec: 0.94	0.91–0.97
				PPV: 0.48	0.31–0.66
				NPV: 0.94	0.91–0.96
				LR+: 8.22	4.59–14.73

PEDS, Parents' Evaluation of Developmental Status; ASQ, Ages and Stages Questionnaire; CI, confidence interval; Sens, sensitivity; Spec, specificity; PPV, positive predictive value; NVP, negative predictive value; LR+, positive likelihood ratio. [a]Developmental delay based on a score of <10th percentile on adaptive function and one other criterion measure.
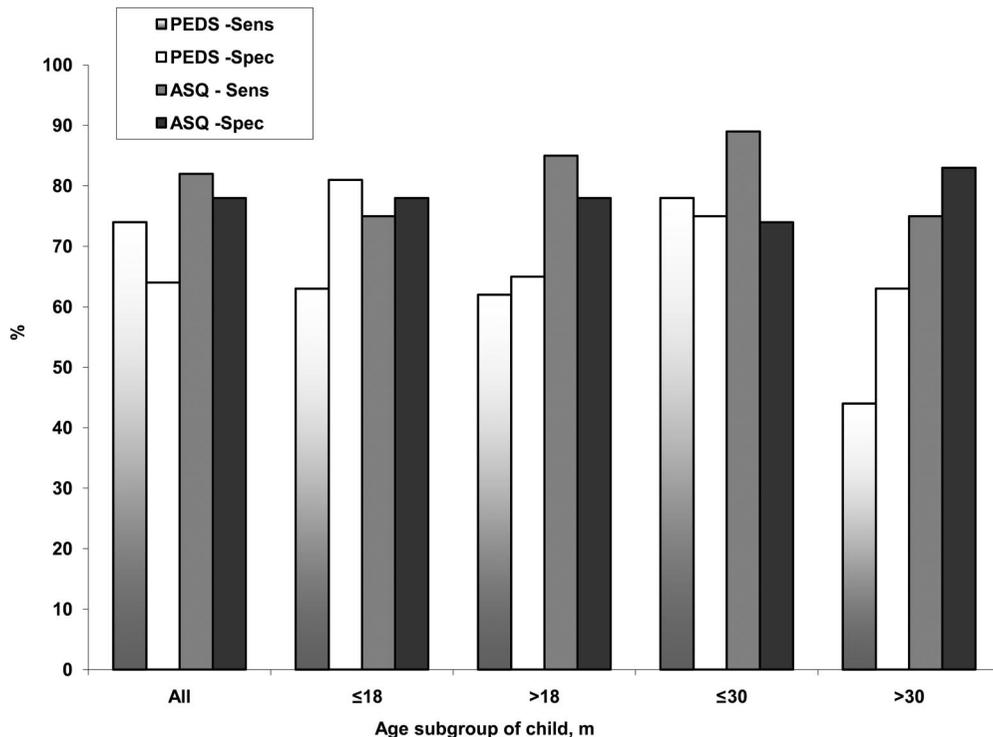
Figure 1. Sensitivity and specificity of the ASQ and PEDS according to age subgroup. Subgroups analyses included the following: (1) All = all children studied (N = 331 for ASQ and PEDS, respectively); (2) ≤18 = children ≤18 months (N = 89 for PEDS and ASQ, respectively); (3) >18 = children 19 months and older (N = 242 for ASQ and PEDS, respectively); (4) ≤30 = children ≤30 months (N = 181 for ASQ, N = 182 for PEDS); and (5) >30 = children 31 months and older (N = 150 for ASQ, N = 149 for PEDS). PEDS, Parents' Evaluation of Developmental Status; Sens, sensitivity; Spec, specificity; ASQ, Ages and Stages Questionnaire.

Results for the PEDS were more variable, showing moderate sensitivity (78%) and specificity (75%) in the ≤30 month subgroup but either a low sensitivity or low specificity in all other age groups: (1) ≤18 months (62% and 81%, respectively), (2) >18 months (62% and 65%, respectively), and (3) >30 months (44% and 63%). Statistically significant differences were found in both the sensitivity ($\chi^2 = 4.2$, $p = .04$) and specificity ($\chi^2 = 7.3$, $p = .01$) of the 2 tests in the >18 months age group, and in the specificity of the tests in the >30 months age group ($\chi^2 = 11.8$, $p = .00$).

Evaluation of PEDS Prescreening with ASQ Follow-Up Testing

The use of the PEDS as a prescreening test has been proposed, with administration of the ASQ only for children who have a predictive/significant concern on the PEDS.[40] The data were examined, post hoc, to determine the psychometric properties of such an approach. Children were classified as having screened negative if they had no predictive concerns on the PEDS or no abnormal domains on the ASQ. Children with ≥1 predictive concerns on the PEDS and ≥1 abnormal domains on the ASQ were classified as meeting the criteria for a positive screen. Using this method had high specificity (90%) but low sensitivity (58%).

Sensitivity and Specificity Using ≥2 Abnormal Domains or Predictive Concerns

Because some clinicians do not refer children for assessment unless they have ≥2 abnormal domains on the ASQ or ≥2 predictive concerns on the PEDS, the test characteristics of the PEDS and ASQ using these alternate cutoff points were examined.[33] Table 3 indicates that the use of these cutoff points significantly improved specificity but sensitivity fell to below acceptable levels for both tests. There was a significant difference between the sensitivity and specificity using the ≥2 compared with the ≥1 abnormal cutoff for each test. However, there was no significant difference between the sensitivity or the specificity of the ASQ compared with the PEDS using this cutoff.

Alternate Criterion Measure Threshold for Classification of Developmental Delay

A total of 14 children were classified as having a developmental delay using the ≤1.5 SD threshold, and only 3 children were classified using the ≤2 SD cutoff. Figure 2 illustrates that at the ≤1.5 SD cutoff, the sensitivity and specificity remained stable for both the ASQ (85% and 75%, respectively) and the PEDS (78% and 68%, respectively). At the ≤2 SD cutoff, sensitivity of both screening tests rose to 100%, whereas specificity remained in the moderate range for the ASQ (72%) and low range for the PEDS (67%).

DISCUSSION

This study adds to the growing literature supporting the use of the Ages and Stages Questionnaire (ASQ) and Parents' Evaluation of Developmental Status (PEDS) for
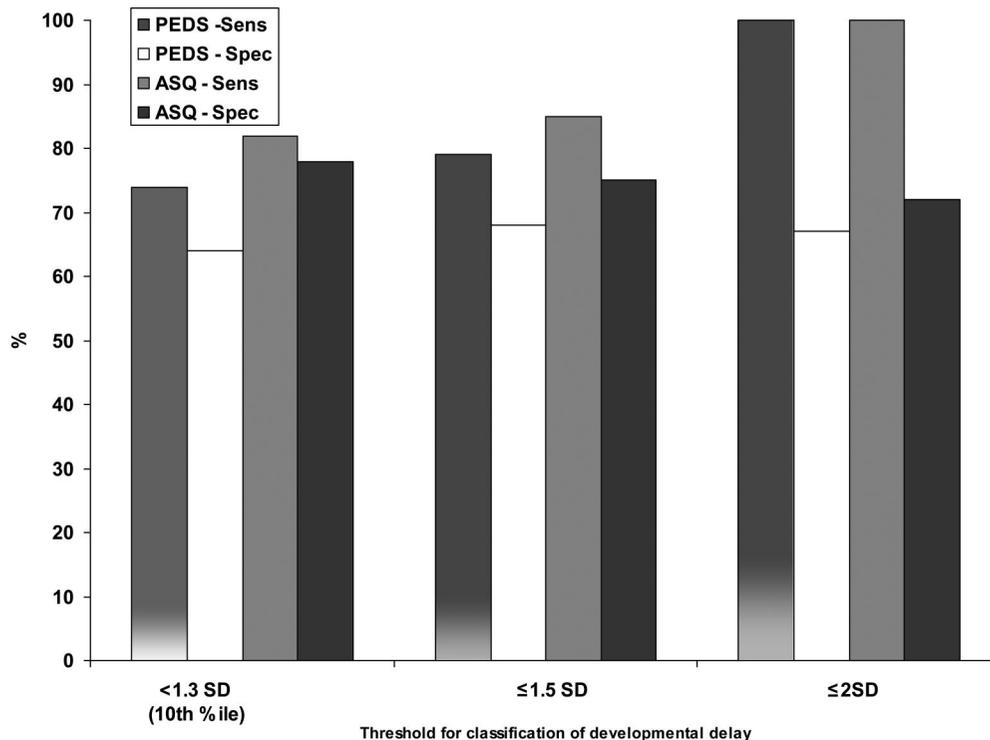
Figure 2. Sensitivity and specificity of the ASQ and PEDS according to threshold for classification of developmental delay. The threshold for classification of a child as having a developmental delay included the following: (1) <1.3 SD = 1.3 SDs below the mean (<10th percentile) on any of the criterion measures and on the Vineland Adaptive Behavior Scales–Second Edition (VABS II); (2) ≤1.5 SD = ≤1.5 SDs below the mean on any of the criterion measures and on the VABS II; and (3) ≤2 SD = ≤2 SDs below the mean on any of the criterion measures and on the VABS II. PEDS, Parents' Evaluation of Developmental Status; Sens, sensitivity; Spec, specificity; ASQ, Ages and Stages Questionnaire.

developmental screening in primary care. Past research has demonstrated the ease of use, acceptability, and validity of each of these brief screening measures in standardization samples but has not compared the utility of the 2 measures. The current study directly compares the sensitivity and specificity of the ASQ and PEDS in a sample of children presenting to their primary care providers.[35-45] The main results of this study that both the ASQ and PEDS have moderate sensitivity (82% and 74%, respectively) for screening for developmental delays, accompanied by moderate specificity (78%) for the ASQ and low specificity (64%) for the PEDS and are consistent with findings of past research conducted on population or standardization samples.[25,39,41,42,44] Sensitivity and specificity of both tests was found to be relatively stable, regardless of the classification of developmental delay used. The findings support the use of the ASQ and, to a lesser extent, the PEDS, for systematic developmental screening, and are in keeping with the American Academy of Pediatrics (AAP)'s policy on early detection of developmental delay in children.[25]

Our results suggest that the ASQ has superior sensitivity and specificity to the PEDS. Indeed, the higher sensitivity and specificity of the ASQ was not only statistically significant but also appears to be clinically significant. Of particular concern was the finding that the specificity of the PEDS fell below 70%, which is the usual cutoff for an acceptable screening measure.[25,47,59,67] Past research on the accuracy of these measures is variable,

but consistent with our findings, and shows a trend toward higher sensitivity and specificity of the ASQ when compared with the PEDS.[25,33,41-44] Analysis of subgroups of children, based on the AAP's recommended ages for screening, demonstrated moderate sensitivity and specificity of the ASQ across age groups. In contrast, the PEDS showed low specificity in the overall group, and low sensitivity and/or specificity in most age subgroups, with the exception of the ≤30 month age group, where there was moderate sensitivity and specificity. Due to small sample sizes, the analysis of subgroups should be interpreted with caution, but the lack of significant difference in sensitivity and specificity of the screening tests in the younger age groups suggests that both tests may be suitable for screening in children aged 30 months and younger. Given these findings, it is important to consider other test factors that will affect the use of these measures by physicians in clinical practice. Although the PEDS showed comparatively lower accuracy in this study, it has in the past been shown that many health care providers do not administer screening tests routinely, and that eliciting parents' concerns about development are strong predictors of developmental problems, and superior to clinical impression alone.[30,48,49,68,69] Furthermore, some studies have suggested that the PEDS may have some practical advantages over the ASQ, primarily related to shorter administration time and the ability to complete the testing easily in the waiting room of a busy clinic. In some practice settings and patient

9

populations, use of the ASQ could be hindered by several factors: (1) difficulties in timing home administration with their well-child appointment, including ensuring the correct age form is administered; (2) relying on parents to attempt multiple developmental tasks with their child; and (3) relying on parents to return completed forms for scoring.[40] Thus, we conclude that although the ASQ has the superior combination of sensitivity and specificity across a wide range of ages, both the ASQ and PEDS appear to be acceptable for developmental screening in children at the ages suggested by the AAP, namely, those 30 months and younger.

Few studies have directly compared developmental screening measures in primary care samples. A recent study by Rydz et al[35] compared the accuracy of the ASQ and the Child Development Inventory (CDI) in children presenting to their pediatrician for an 18-month well-child visit. Of the 183 children screened with the ASQ, there was both low sensitivity (67%) and specificity (39%) in identifying developmental delay. In contrast to our study, the authors concluded that the ASQ did not meet standards for developmental screening. Although the study by Rydz et al had the advantage of replicating day-to-day practice, by having parents complete the screening tests in the waiting room of their pediatrician's office, there are several limitations which may explain their divergent findings. First, while the current study followed the methods outlined in the instruction manual, the study by Rydz et al used an abridged method of administration of the ASQ. Although the modified method of administration may be practical for use in busy office settings, it may have contributed to lowering the performance of the test. Second, the criterion measure used in the current study was administered on the same day as the screening tests by a clinical psychologist and included a comprehensive, clinically relevant battery of development and adaptive functioning tests. In contrast, the study by Rydz et al used as their gold standard a single developmental inventory administered 3 months after the date of the screening test. This would likely have the effect of misclassifying some children as developmentally delayed when they were not, and vice versa, lowering the measured sensitivity and specificity of the ASQ.[47] Most importantly, however, was the failure to administer the criterion measure to all children. Only those children who tested positive on the ASQ were administered the criterion measure, and as a result, the determination of which children were classified as truly negative or falsely negative could not be adequately determined, making calculation of sensitivity and specificity inaccurate. Consequently, the current study's findings of moderate sensitivity and specificity of the ASQ, across a range of ages, and using a variety of definitions of developmental delay, are likely more representative of the actual performance of the test in clinical practice.

Sices et al[33] also recently compared the performance of the ASQ and PEDS in 60 children, aged 9 to 31 months, presenting to their primary care pediatrician.

No criterion measure for developmental delay was used, but agreement in classification of the PEDS and ASQ was directly compared. Consistent with our study, they found disagreements in classification of one-third of children and only slight to fair agreement between the PEDS and ASQ. Disagreement in classification in the current study was higher among those without a developmental delay (39%) than for those with a developmental delay (27%). The findings raise the issue that while intended to identify a similar group of children at risk for developmental problems, the two tests may identify different children, and this could impact which screening test is selected by physicians for use in clinical practice. However, although an understanding of the agreement between the PEDS and ASQ is informative, it is likely that clinicians will primarily use one measure or the other for developmental screening. As such, it is more informative to determine which test has the superior combination of sensitivity and specificity for identifying children with developmental delay. The current study begins to answer this question and addresses many of the limitations of previous comparative studies of developmental screening tests by using a larger sample size, administering both screening tests and criterion measures to all children, using a comprehensive psychological evaluation as the criterion measure, and including no time lag between administration of criterion and screening measures.[47] As such, this study provides clinicians with valuable information about the ability of the tests to definitively classify children with developmental delay in primary care settings.

The findings of our study shed light on several procedures that have been proposed in clinical practice but to date have not been studied. First, use of the PEDS as a prescreening test has been proposed with only those children having significant concerns going on to testing with the ASQ.[43] Our analysis found that although this method improved specificity substantially, sensitivity was only 59%, suggesting that such a method has a disadvantage over using a single measure for developmental screening. Second, although the manuals for the ASQ and PEDS recommend use of a single abnormal domain or predictive concern as the cutoff for referral, many pediatricians in practice choose not to initiate further assessment unless a child has ≥ 2 abnormal areas.[33,41,42,44,45] However, data used for determining the sensitivity and specificity in validation studies of the PEDS were based on the presence of ≥ 1 predictive concerns, and to date there has been no research to guide the clinical practice of referral when there are ≥ 2 abnormal domains on the ASQ.[44] The findings of the current study strongly support the use of a single domain or predictive concern cutoff point. Although use of a more stringent cutoff improved the specificity of the tests dramatically (89% for PEDS and 94% for ASQ), sensitivity decreased to <50% for both tests and is clearly inadequate for a screening test. Pediatricians in practice should be aware that selecting only those chil-

dren who score abnormal in ≥ 2 areas will result in significant underreferral of children with developmental delay.

There are several limitations to the current study. Although an attempt was made to replicate the real-world clinical application of the PEDS and ASQ, by recruitment of children from primary care offices, administration of screening tests by parents primarily at home, and interpretation of results by a family physician, all aspects were overseen and carried out by members of the research team. Prior studies have already, however, confirmed the potential for implementing similar screening strategies in real-world settings.[35-37] While we used a valid and reliable criterion measure, the assessment took place at one point in time and, ideally, longitudinal assessments would be conducted to confirm that the developmental delay is stable.[47] Although psychological testing of the child was conducted in the language most spoken by the child, only English versions of the PEDS and ASQ were completed by parents. However, all parents reported fluency in English, and subgroup analyses did not raise concerns about performance of the tests in different language groups. Finally, although the sample size was large enough to demonstrate differences between the screening tests, the absolute number of children with developmental delay was relatively small, and a larger sample size would be ideal to give support to the conclusions. In particular, although the majority of children in this study were younger than 3 years, the finding of trends toward improved performance of the PEDS in younger age groups should be replicated in a larger sample of children presenting for their 18- and 30-month developmental screens.

CONCLUSIONS

The findings of this study add support to the guidelines of the American Academy of Pediatrics (AAP) and demonstrate that both the Ages and Stages Questionnaire (ASQ), and to some extent the Parents' Evaluation of Developmental Status (PEDS), have reasonable test characteristics for developmental screening in primary care settings. Although the ASQ seems to have higher sensitivity and specificity, the choice of which measure to use should be determined by the practice setting, population served, and preference of the physician. Guided by the favorable results of the current study, future research should engage practicing primary care physicians in the administration, scoring, interpretation, and follow-up of the ASQ and PEDS, to better understand the performance of these tests in real-world clinical practice settings.

REFERENCES

1. Boyle CA, Decoufle P, Yeargin-Allsopp M. Prevalence and health impact of developmental disabilities in US children. *Pediatrics.* 1994;93:399–403.
2. Yeargin-Allsopp M, Boyle C. Overview: the epidemiology of neurodevelopmental disorders. *Ment Retard Dev Disabil Res Rev.* 2002;8:113–116.
3. McCain M, Mustard F. *Early Years Study.* Toronto, ON: Government of Ontario; 1999.
4. Greenspan S, Shanker S. *The First Idea: How Symbols, Language and Intelligence Evolved from Our Primate Ancestors to Modern Humans.* Cambridge, MA: Da Cap Press; 2004.
5. Anderson LM, Shinn C, Fullilove MT, et al. The effectiveness of early childhood development programs: a systematic review. *Am J Prev Med.* 2003;24(suppl 3):S32–S46.
6. Berlin LJ, Brooks-Gunn J, McCarton C, McCormick MC. The effectiveness of early intervention: examining risk factors and pathways to enhanced development. *Prev Med.* 1998;27:238–245.
7. Hill JL, Brooks-Gunn J, Waldfogel J. Sustained effects of high participation in an early intervention for low-birth-weight premature infants. *Dev Psychol.* 2003;39:730–744.
8. Ontario Children's Health Network/Ontario College of Family Physicians. *Getting it Right at 18 Months … Making it Right for a Lifetime: Report of the Expert Panel on the 18 Month Well Baby Visit.* Toronto, ON; September, 2005.
9. Schweinhart LJ, Montie J, Xizng Z, et al. *Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40.* Ypsilanti, MI: HighScope Press; 2005.
10. Heckman JJ. *Policies to Foster Human Capital.* Working Paper No. 7288. Cambridge, Mass: National Bureau of Economic Research; 1999.
11. Halfon N, Inkela M. Optimizing the health and development of children. *JAMA.* 2003;290:3136–3138.
12. Screening infants and young children for developmental disabilities. American Academy of Pediatrics Committee on Children with Disabilities. *Pediatrics.* 1994;93:863–865.
13. Chaplais JD, Macfarlane JA. A review of 404 late walkers. *Arch Dis Child.* 1984;59:512–516.
14. Mustard JF. *Early Child Development and Experience-based Brain Development: The Scientific Underpinnings of the Importance of Early Child Development in a Globalized World.* Brookings Institute; Toronto, Canada: 2006.
15. Simeonsson RJ, Cooper DH, Scheiner AP. A review and analysis of the effectiveness of early intervention programs. *Pediatrics.* 1982;69:635–641.
16. Zigler E. Project Head Start: success or failure? In: Zigler E, Valentine J, eds. *Project Head Start: A Legacy of the War on Poverty.* New York, NY: The Free Press; 1979:495–507.
17. Palfrey JS, Singer JD, Walker DK, Butler JA. Early identification of children's special needs: a study in five metropolitan communities. *J Pediatr.* 1987;111:651–659.
18. Berrueta-Clement J, Schweinhart W, Barnett S, et al. *Changed Lives: The Effects of the Perry Preschool Program on Youths Through Age 19.* Ypsilanti, MI: The High/Scope Press; 1984.
19. Seitz U, Rosenblum L, Apfel N. Effects of family support intervention: a ten-year follow-up. *Child Dev.* 1985;56:376–91.
20. Shonkoff JP, Hauser-Cram P. Early intervention for disabled infants and their families: a quantitative analysis. *Pediatrics.* 1987;80:650–658.
21. Bennett FC, Guralnick MJ. Effectiveness of developmental intervention in the first five years of life. *Pediatr Clin North Am.* 1991;38:1513–1528.
22. Drillien CM, Pickering RM, Drummond MB. Predictive value of screening for difficult areas of development. *Dev Med Child Neurol.* 1988;30:294–305.

23. Best Start Expert Panel on Early Learning. *Early Learning for Every Child Today: A Framework for Ontario Early Childhood Settings*. Toronto, ON; 2006.

24. Minkovitz CS, Hughart N, Strobino D, et al. A practice-based intervention to enhance quality of care in the first 3 years of life: the Healthy Steps for Young Children Program. *JAMA*. 2003; 290:3081–3091.

25. American Academy of Pediatrics, Council on Children With Disabilities, Section on Developmental Behavioral Pediatrics, Bright Futures Steering Committee, Medical Home Initiatives for Children With Special Needs Project Advisory Committee. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening [Erratum in *Pediatrics*. 2006;118:1808–1809]. *Pediatrics*. 2006;118:405–420.

26. Wiggins LD, Baio J, Rice C. Examination of the time between first evaluation and first autism spectrum diagnosis in a population-based sample. *J Dev Behav Pediatr*. 2006;27(suppl 2):S79–S87.

27. Tomblin JB, Records NL, Buckwalter P, Zhang X, Smith E, O'Brien M. Prevalence of specific language impairment in kindergarten children. *J Speech Lang Hear Res*. 1997;40:1245–1260.

28. Glascoe FP, Shapiro HL. Introduction to developmental and behavioral screening. Available at: www.dbpeds.org/tutorial. Accessed September 2010.

29. Janus M. Measuring community early child development. *CAP Journal*. 2006;14:1–4.

30. Sices L, Feudtner C, McLaughlin J, Drotar D, Williams M. How do primary care physicians identify young children with developmental delays? A national survey. *J Dev Behav Pediatr*. 2003;24:409–417.

31. Sand N, Silverstein M, Glascoe FP, Gupta VB, Tonniges TP, O'Connor KG. Pediatricians' reported practices regarding developmental screening: do guidelines work? Do they help? *Pediatrics*. 2005;116:174–179.

32. Limbos MM, Joyce DP, Roberts GJ. Nipissing District Developmental Screen. Patterns of use by physicians in Ontario. *Can Fam Physician*. 2009;56:e66–e72.

33. Sices L, Stancin T, Kirchner L, Bauchner H. PEDS and ASQ developmental screening tests may not identify the same children. *Pediatrics*. 2009;124:e640–e647.

34. Feightner JW, for the Canadian Task Force on the Periodic Health Examination. Preschool screening for developmental problems. In: Ministry of Supply and Services, eds. *The Canadian Guide to Clinical Preventive Health Care*. Ottawa, Canada: Ministry of Supply and Services; 1994.

35. Rydz D, Srour M, Oskoui M, et al. Screening for developmental delay in the setting of a community pediatric clinic: a prospective assessment of parent-report questionnaires. *Pediatrics*. 2006;118:e1178–e1186.

36. Hix-Small H, Marks K, Squires J, Nickel R. Impact of implementing developmental screening at 12 and 24 months in a pediatric practice. *Pediatrics*. 2007;120:381–389.

37. Earls MF, Hay SS. Setting the stage for success: implementation of developmental and behavioral screening and surveillance in primary care practice—the North Carolina Assuring Better Child Health and Development (ABCD) Project. *Pediatrics*. 2006;118: e183–e188.

38. Schonwald A, Huntington N, Chan E, Risko W, Bridgemohan C. Routine developmental screening implemented in urban primary care settings: more evidence of feasibility and effectiveness. *Pediatrics*. 2009;123:660–668.

39. Squires J, Bricker D, Potter L. Revision of a parent-completed development screening tool: Ages and Stages Questionnaires. *J Pediatr Psychol*. 1997;22:313–328.

40. Brothers KB, Glascoe FP, Robershaw NS. PEDS: developmental milestones—an accurate brief tool for surveillance and screening. *Clin Pediatr (Phila)*. 2008;47:271–279.

41. Glascoe FP. *Parents' Evaluation of Developmental Status*. Nashville, TN: Ellsworth & Vandermeer Press, Ltd; 1997.

42. Bricker D, Squires J. *Ages & Stages Questionnaires: A Parent-Completed, Child-Monitoring System*. 2nd ed. Baltimore, MD: Paul H. Brookes Publishing Co; 1999.

43. Glascoe FP. Evidence-based approach to developmental and behavioural surveillance using parents' concerns. *Child Care Health Dev*. 2000;26:137–149.

44. Glascoe FP. *Collaborating with Parents: Using Parents' Evaluations of Developmental Status to Detect and Address Developmental and Behavioral Problems*. Nashville, TN: Ellsworth & Vandermeer Press LLC; 2002.

45. Squires J, Potter L, Bricker D. *User's Guide for the Ages & Stages Questionnaires: A Parent-Completed, Child-Monitoring System*. 2nd ed. Baltimore, MD: Paul H. Brookes Publishing Co; 1999.

46. Marks K. Should general pediatricians not select the Ages & Stages Questionnaire in light of the Rydz et al study? *Pediatrics*. 2007;120:457–458.

47. Marks K, Glascoe FP, Aylward GP, Shevell MI, Lipkin PH, Squires JK. The thorny nature of predictive validity studies on screening tests for developmental-behavioral problems. *Pediatrics*. 2008;122:866–868.

48. Glascoe FP. Parents' concerns about children's development: prescreening technique or screening test? *Pediatrics*. 1997;99: 522–528.

49. Glascoe FP, MacLean WE, Stone WL. The importance of parents' concerns about their child's behavior. *Clin Pediatr (Phila)*. 1991;30:8–11.

50. Sattler JM. *Assessment of Children. Revised and Updated Third Edition*. San Diego, CA: Jerome Sattler; 1992.

51. Anastasia A. *Psychological Testing*. New York, NY: Macmillan; 1976.

52. Bayley N. *Bayley Scales of Infant Developmental*. 3rd ed. Toronto, Canada: Psychological Corporation; 2006.

53. Dahinten SV, Ford L. *Validation of the Nipissing District Developmental Screen for Use with Children and Toddlers—Working Paper*. Vancouver, BC: Human Early Learning Partnership (HELP); 2004.

54. Wechsler D. *Wechsler Preschool and Primary Scale of Intelligence*. 3rd ed. San Antonio, TX: Harcourt Assessment, Inc.; 2003.

55. McFie J. *Assessment of Organic Intellectual Impairment*. Oxford, England: Academic Press; 1975.

56. Beres KA, Kaufman AA, Perlman MD. Assessment of child intelligence. In: Goldstein G, Hersen M, eds. *Handbook of Psychological Assessment*. 3rd ed. Kidlington, UK: Elsevier Science Ltd.;2000: 65–96.

57. Halperin JM, McKay KE. Psychological testing for child and adolescent psychiatrists: a review of the past 10 years. *J Am Acad Child Adolesc Psychiatry*. 1998;37:575–584.

58. Sparrow SS, Cicchetti DV, Balla DA. *Vineland Adaptive Behavior Scales (Second Edition) Survey Inteview Form Manual*. Minneapolis, MN: NCS Pearson Inc.; 2005.

59. Glascoe FP, Byrne KE, Ashford LG, Johnson KL, Chang B, Strickland B. Accuracy of the Denver II in developmental screening. *Pediatrics*. 1992;89:1221–1225.

60. Zimmerman IL, Steiner VG, Pond RE. *Preschool Language Scale*. 4th ed. San Antonio, TX: Psychological Corporation; 2002.

61. Zimmerman IL, Castilleja NF. The role of a language scale for infant and preschool assessment. *Ment Retard Dev Disabil Res Rev*. 2005;11:238–246.

62. Robins DL, Fein D, Barton M, Green J. The Modified Checklist for Autism in Toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders. *J Autism Dev Disord*. 2001;31:131–151.

63. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Text Rev. Washington, DC: American Psychiatric Association; 2000.

64. Wang H, Chow SC, Li G. On sample size calculation based on odds ratio in clinical trials. *J Biophar Stat.* 2002;12:471–483.

65. Hawass NED. Comparing the sensitivities and specificities of two diagnostic procedures performed on the same group of patients. *Br J Radiol.* 1997;70:360–366.

66. Mackinnon A. A spreadsheet for the calculation of comprehensive statistics for the assessment of diagnostic tests and inter-rater agreement. *Comput Biol Med.* 2000;30:127–134.

67. American Psychological Association. *Standards for Educational and Psychological Tests.* Washington, DC: American Psychiatric Association; 1985.

68. Glascoe FP, Sandler H. Value of parents' estimates of children's developmental ages. *J Pediatr.* 1995;127:831–835.

69. Glascoe FP. It's not what it seems: the relationship between parents' concerns and children with global delays. *Clin Pediatr (Phila).* 1994;33:292–296.