

# ***Attachment 12a***

## ***Pubertal Maturation Self-Assessment Feasibility Study***

## Children's National Medical Center Feasibility Study

From March through August, 2009, a double-blinded feasibility study of the NHANES pubertal maturation self-assessment module was conducted under a contractual agreement with Children's National Medical Center (CNMC), Washington, DC. The objectives of the study were 1) to assess agreement between the self-assessment method and a physical examination and 2) to report ACASI module usability including completion time, acceptability of the questions, ease of use, and ease of self-assessment. Self-assessment results were compared with ratings recorded by CNMC nurse practitioners, physicians, fellows, and medical students who worked with physician preceptors.

### Methods

A convenience sample of children and adolescents 8 to 18 years of age who were scheduled for well-child physical or gynecological examinations were recruited from the CNMC clinic waiting room. Girls who were pregnant or had given birth previously, and children who had medical conditions, physical constraints or cognitive problems that would interfere with normal pubertal development or their ability to use the computer were excluded from the study. Parental permission and consent and assent materials were drafted using the templates required by CHNMC. The CNMC consent and assent forms were approved by the NCHS ERB (1) and the CNMC Institutional Review Board. At the request of the NCHS ERB, the drawings in the module were modified so that they were more muted and less explicit.

Participants completed the pubertal maturation self-assessment ACASI module either before or immediately after the physical examination. They completed the ACASI module alone in a private room equipped with a laptop computer, headphones, and a handheld mirror. Two sexual maturation self-assessment questions were asked: girls were asked to select one breast drawing and one pubic hair drawing that looked the most like their bodies, and boys were asked to select one genital drawing and one pubic hair drawing that looked the most like their bodies. The number of drawings presented varied by the age of the participants. Children age 8–9 years selected from drawings of Tanner stages 1 through 4, and youth age 10–18 years selected from drawings of Tanner stages 1 through 5. Two additional follow-up questions were asked regarding the difficulty of using the computer and of choosing a drawing.

The well-child physical examinations were performed by 70 different nurse practitioners, physicians, fellows, and medical students who worked with physician preceptors. Tanner stages for boys' genitals and pubic hair and girls' breast and pubic hair were recorded by the examiners using the Tanner staging criteria defined by Marshall and Tanner (2, 3). All examiners had been trained in Tanner stage assessment as a part of their routine medical education and Tanner stage drawings were available for their reference. A study investigator and/or a pediatrician or pediatric nurse practitioner trained by study investigators were available to answer examiners' questions regarding the study and Tanner staging. However, examiners were not standardized to Tanner staging prior to the study and no NCHS staff was present during the examination to identify differences in how the Tanner stages were assessed. Examiners and study participants were blinded to self-assessment and examiner ratings respectively.

Fleiss-Cohen weighted kappa coefficients were used to assess self vs. examiner agreement after adjusting for chance agreement (4, 5). Fleiss-Cohen weighted kappa statistics give a higher weight to responses that are within one Tanner stage and a lower weight to responses that differ by two or more Tanner stages (6, 7). A weighed kappa coefficient  $> 0.75$  is considered excellent agreement; a kappa coefficient of  $0.4-0.75$  good agreement, and a kappa coefficient of  $< 0.4$  marginal agreement (8). Comparisons of kappa statistics for obese versus non-obese youth and between age groups (8-10 years, 11-14 years, and 15-18 years) were tested using a Student's t statistic. Significance was based on a two-sided p-value  $< 0.05$ , adjusted for multiple comparisons.

Self- and examiner assessment of Tanner stages are both somewhat subjective as there is no exact cutoff to differentiate the Tanner stages. If the validity and feasibility of a pubertal maturation self-assessment tool were judged only on exact agreement (or with a simple kappa, which gives credit only for exact agreement), then low to moderate agreement between the self-assessment and the examiner ratings would be expected. Therefore, researchers in the field of pubertal development consider agreement within one stage (near agreement) to be sufficient and often report results as agreement within one stage and as Fleiss-Cohen weighted kappa statistics (6, 7, 9, 10).

## Results

Two hundred and thirty-four of the 284 eligible youth completed both the ACASI module and the physical examination, for an 82.4% response rate. Reasons given for choosing not to participate in the study included the child being ill, no time, parent/guardian not interested, child not interested, or the parent/guardian thought the child was too young to complete the ACASI module alone. Study participants were 51% male, 94% black (race defined by examiner, not participant). Approximately 24% of the boys and 27% of the girls were age 8–10 years, 49% of the boys and 40% of the girls were age 11–14 years, and 27% of the boys and 33% of the girls were age 15–18 years. Over 17% of boys and 31% of girls were obese (Table 1).

Results of cross-tabulations between child and examiner assessment are presented in Tables 2 and 3. For boys, 38% exact agreement was seen for genital staging and 53% exact agreement for pubic hair staging. For girls, 53% exact agreement was seen for breast stage and 46% exact agreement for pubic hair stage. However, agreement increased greatly when exact agreement was combined with agreement within one stage. Over 79% of boys self-assessed their genital development within one stage of the examiner's assessment, and 87% self-assessed their pubic hair development within one stage of the examiner's assessment. Likewise, 90% of girls self-assessed their breast development within one stage of the examiner's assessment and 93% self-assessed their pubic hair development within one stage of the examiner's assessment.

Based on Fleiss-Cohen weighted kappa coefficients, boys' assessment of genital development was in good agreement with the examiner's assessment (weighted kappa: 0.65, 95% Confidence Interval (CI): 0.55–0.75), while their assessment of their pubic hair development was in excellent agreement with the examiner's assessment (weighted kappa: 0.78, CI: 0.70–0.86). Girls' assessment of their breast development was in excellent agreement with the examiner's assessment (weighted kappa: 0.81, 95% CI: 0.74–0.87), as was their assessment of their pubic hair development (weighted kappa: 0.78, CI: 0.71–0.86) (Table 4). Agreement with the examiner did not vary between obese and non-obese youth. However, girls age 15–18 years showed significantly less agreement with the examiner for breast and pubic hair development than girls age 8–10 years ( $p \leq 0.01$ ) (Table 4).

Assessment of agreement for some subgroups was limited in situations in which a Tanner stage was not selected by at least one child and one examiner. For example, four of the 8–10 year old boys said they were in the fourth Tanner stage of genital development, while none of the examiners assessed any 8–10 year old boys as stage four of genital development. The kappa calculation included a cell equal to zero and resulted in a very small statistic of 0.03 (Table 4). A similar situation was observed for all 15–18 year olds and analyses of the assessment of pubic hair by 8–10 year-old boys. The use of small samples such as that in the CNMC study can result in an underestimation of the weighted kappa statistic (11).

It should be noted that the overall sample size was based on a 70% power calculation. The study was not powered for gender and age group sub-analyses, thus the very small analytic cell sizes (resulting in zero cells in the diagonal and statistical underestimation of the weighted kappa) for all 15–18 year olds and for all 8–10 year old boys. Analyses by age group were conducted in order to determine if the youngest children would have more difficulty self-assessing their Tanner stage than the older children. The study found that children age 8-10 years did not perform worse than older children, and that the youngest girls actually performed significantly better than the oldest girls.

Participants completed the ACASI module quickly and reported that it was very easy to use. The two sexual maturation questions were completed in an average of 1.2 minutes (SD 0.6 minutes). Time to complete the two puberty questions did not differ by sex or age group. The vast majority of youth reported that the computer was 'easy to use' and that they 'did not have trouble' choosing a drawing (Table 5).

### Discussion

A number of self-assessment questionnaires using labeled drawings or photographs depicting the five Tanner stages of sexual maturation have been developed and tested previously (6, 12-15). The first study to test a self-assessment questionnaire using line drawings of Tanner stages found good to excellent examiner-self-assessment correlations of 0.81 for girls' pubic hair, 0.63 for breast development, 0.63 for boys' pubic hair and 0.59 for genital development (12). Similarly, the first study to test a self-assessment questionnaire using the Tanner photographs found excellent examiner self-assessment correlations (kappa = 0.81 for girl's breast stage, 0.91 for girl's pubic hair stage, and 0.88 for boy's combined pubic hair and genital stage) (14).

Subsequent studies found that self-assessment tools utilizing line drawings or Tanner photographs varied widely in their reliability. In girls, correlations between self- and examiner assessment of breast development have ranged from 0.16–0.87 and correlations for pubic hair stage from 0.33–0.91 (6, 9, 10, 12-25). Likewise, in boys, correlations between self- and examiner assessment of genital development have ranged from 0.18–0.72 and correlations for pubic hair stage from 0.36–0.91 (6, 9, 10, 12, 14-19, 22-24, 26). The clarity of line drawings and photographs is often reduced by photocopying of paper data collection forms, making it difficult for children to discern the difference between the Tanner stages. The more explicit Tanner photographs are no longer considered appropriate for use in many settings.

The CNMC study found that obese and non-obese youth did not differ in agreement between the examiner's assessment and the self-assessment for any of the sexual maturation measures. This was a favorable finding since two earlier studies had found that obese youth, particularly girls, were more likely to overestimate their sexual maturation status (7, 27). However, other studies also found no difference between examiner and self-assessment of sexual maturation by body fatness or BMI (10, 16, 17). The reported variability in self-assessment accuracy by obesity status may be due to altered body perception or difficulty girls have in determining the difference between breast tissue and fatty tissue.

The CNMC study found that the agreement between examiner and self-assessment of sexual maturation differed by age, with older girls showing lower agreement than younger girls. Both Taylor et al. and Norris and Richter also found agreement to be better in younger children than older children (6, 24). However Desmangles et al. and Stephen et al. found that agreement did not differ by age (10, 16). In the CNMC, more mature youth often underestimated their pubertal stage and rated their development as Tanner stage 4 when the examiner rated them as Tanner stage 5. This was seen in both boys and girls, and also has been observed in a few other studies (6, 20).

Since 94% of the participants were black, the CNMC study was unable to examine differences by race. However, other studies have observed variability in self-assessment accuracy by race/ethnicity (20, 21, 28). Fair to moderate agreement between 11–14 year old girls and examiners was observed for areola and pubic hair assessments ( $\kappa = 0.32–0.55$ ) in the National Heart, Lung, and Blood Institute (NHLBI) Growth and Health Study, a 10-year

longitudinal study (21). Although black girls who participated in that study were more mature than white girls at all ages, they underestimated their development compared to white girls. Another study reported that ratings of agreement with physicians were lower for Hispanic girls than for white and black girls (20). In contrast, Stephen et al. found that race did not influence the child's ability to self-assess sexual maturation status (10).

Limitations of the CNMC study include the lack of inter-rater reliability assessment due to the large number of clinical examiners who rotated through the outpatient clinics during the data collection time period (March-August 2009) and the small number of physical examinations conducted by each examiner. Because the examiners were not standardized in Tanner staging prior to the study, the physical examinations could not be described as 'gold standard' examinations, restricting the results of the study to a description of the feasibility, and not the validity, of the ACASI module. Repeat self-assessments were not able to be done to determine intra-rater reliability. Finally, because the study population was primarily African-American, the results could not be generalized to youth of other races/ethnicities.

In summary, children age 8–18 years used the NHANES ACASI module to report sexual maturation self-assessments that showed overall good to excellent agreement with a clinical examiner's assessment of their sexual maturation. Results are similar to those from studies using other self-assessment methods. While the ACASI data will not provide the same definitive assessment as would be obtained using a physical examination, the self-assessment information will be valuable in grouping children into early, middle, and late puberty groups. The results suggest that the module is a feasible replacement of physician examination for obtaining sexual maturation information in large surveys such as NHANES. The next step is the pilot test of this module.

## References

1. Report of NCHS Research Ethics Review Board (ERB). NCHS Protocol #2008-01: "Validation of an Automated Computer-Assisted Self Interview Questionnaire to Assess Pubertal Maturation." July 9, 2008.
2. Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. *Arch Dis Child.* 1969;44:291-303.
3. Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. *Arch Dis Child.* 1970;45:13-23.
4. Rosner B. *Fundamentals of Biostatistics.* 5th ed. Pacific Grove: Duxbury; 2000.
5. Fleiss JL, Cohen J. The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability. *Educ Psychol Meas* 1973;33:613-9.
6. Taylor SJ, Whincup PH, Hindmarsh PC, Lampe F, Odoki K, Cook DG. Performance of a new pubertal self-assessment questionnaire: a preliminary study. *Paediatr Perinat Epidemiol.* 2001;15:88-94.
7. Schall JI, Semeao EJ, Stallings VA, Zemel BS. Self-assessment of sexual maturity status in children with Crohn's disease. *J Pediatr.* 2002;141:223-9.
8. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
9. Chan NP, Sung RY, Kong AP, Goggins WB, So HK, Nelson EA. Reliability of pubertal self-assessment in Hong Kong Chinese children. *J Pediatr Child Health* 2008;44:353-8.
10. Stephen MD, Bryant WP, Wilson DP. Self-assessment of sexual maturation in children and adolescents with diabetes mellitus. *Endocr Prac.* 2008;14:840-5.
11. Shankar V, Bangdiwala SI. Behavior of agreement measures in the presence of zero cells and biased marginal distributions. *J Appl Stat.* 2008;35:445-64.



12. Morris NM and and Udry Jr. Adolescents' self-assessment of sexual maturation. *Pediatrics*. 1980;66:918-20.
13. Brooks-Gunn J, Warren MP, Rosso J, Gargiulo J. Validity of self-report measures of girls' pubertal status. *Child Dev*. 1987;58:829-41.
14. Duke PM, Litt IF, Gross RT. Adolescents' self-assessment of sexual maturation. *Pediatrics* 1980;66:918-20.
15. Neinstein LS. Adolescent self-assessment of sexual maturation: reassessment and evaluation in a mixed ethnic urban population. *Clin Pediatr*. 1982;21:482-4.
16. Desmangles JC, Lappe JM, Lipaczewski G, Haynatzki G. Accuracy of pubertal Tanner staging self-reporting. *J Pediatr Endocrinol Metab*. 2006;19:213-21.
17. Williams CL, Cheyne KL, Houtkooper LK, Lohman TG. Adolescent self-assessment of sexual maturation. Effects of fatness classification and actual sexual maturation stage. *J Adolesc Health Care*. 1988;9:480-2.
18. Miller CL, Tucker ML, Pasch L, Eccles JS. Measuring pubertal development: A comparison of difference scales and different sources. The University of Michigan SRA 1988.
19. Schlossberger NM, Turner RA, Irwin CE. Validity of self-report of pubertal maturation in early adolescents. *J Adolesc Health*. 1992;13:109-13.
20. Hergenroeder AC, Hill RB, Wong WW, Sangi-Haghpeykar H, Taylor W. Validity of self-assessment of pubertal maturation in African American and European American adolescents. *J Adolesc Health*. 1999;24:201-5.
21. Wu Y, Schreiber GB, Klementowicz V, Biro F, Wright D. Racial differences in accuracy of self-assessment of sexual maturation among young black and white girls. *J Adolesc Health*. 2001;28:197-203.
22. Wacharasindhu S, Pri-Ngam P, Kongchonrak T. Self-assessment of sexual maturation in Thai children by Tanner photograph. *J Med Assoc Thai*. 2002;85:308-19.

23. Schmitz KE, Hovell MF, Nichols JF, Irvin VL, Keating K, Simon GM, et al. A validation study of early adolescents' pubertal self assessment. *J Early Adolesc.* 2004;24:357-84.
24. Norris SA, Richter LM. Usefulness and Reliability of Tanner Pubertal Self-Rating to Urban Black Adolescents in South Africa. *J Res Adolesc.* 2005;15:609-24.
25. Leone M, Comtois S. Validity and reliability of self-assessment of sexual maturity in elite adolescent athletes. *J Sports Med Phys Fitness.* 2007;47:361-5.
26. Azevedo JC, Brasil LM, Macedo TB, Pedrosa LF, Arrais RF. Comparison between objective assessment and self-assessment of sexual maturation in children and adolescents. *Jornal de Pediatria* 2009;85:135-42.
27. Boas SR, Falsetti D, Murphy TD, Orenstein DM. Validity of self-assessment of sexual maturation in adolescent male patients with cystic fibrosis. *J Adolesc Health.* 1995;17:42-5.
28. Bonat S, Pathomvanich A, Keil MF, Field AE, Yanovski JA. Self-assessment of pubertal stage in overweight children. *Pediatrics* 2002;110:743-7.
29. Freyre EA, Llaza G, Rebaza-Gonzalez R, Sami D. Tanner's pubic hair staging is not applicable to all adolescents. *J Adolesc Health.* 2002;30:144-5.

**Table 1: Sample size and demographics of the study sample**

	<b>Total (N = 234)</b>		<b>Boys (N = 119)</b>		<b>Girls (N = 115)</b>	
	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>
<b>Race = black</b>	215	91.9	110	94.0	105	92.9
<b>Age 8–10 years</b>	60	25.6	29	24.4	31	27.0
<b>Age 11–14 years</b>	104	44.4	58	48.7	46	40.0
<b>Age 15–18 years</b>	70	29.9	32	26.9	38	33.0
<b>Obese</b>	56	23.9	20	17.1	36	31.3

**Table 2: Cross tabulation between boys’ self assessment (via ACASI module) vs. examiner assessment of genital and pubic hair Tanner stage**

<b>Examiner’s assessment</b>	<b>Boy’s self-assessment</b>					<b>Total</b>
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	
<b>Genital stage</b>						
<b>1</b>	5	14	2	5	0	26
<b>2</b>	2	8	4	3	1	18
<b>3</b>	0	3	7	4	6	20
<b>4</b>	1	2	3	1	2	9
<b>5</b>	0	0	5	15	24	44
<b>Total</b>	8	27	21	28	33	117
<b>Pubic hair stage</b>						
<b>1</b>	15	7	4	4	0	30
<b>2</b>	5	7	6	1	1	20
<b>3</b>	0	0	4	5	4	13
<b>4</b>	0	0	1	12	3	16
<b>5</b>	0	0	1	12	25	38
<b>Total</b>	20	14	16	34	33	117

**Table 3: Cross-tabulation between girls’ self assessment (via ACASI module) vs. examiner assessment of breast and pubic hair Tanner stage**

<b>Examiner’s assessment</b>	<b>Girl’s self-assessment</b>					<b>Total</b>
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	
<b>Breast stage</b>						
<b>1</b>	8	2	0	0	0	10
<b>2</b>	6	10	3	0	0	19
<b>3</b>	0	7	7	1	2	17
<b>4</b>	0	1	2	7	3	13
<b>5</b>	0	0	9	18	29	56
<b>Total</b>	14	20	21	26	34	115
<b>Pubic hair stage</b>						
<b>1</b>	10	1	1	0	0	12

2	1	8	4	1	0	14
3	0	1	5	6	2	14
4	0	0	4	10	12	26
5	0	0	4	24	19	47
<b>Total</b>	11	10	18	41	33	113

**Table 4: Weighted Kappa and 95% Confidence Intervals of self vs. examiner assessment of Tanner stage**

	Boys		Girls	
	Genitals	Pubic hair	Breast	Pubic hair
<b>Overall</b>	0.65 (0.55 - 0.75)	0.78 (0.70 - 0.86)	0.81(0.74 - 0.87)	0.78 (0.71 - 0.86)
<b>8–10 years</b>	0.03 (0.00 - 0.22)	0.32 (0.00 - 0.75)	0.72 (0.53 - 0.91)	0.78 (0.66 - 0.91)
<b>11–14 years</b>	0.47 (0.27 - 0.66)	0.57 (0.40 - 0.74)	0.68 (0.50 - 0.85)	0.52 (0.29 - 0.75)
<b>15–18 years</b>	0.35 (0.02 - 0.69)	0.36 (0.00 - 0.87)	0.30 (0.00-0.61)*	0.20 (0.00-0.56)*
<b>Obese</b>	0.73 (0.56 - 0.90)	0.74 (0.51 - 0.97)	0.78 (0.63 - 0.93)	0.59 (0.34 - 0.84)
<b>Non-obese</b>	0.65 (0.54 - 0.76)	0.78 (0.70 - 0.87)	0.81(0.73 - 0.88)	0.83(0.75 - 0.90)

\* Girls age 15–18 years showed significantly less agreement with the examiner for breast and pubic hair Tanner stage than girls age 8–10 years ( $p \leq 0.01$ )

**Table 5: Time to complete the ACASI, ease of computer use, and ease of selecting a drawing that best matched their body**

	Overall	Boys	Girls	8 – 10 yrs	11 – 14 yrs	15-18 yrs
<b>Minutes to complete the two puberty questions (mean, SD)</b>	1.20 (0.63)	1.27 (0.73)	1.14 (0.49)	1.12 (0.75)	1.24 (0.54)	1.22 (0.64)
<b>Computer was:</b>						
<b>Easy to use</b>	97.4%	95.8%	99.1%	94.9%	97.1%	100%
<b>Somewhat difficult to use</b>	2.2%	3.4%	0.9%	5.1%	1.9%	0%
<b>Hard to use</b>	0.4%	0.9%	0%	0%	1.0%	0%
<b>Choosing a drawing:</b>						
<b>Did not have trouble</b>	79.5%	76.5%	82.6%	76.7%	76.0%	87.1%
<b>Had some trouble</b>	19.7%	22.7%	16.5%	21.7%	23.1%	12.9%
<b>Had a lot of trouble</b>	0.9%	0.8%	0.9%	1.7%	1.0%	0%