0021-7557/09/85-02/135
Jornal de Pediatria

Comparison between objective assessment and self-assessment of sexual maturation in children and adolescents

Jenner C. V. Azevedo,[1] Lana M. P. Brasil,[2] Taísa B. M. A. Macedo,[3] Lucia F. C. Pedrosa,[4] Ricardo F. Arrais[5]

Abstract

Objective: To assess the correlations between objective assessment and self-assessment of sexual maturation in the outpatient and school settings.

Methods: Three hundred and nineteen individuals, 178 (96 boys and 82 girls) from an outpatient clinic and 141 (73 boys and 68 girls) from public schools (8.3-18.7 years), of whom 73 individuals (39 girls an d 34 boys) had a body mass index above the 85th percentile, according to 2000 CDC Growth Chart, were analyzed. All of them were examined sequentially and individually by two trained physicians after a written consent form was signed by parents or surrogates, and then submitted to self-assessment using pictures (Tanner stages). Kappa coefficients between examiners and the self-assessment were calculated based on the collected data. A p value < 5% was established as statistically significant.

Results: No significant difference was observed between correlations obtained from the outpatient clinic and schools, and both groups were combined for analysis. The correlations obtained by examiners were significantly higher than those from self-assessment, with a kappa coefficient (and confidence interval) of 0.75 (0.8-0.69) for breasts/genitals across examiners against 0.27 (0.34-0.20) and 0.29 (0.36-0.22) between the two examiners and the self-assessment ($p < 0.0001$).

Conclusions: In the studied sample, self-assessment of the pubertal stage should not replace the objective assessment made by trained professionals. Improvement of the self-assessment method may validate its use in population-based studies.

J Pediatr (Rio J). 2009;85(2):135-142: Puberty, sexual development, assessment.

Introduction

Assessment of sexual development and maturation is essential for the appropriate assessment of growth in children and adolescents, and is of paramount importance to the analysis of adequate somatic growth and of adequate timing of pubertal development of an individual.[1] In the early 1960s, Tanner[2] proposed a simple and practical method for pubertal staging, developed through the assessment of English children and adolescents and thoroughly revised by Marshall & Tanner[3,4] in 1969 and 1970. This classification, albeit quite objective, includes individuals who find themselves in intermediate developmental stages, therefore resulting in classification differences.

Such differences, in males,[3] are minimized by the classification of intermediate stages (G2, G3, G4), based on testicular volume, according to which those boys with a volume equal to or greater than 4 cm^3 are at the beginning of the

1. Aperfeiçoando, Endocrinologia Pediátrica. Especialista, Pediatria, Universidade Federal do Rio Grande do Norte (UFRN), Natal, RN, Brazil.
2. Mestre. Ciências da Saúde, UFRN, Natal, RN, Brazil.
3. Especialista, Endocrinologia Pediátrica , UFRN, Natal, RN, Brazil.
4. Doutora. Ciências (Nutrição), UFRN, Natal, RN, Brazil.
5. Doutor. Endocrinologia Clínica, UFRN, Natal, RN, Brazil.

No conflicts of interest declared concerning the publication of this article.

Suggested citation: Azevedo JC, Brasil LM, Macedo TB, Pedrosa LF, Arrais RF. Comparison between objective assessment and self-assessment of sexual maturation in children and adolescents. J Pediatr (Rio J). 2009;85(2):135-142.

Manuscript received Sep 15 2008, accepted for publication Jan 07 2009.

doi:10.2223/JPED.1875

pubertal stage. For both sexes, Tanner stage 1 is characterized by the absence of pubertal signs, whereas stage 5 corresponds to the final pubertal stage. Under these circumstances, misclassification occurs very rarely.

Despite the criticisms against the clinical assessment of pubertal stage by this method, its use has been widely spread in outpatient routine, without any other concurrent clinical assessment methods.

The ease with which the method can be used, combined with situations that hinder the examination of genitals in clinical practice (inappropriate facilities, cultural and emotional factors, among others) and that often do not allow for an objective assessment by the examiner, has prompted the development of research studies that could correlate self-assessment and the objective assessment carried out by trained professionals. The validation of self-assessment would permit its inclusion in research protocols where the objective assessment by qualified physicians is unavailable or unsuitable.

The results described in the literature are inconsistent with regard to the efficiency of self-assessment comparatively to the objective assessment of sexual maturation. Factors such as the method chosen, cultural characteristics, notions of self-image (influenced by the presence or not of overweight, nutritional condition that is knowingly associated with problems related to self-image perception) are certainly accountable for discrepant results in different populations.

The first studies, back in the early 1980s, showed apparently thriving results,[5,6] with correlation indices greater than 80% in two studies. One of these studies was conducted by Saito[6] in a Brazilian sample using the same self-assessment method described in the present study. As studies advanced, the first impressions of good accuracy began to be questioned, in general terms[7,8] and in specific situations involving racial[9-11] and sociocultural[12-14] differences and even in the presence of diverse clinical conditions that potentially led to discrepancies, such as obesity[15,16] and chronic diseases that impair normal development and self-image (cystic fibrosis,[17] anorexia nervosa[18] or delayed pubertal development[19]), revealing variable degrees of accuracy of self-assessment, usually favoring objective assessment to the detriment of the former. A recent study, conducted with 130 girls and 110 boys, compared self-assessment with the objective assessment made by a pediatric endocrinologist. There was misclassification of the girls into Tanner stages in 40% with regard to breasts and in 23% with regard to pubic hair. As to boys, 39% misclassified their pubic hair stage. No statistical difference was observed as to the ages of the children who self-assessed correctly and those who self-assessed incorrectly, and no independent predictors were identified for the correction of Tanner stages. These findings prompted the authors to conclude that pubertal self-assessment in children and adolescents is not reliable.[20]

The present study aims to determine the correlation between self-assessment and objective assessment of pubertal maturation in a sample of children and adolescents examined at an outpatient clinic and at public schools of Natal, state capital of Rio Grande do Norte, Brazil. In addition, it verifies whether overweight plays any role in such correlations. The final results of these studies should provide subsidies for the adoption or not of pubertal self-assessment in epidemiological studies in this population.

Methods

The research protocols of outpatient individuals were approved by the Research and Ethics Committee of Universidade Federal do Rio Grande do Norte (UFRN) (CONEP 114314/CEP-UFRN 187-06), and in schools, by the Research and Ethics Committee of Hospital Universitário Onofre Lopes, affiliated with UFRN (CONEP 141855/CEP-HUOL 67-07). After the signature of the informed consent form, girls aged 8 years or older and boys aged 9 years or older, without chronic diseases, were included in the study. The following individuals were excluded: Patients being treated with growth hormone (GH), luteinizing hormone-releasing hormone (LHRH) analogs and sex steroids; patients followed up for delayed or early sexual development, even without any treatment; patients with confirmed genetic syndromes; and patients with confirmed cognitive deficit. Therefore, 319 patients were included: 178 assessed at the Pediatric Endocrinology Outpatient Clinic Heriberto Ferreira Bezerra, affiliated with the hospital complex of UFRN, and 141 students enrolled in municipal public schools of Natal. Inclusion criteria were similar at the outpatient clinic and at schools. At the outpatient clinic, individuals were assessed sequentially and all participants who matched the age thresholds and who did not meet any exclusion criteria were invited to take part in the study, with minimal loss. At schools, individuals were randomly selected at first, and the selected ones, who participated in the research protocol (which included blood tests) represented losses of 15 and 20% (refusal to participate and loss to follow-up).

Participants were evaluated in terms of pubertal development by Tanner criteria applied by two physicians, in alternate turns during normal physical examination. The first physician was a properly trained resident doctor/medical graduate student (examiner 1) and the second one was the preceptor (examiner 2). The assessment of sexual maturation was carried out individually at the end of the clinical examination, in the presence of parents or surrogates, with the child/adolescent lying on an appropriate stretcher and in an environment that allowed for proper privacy. Testicular volume was measured in order to help distinguish between stages, especially between G1 and G2 (at least one of the testicles measuring 4 cm^3 or more), using Prader orchidometer and the classification proposed by Marshall & Tanner[3] for characterization of testicular volume. However, testicular volume

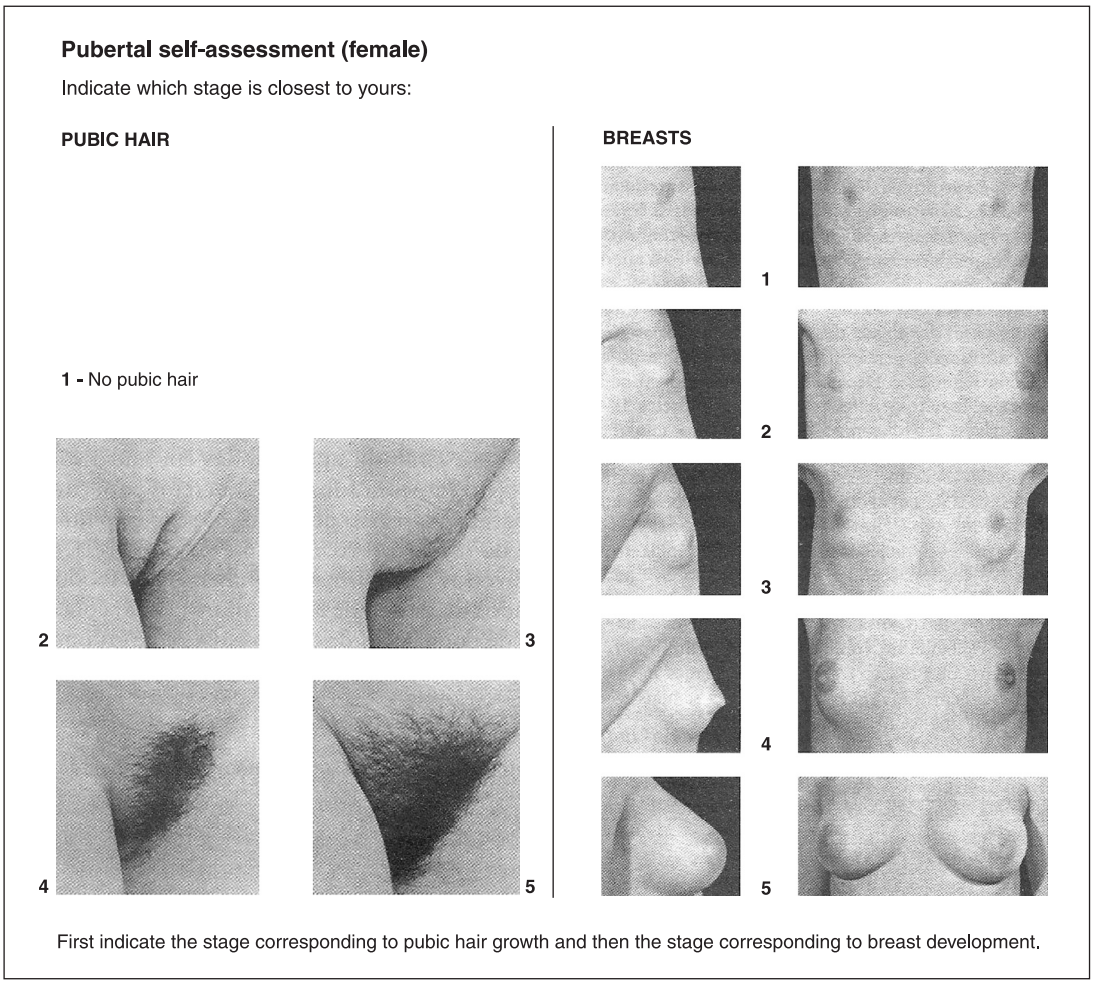

Figure 1 - Pictures used for self-assessment of girls[2]

was not tabulated. After that, self-assessment was performed by having participants choose the specific pictures for Tanner stages of sexual maturation (breasts and pubic hair for girls; genitals and pubic hair for boys) that best described their current stage of development (Figures 1 and 2).

The answers were written in a specific table containing patient's initials, medical chart number (in case of outpatients), age, sex, body mass index (BMI) and assessment of sexual maturation by examiner 1 and examiner 2, along with the self-assessment. For BMI, the criteria for risk of overweight and overweight were used, according to the cutoff points suggested by 2000 CDC Growth Chart.[21] The data were analyzed by calculating the kappa coefficient, using StatistiXL software (version 1.7 beta). The analysis was performed by all participants and by groups divided into only boys, only girls, those with BMI ≥ 85th percentile and boys and girls with BMI ≥ 85th percentile. They were categorized into: poor concordance, coefficients < 0.2; fair concordance, between 0.2 and 0.4; moderate concordance, between 0.4 and 0.6; good concordance, between 0.6 and 0.8; and excellent concordance, between 0.8 and 1.0. Associations whose confidence intervals did not overlap across the compared groups were considered to be statistically different. All correlations obtained with $p < 0.05$ were statistically significant.

Results

The best coefficient in the assessment of breasts or genitals including all 319 participants was obtained when the examination was performed by the examiners, with a kappa coefficient of 0.75 (good concordance). The coefficients between the examiners' assessment and self-assessment showed fair concordance (kappa coefficient of 0.27 between examiner 1 and the participants, and 0.29 between examiner 2 and the participants), and were statistically lower than those provided by the objective assessment of physicians.

When the whole group of participants was analyzed, there was greater symmetry in the assessment of pubic hair between both examiners, with a kappa coefficient of 0.79 (good concordance). In the assessments between examiner 1 and the participants and between examiner 2 and the participants, the coefficients corresponded to 0.42 and 0.46, respectively (both showing moderate concordance)

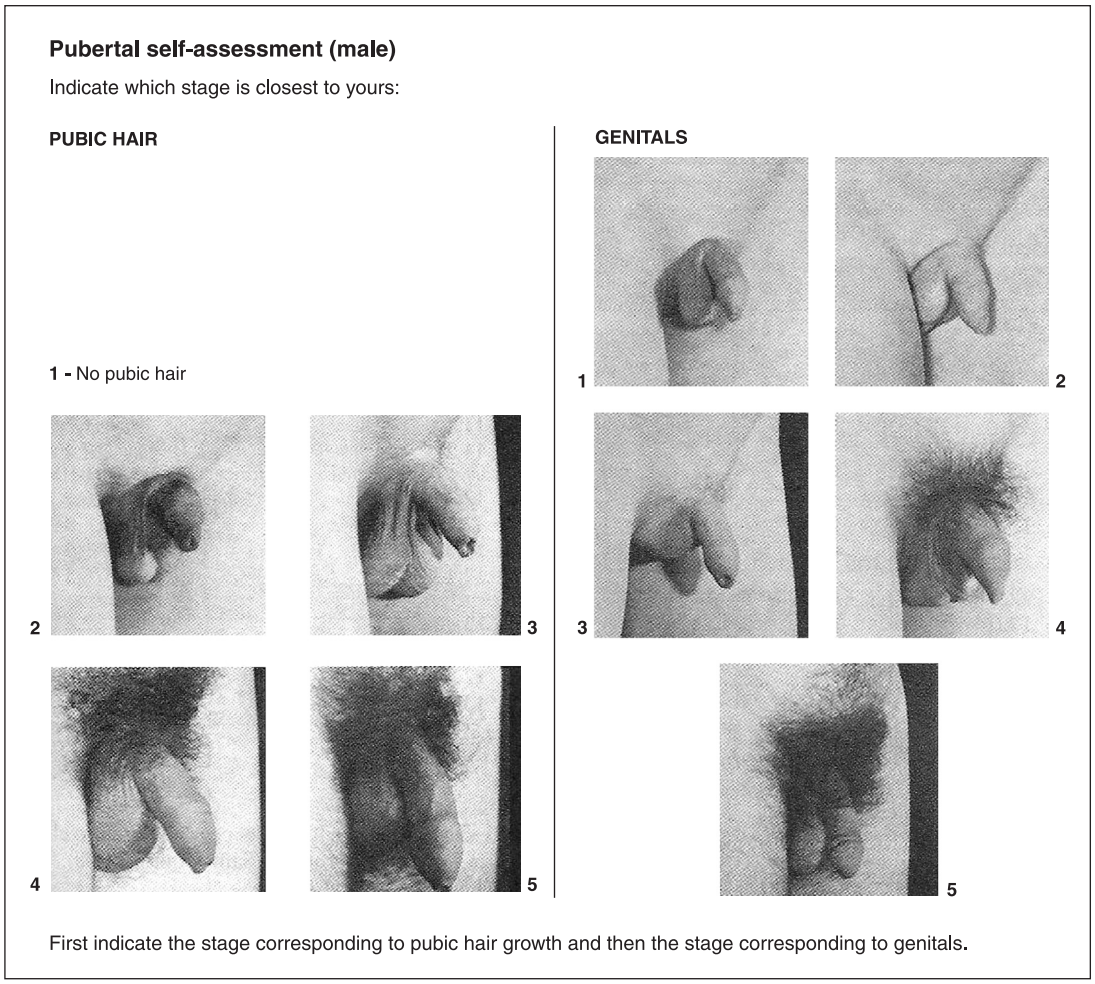

Figure 2 - Pictures used for self-assessment of boys[2]

By analyzing only the 169 male participants, the concordance between examiners yielded a better coefficient, both in terms of pubic hair and genitals (0.75 and 0.71, respectively). In the assessment of genitals, boys had a coefficient of 0.21 and 0.22 in the examinations made by examiners 1 and 2, respectively. Likewise, for pubic hair, the coefficients corresponded to 0.36 and 0.39 when the same groups were evaluated.

Among the 150 girls analyzed, the coefficients between medical professionals with regard to the examination of breasts and pubic hair corresponded to 0.77 (good concordance) and 0.82 (excellent concordance), respectively. In the self-assessment of breasts, the coefficients were 0.28 and 0.32 when examined by participants and examiner 1 and by participants and preceptors, with fair concordance in both cases. Girls self-assessed the staging of pubic hair with moderate concordance, obtaining coefficients of 0.47 between participants and examiner 1 and of 0.52 between participants and examiner 2. Albeit significant, these associations are lower than those obtained between examiners.

When only the group of 73 participants (boys and girls) with overweight or risk of overweight were considered, examiners found higher coefficients for the examination of breasts/genitals and pubic hair (0.75 and 0.72, respectively). The comparison between participants and examiner 1 yielded a coefficient of 0.25 for the assessment of breasts/genitals and of 0.39 for pubic hair. The self-assessment of breasts/genitals and of pubic hair compared to the examination performed by examiner 2 yielded coefficients of 0.38 and 0.42, respectively.

Among the 36 male participants with BMI > 85th percentile, examiners found a coefficient of 0.77 for the assessment of genitals and of 0.66 for pubic hair. The coefficients between examiners and participants could not be determined due to technical limitations (there were no individuals representing at least one of Tanner stages, thus hindering the calculation of the kappa coefficient).

Among the 37 female participants with BMI > 85th percentile, the coefficients, although higher between examiners (0.71 for breasts and 0.75 for pubic hair), when compared to

Table 1 - Kappa coefficient for participants of both sexes and the group with BMI > 85th percentile of weight, including confidence intervals

Groups	kappa	95%CI
Boys and girls, n = 319 (8.3-18.7 years)		
Genitals or breasts (Tanner)		
E1 x E2	0.75	0.80-0.69
E1 x SA	0.27	0.34-0.20
E2 x SA	0.29	0.36-0.22
Pubic hair (Tanner)		
E1 x E2	0.79	0.84-0.74
E1 x SA	0.42	0.49-0.35
E2 x SA	0.46	0.53-0.39
Boys and girls with BMI > 85th percentile n = 73 (8.8-18.7 years)		
Genitals or breasts (Tanner)		
E1 x E2	0.75	0.86-0.63
E1 x SA	0.25	0.39-0.10
E2 x SA	0.38	0.52-0.23
Pubic hair (Tanner)		
E1 x E2	0.72	0.84-0.60
E1 x SA	0.39	0.53-0.24
E2 x SA	0.42	0.56-0.27

95%CI = 95% confidence interval; BMI = body mass index; E1 = examiner 1; E2 = examiner 2; SA = self-assessment.

those obtained by the comparison with the self-assessment (0.33 and 0.45 for breasts and pubic hair by examiner 1 and 0.43 and 0.45 for breasts and pubic hair by examiner 2), do not allow us to affirm that these values were significantly different because the adopted confidence intervals were interpolated. The data on the associations between all compared groups are shown in Tables 1 to 3.

Discussion

The data obtained show that the assessment of pubertal stages in the populations of students and outpatients, performed by trained physicians, yields more reliable results than self-assessment, especially regarding the assessment of breasts, with slightly higher coefficients when compared to the assessment of genitals (without significant difference between sexes). Also, in the assessment of pubic hair, the most reliable results were those found by examiners, although participants usually assessed pubic hair growth more correctly than they did with respect to their breasts/genitals. The associations, although significant, are effectively lower in self-assessments, yielding fair coefficients in relation to the good coefficients provided by medical evaluation. These data significantly differ from those published by Saito,[6] who found good concordance between objective assessment and self-assessment of pubertal development in a considerably smaller outpatient population of adolescents using similar pictures to those shown in the present study. The results also differ from the data described by Desmangles et al.[20] in 2006, in which these authors concluded that self-assessment of pubertal development in a North American population was not reliable, bearing some resemblance to the results published by Martin et al.,[22] who observed significant, but moderate, concordance between objective assessment and self-assessment of boys by means of pictures and drawings. The sequence of their work, with girls assessed by physical education teachers, in 2002, yielded the same results as their study with boys, with moderate concordance, and the authors concluded that self-assessment was useful.[23]

In boys who were assessed separately, the reliability of self-assessment was also lower in relation to the examination performed by trained physicians given that only fair concordance was obtained for the self-assessment of genitals and pubic hair. When only the results obtained for the girls were considered, the importance of Tanner staging by a physician was shown by excellent concordance between the two examiners with respect to public hair and by good concordance with respect to breasts. Female participants also showed that they performed a better self-assessment of their pubic hair than of their breasts, with moderate versus fair concordance, respectively. This finding may be related to better perception of self-image by the girls than by the boys.

Table 2 - Kappa coefficient for male participants and the group with BMI > 85th percentile, including confidence intervals

Groups	kappa	95%CI
Boys, n = 169 (9-18.7 years)		
Genitals (Tanner)		
E1 x E2	0.71	0.79-0.63
E1 x SA	0.21	0.31-0.11
E2 x SA	0.22	0.32-0.12
Pubic hair (Tanner)		
E1 x E2	0.75	0.83-0.67
E1 x SA	0.36	0.46-0.26
E2 x SA	0.39	0.49-0.29
Boys with BMI > 85th percentile n = 36 (9-18.7 years)		
Genitals (Tanner)		
E1 x E2	0.77	0.92-0.58
E1 x SA	*	
E2 x SA	*	
Pubic hair (Tanner)		
E1 x E2	0.66	0.85-0.47
E1 x SA	*	
E2 x SA	*	

95%CI = 95% confidence interval; BMI = body mass index; E1 = examiner 1; E2 = examiner 2; SA = self-assessment.
* Not calculated due to lack of corresponding values (Tanner stages) in some subgroups.

The coefficient between medical assessment was similar (kappa = 0.75) in participants with or without overweight or risk of overweight, especially in the assessment of breasts/genitals, which had good concordance. However, medical assessment is the most reliable in Tanner staging, showing higher coefficients, as self-assessment yielded fair concordance for breasts or genitals and moderate concordance with respect to pubic hair. Bonat et al.,[16] in 2002, concluded that obese girls overestimated Tanner stage with regard to their breasts and that obese and nonobese boys significantly overestimated the staging of pubic hair. Therefore, the authors decided not to recommend pubertal self-assessment of breasts in girls and of pubic hair in boys. The statistical analysis in the present study shows that, in the group of girls with overweight or risk of overweight, self-assessment has lower correlations, although without significant difference in relation to the examination performed by trained physicians, but among these, correlation was good with respect to the assessment of pubic hair and breasts, whereas self-assessment showed moderate concordance for both variables. In the group of boys with BMI ≥ 85th percentile, the correlation between the assessment made by physicians was good with respect to genitals and pubic hair.

A study undertaken in Thailand by Wacharasindhu et al.,[14] in 2002, revealed that the accuracy of self-assessment was better when patients were given a longer time for self-assessment. Another factor that could possibly increase the efficiency and reliability of self-assessment concerns the quality of the graphical material used, despite the fact that two similar studies carried out in 2001 and 2002 by Martin et al.[22] and by Bojikian et al.,[23] respectively, did not demonstrate a significant difference between self-assessment of boys and girls by means of pictures or drawings. Evidence of good-quality graphical material as a factor of better correlation in self-assessment of pubertal stages was suggested by a recent work performed in Hong Kong by Chan et al.,[24] who assessed 354 Chinese children and adolescents aged 8 to 18 years, comparing the self-assessment made with the help of explanatory drawings of Tanner stages with a later objective assessment carried out by properly trained physicians. Apparently, it is not enough just to improve the graphical material by adding explanatory texts if the limited intellectual capacity still observed in part of our population is not taken into account. This is, indeed, a more complex task, as pointed out by Guimarães & Costa Passos.[12]

The present study shows that, in the studied population, the use of images from Tanner's original classification yields significant associations; however, these associations are effectively lower in self-assessments, and therefore, objective assessment by trained professionals should not be

Table 3 - Kappa coefficient for female participants and the group with BMI > 85th percentile, including confidence intervals

Groups	kappa	95%CI
Girls, n = 150 (8.3-17.9 years)		
Breasts (Tanner)		
E1 x E2	0.77	0.85-0.69
E1 x SA	0.28	0.38-0.18
E2 x SA	0.32	0.42-0.22
Pubic hair (Tanner)		
E1 x E2	0.82	0.89-0.75
E1 x SA	0.47	0.57-0.37
E2 x SA	0.52	0.62-0.42
Girls with BMI > 85th percentile n = 37 (8.8-17.9 years)		
Breasts (Tanner)		
E1 x E2	0.71	0.89-0.53
E1 x SA	0.33	0.54-0.11*
E2 x SA	0.43	0.64-0.21*
Pubic hair (Tanner)		
E1 x E2	0.75	0.91-0.59
E1 x SA	0.45	0.65-0.25*
E2 x SA	0.45	0.65-0.25*

95%CI = 95% confidence interval; BMI = body mass index; E1 = examiner 1; E2 = examiner 2; SA = self-assessment.
* Groups with interpolated confidence intervals between the correlations obtained from examiners' assessment and from the comparison between examiners and self-assessment.

replaced with self-assessment of sexual maturation by Tanner stages.

Acknowledgements

We are grateful to CNPq (Project 478287-06/02) for partially financing the research, in the selection and access of participants from the schools included in the study. We are also indebted to Dr. Eveline de Alencar Oliveira, Dr. Marina Romero Costa and Dr. Tânia Azevedo Cabral, who actively participated in the collection of data from the schools and from the outpatient clinic.

References

1. Chipkevitch E. Avaliação clínica da maturação sexual na adolescência. J Pediatr (Rio J). 2001;77:S135-42.
2. Tanner, JM. Growth at adolescence. 2nd ed. Oxford: Blackwell Scientific; 1962.
3. Marshall W A, Tanner JM. Variation in the pattern of pubertal changes in boys. Arch Dis Child. 1970;45:13-23.
4. Marshall WA, Tanner JM. Variation in the pattern of pubertal changes in girls. Arch Dis Child. 1969;44:291-303.
5. Duke PM, Litt IF, Gross RT. Adolescents' self-assessment of sexual maturation. Pediatrics. 1980;66:918-20.
6. Saito MI. Maturação sexual: auto-avaliação do adolescente. Pediatr (São Paulo). 1984;6:111-5.
7. Brooks-Gunn J, Warren MP, Rosso J, Gargiulo J. Validity of self-report measures of girls' pubertal status. Child Dev. 1987; 58:829-41.
8. Schlossberger NM, Turner RA, Irwin CE Jr. Validity of self-report of pubertal maturation in early adolescents. J Adolesc Health. 1992;13:109-13.
9. Neinstein LS. Adolescent self-assessment of sexual maturation: reassessment and evaluation in a mixed ethnic urban population. Clin Pediatr (Phila). 1982;21:482-4.
10. 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.
11. 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.
12. Guimarães JP, Costa Passos AD. Análise de concordância entre informações referidas e observadas acerca do estadiamento pubertário entre escolares do sexo feminino. Rev Saude Publica. 1997;31:263-71.
13. Verona-Lopez W, Guillemot M, Spyckerelle Y, Mulot B, Deschamps JP. Self assessment of the stages of sex maturation in male adolescents. Pediatrie. 1988;43:245-9.

14. 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.

15. Williams RL, 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.

16. Bonat S, Pathomvanich A, Keil MF, Field AE, Yanovski JA. Self-assessment of pubertal stage in overweight children. Pediatrics. 2002;110:743-7.

17. 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.

18. Hick KM, Katzman DK. Self-assessment of sexual maturation in adolescent females with anorexia nervosa. J Adolesc Health. 1999;24:206-11.

19. Finkelstein JW, D'Arcangelo MR, Susman EJ, Chinchilli VM, Kunselman SJ, Schwab J, et al. Self-assessment of sexual physical maturation in boys and girls with delayed puberty. J Adolesc Health. 1999;25:379-81.

20. Desmangles JC, Lappe JM, Lipaczewski G, Haynatzki G. Accuracy of pubertal Tanner staging self-reporting. J Pediatr Endocrinol Metab. 2006;19:213-21.

21. National Health and Nutrition Examination Survey (NHANES). 2000 CDC Growth Charts: United States. [page on the Internet]. http://www.cdc.gov/growthcharts. Access: 25/08/2007.

22. Martin RH, Uezu R, Parra AS, Arena SS, Bojikian LP, Bohme MT. Auto-avaliação da maturação sexual masculina por meio da utilização de desenhos e fotos. Rev Paul Educ Fis. 2001; 15: 212-22.

23. Bojikian LP, Massa M, Martin RH, Teixeira CP, Kiss MA, Bohme MT. Auto-avaliação puberal feminina por meio de desenhos e fotos. Rev Bras Ativ Fis Saude. 2002;7:24-34.

24. Chan N, Sung RY, Kong AP, Goggins WB, So HK, Nelson AS. Reliability of pubertal self-assessment in Hong Kong. J Pediatr Child Health. 2008; 44:353-8.

Correspondence:
Ricardo Fernando Arrais
Rua Antomar de Brito Freitas, 3644
CEP 59064-590 - Natal, RN - Brazil
E-mail: rfarrais@ufrnet.br