

A Self-Administered Rating Scale for Pubertal Development

MARY A. CARSKADON, Ph.D., AND CHRISTINE ACEBO, Ph.D.

The purpose of this study was to assess the reliability and validity of a new self-rating scale to measure children's pubertal status without pictorial representations or interviews. The scale is an adaptation of an interview-based puberty-rating scale by Petersen, and included scores for each of five items rating physical development, an overall maturation measure, and a categorical maturation score designed to be similar to Tanner staging categories. Each measure was obtained from independent ratings by students and parents, and a 3-point categorical scale was also obtained from teachers. Subjects included 698 5th- and 6th-grade students (323 boys and 375 girls) from 61 schools and their parents and teachers. Fifth-grade students rated themselves and were rated by parents as less mature than 6th graders; 6th-grade girls were consistently rated more mature than boys of the same age. Significant correlations were found between parents and students for all of the measures for 6th-graders and 5th-grade girls and several measures for 5th-grade boys. This new scale is a useful tool for assessing pubertal status in settings that require noninvasive measures.

KEY WORDS:

Puberty scale
Adolescent development
Maturation
Reliability
Validity

Measurement and documentation of pubertal maturation with noninvasive techniques is a major concern for research in many disciplines where puberty is likely to have a significant impact. A number of approaches have been developed to address this concern.

Among the most well-recognized systems to assess physical maturation is that developed by Tanner (1). Application of this methodology requires inspection of the naked child by a trained medical expert, a procedure that is often unavailable to a given research project or not acceptable owing to other study needs. Other investigators have presented alternative puberty rating methods that do not require inspection of naked children but instead rely upon a child's self-report. All systems use Tanner staging as the benchmark. Neither visual evaluation nor use of pictorial representations (2,3) were available to Petersen and her colleagues (4) for a school-based early adolescent study due to objections raised by parents and school officials. As a result, they developed an interview version of a Pubertal Development Scale (PDS), for which they obtained measures of reliability and validity in 6th, 7th, and 8th graders (5,6).

None of these techniques was appropriate for our childhood sleep study, which required a self-administered questionnaire to be completed in the classroom. Therefore, we revised the Petersen et al. (5) verbal descriptions into boy's and girl's pubertal development items that we embedded in our sleep habits survey. Because the childhood sleep study also involved parental assessments, a parallel parent form of the items was developed as well. Finally, teacher ratings of students' physical maturity on a simple 3-point scale were also collected. The reliability of this self-administered instrument was assessed

From the Sleep Research Laboratory, E. P. Bradley Hospital, Department of Psychiatry and Human Behavior, Brown University School of Medicine, East Providence, Rhode Island.

Address reprint requests to: Mary A. Carskadon, Ph.D., Sleep Research Laboratory, E. P. Bradley Hospital, Department of Psychiatry and Human Behavior, Brown University School of Medicine, 1011 Veterans Memorial Parkway, East Providence, RI 02915.

Manuscript accepted September 29, 1992.

and correlations with parent and teacher ratings examined.

Methods

The pubertal scale utilized in the survey study (described below and available from the author upon request) was pretested using 38 adolescent subjects seen in our laboratory as part of another study. This sample included 17 girls (ages 10–16 years) and 21 boys (ages 9–16 years), who were also examined by a pediatrician for conventional Tanner staging. The pediatrician was blind to the subjects' self-evaluations. Mean self-rating scores from the children were compared to the pediatrician's assessment of Tanner stage based upon pubic hair growth. The Spearman correlation coefficient between self-rated and pediatrician-rated physical development was high ($r = 0.868-0.841$, $p < 0.001$), and a high correspondence between the two measures of physical development was indicated.

Pubertal development data were gathered as part of a larger survey study of sleep patterns in grades 4, 5, and 6 (7). The source of subjects for the sleep survey was teachers who responded to a letter in the teacher's edition of *SuperScience Blue*, a science magazine for primary schools. Teachers who expressed interest in having their classes participate in a study about "physical maturation and body clocks" were sent preliminary materials to obtain approval for the study from local school officials. Teachers from 87 schools were mailed questionnaires; returns came from 78 schools in 36 states.

Packets of materials were sent to teachers on 11 March 1992; packets contained an instruction manual for the teacher and a postage-paid envelope for each student marked "Boy" or "Girl" on the front. Teachers were instructed to give "boy" envelopes to boys and "girl" envelopes to girls. Each envelope contained a 6-page form for students to complete about themselves (with questions about sleep, mood, morningness/eveningness, physical development, and so forth), a letter to parents, a 2-page form for parents to complete about the student (including sleep, family functioning, seasonal patterns, physical development, etc.), a 3-page form (2 each) for parents to complete about their own sleep, and a half-page form including 5 items for the teacher to complete about the student. Forms within an envelope were given a common numerical code, so that student, parent, and teacher data could be linked for each student. Students completed their own

forms as a classroom exercise. After completing their forms, students were instructed to remove the Teacher Form from the packet, write their name on a sticker located on this form, and pass it in to the teacher. Teachers completed the Teacher Forms and removed the labels before mailing all these forms in a single envelope.

After completing his or her form, each student folded and sealed it, returned it to the envelope, and took the envelope home. The letter to parents described the study and its purpose and requested that parents complete the remaining forms and provide consent for the inclusion of their surveys and those of their child in the research project. This consent procedure was approved by the Rhode Island Hospital/Bradley Hospital joint Institutional Review Board.

The following response rates were achieved over the entire population (total forms sent to grades 4, 5, and 6 = 3,942); for girls, 52% returned their own forms, 49% of parents about student, and 73% of teachers; for boys, 44% returned their own forms, 41% of parents about student, 67% of teachers.

We were able to compare children who did and did not return their forms by evaluating data from the teacher forms of all the 4th, 5th, and 6th graders, as teacher forms were returned for 1,786 students who responded (839 boys and 947 girls) as well as from 956 students who did not respond (503 boys and 453 girls). Students who returned their forms were rated as more socially mature, functioning better academically, more active and engaged in class, and more alert and wide awake in class. These differences were statistically significant for both boys and girls. There was no difference on the teacher ratings of physical maturity between those who did and did not return their forms.

The sample included all the students aged 10, 11, and 12 years from the 5th-grade classes and all the 11- and 12-year-old 6th-grade students who completed the basic demographic items on the survey as well as all five of the physical development items. This resulted in a group including 39 5th-grade classes and 33 6th-grade classes from 61 schools in 31 states.

Table 1 lists the items that were included in the physical development assessment on the students' forms. Items 4 and 5 were unique to the gender-specific student forms. Items were counted as "missing" if the student failed to respond or responded "I don't know." Thus, of all the fifth graders, 43% of boys and 46% of girls had a score of missing for at least one item; of all the sixth graders, 37% of boys

Table 1. Physical Development Items

Introduction: The next questions are about changes that may be happening to your body. These changes normally happen to different young people at different ages. Since they may have something to do with your sleep patterns, do your best to answer carefully. If you do not understand a question or do not know the answer, just mark "I don't know."

Questions

1. Would you say that your growth in height:
2. And how about the growth of your body hair? ("Body hair" means hair any place other than your head, such as under your arms.) Would you say that your body hair growth:
3. Have you noticed any skin changes, especially pimples?

FORM FOR BOYS:

4. Have you noticed a deepening of your voice?
5. Have you begun to grow hair on your face?

FORM FOR GIRLS:

4. Have you noticed that your breasts have begun to grow?
- 5a. Have you begun to menstruate (started to have your period)?
- 5b. If yes, how old were you when you started to menstruate?

For Items 1 through 4 on the girls' version and all items on the boys' version, response options were: not yet started (1 point); barely started (2 points); definitely started (3 points); seems complete (4 points); I don't know (missing). Yes on the menstruation item = 4 points; no = 1 point. Point values are averaged for all items to give a Pubertal Development Scale (PDS). Puberty Category Scores are computed using the criteria of Crockett et al. [6]. Puberty Category Scores for boys used *body hair growth*, *voice change*, and *facial hair growth* as follows: Prepubertal = 3; Early Pubertal = 4 or 5 (no 3-point responses); Midpubertal = 6, 7, or 8 (no 4-points); Late pubertal = 9-11; Postpubertal = 12. For girls, Puberty Category Scores used *body hair growth*, *breast development*, and *menarche* as follows: Prepubertal = 3; Early Puberty = 3 and no menarche; Midpubertal = 4 and no menarche; Late Puberty = ≤ 7 and menarche; Postpubertal = 8 and menarche.

and 37% of girls had a missing score on at least one item. [Computation of Pubertal Development Scores and Puberty Category Scores are described in the footnote to Table 1.] The physical development items on parent forms were comparable to those on the student forms. PDS values for the parents' data were computed as for the students' data (see Table 1). The physical maturation item on the teacher's form asked: "At what level of physical maturity is the student?" The following response options were provided: (1) Immature (no signs of puberty); (2) Maturing (obvious signs of puberty, e.g., skin changes, especially pimples; voice changes or facial hair growth in boys); (3) Mature (changes nearly complete); or (4) I do not know the student well enough to rate. Missing data were identified for parents and teachers in the same manner as for stu-

Table 2. Sample Description

Group	n	Mean Age in Years		
		(SD in Months)	Parent n	Teacher n
Grade 5 boys	172	11.2 (5.2)	144	168
Grade 5 girls	178	11.0 (4.9)	158	177
Grade 6 boys	151	12.1 (4.6)	121	145
Grade 6 girls	197	12.0 (4.8)	176	192

dents. Table 2 describes the final groups used in the analyses.

The participants were primarily Caucasian (87%), living in single-family homes (83%) primarily in population areas of less than 50,000 people (71%). Forty-five percent were firstborn. Mothers alone completed 79% of the parent forms, with both mother and father and father alone each accounting for 10%, and another guardian for the remaining 1%.

Results

Internal Consistency

Internal consistency of the student and parent versions of the pubertal development measures were assessed using Cronbach's coefficient α (8); α values ranged from 0.67 to 0.70 for the student versions and from 0.68 to 0.78 in the parent versions. All versions of the PDS showed similar levels of internal consistency, indicating that the students' self-administered ratings in the present study were similar to those of the Petersen et al. interview ratings (5). Furthermore, the parental PDS showed comparable values. It is noteworthy that even in the younger groups, the student and parent PDS maintained internal reliability.

Item and Full-Scale Scores—Students and Parents

Table 3 shows the mean and SD of item scores and overall mean PDS for the student and parent ratings of 5th and 6th graders. A consistent pattern was found, with significantly higher (more mature) mean item values in the older than younger students for boys [$F(1, 321) = 3.98, p < 0.05$] and particularly so for girls [$F(1, 373) = 49.41, p < 0.001$]. Furthermore, the overall mean PDS in the 6th-grade girls was significantly higher than for the 6th-grade boys ($t = 6.61; df = 336; p < 0.001$).

Comparisons between student and parent ratings

Table 3. Group Mean (SD) Scores on Individual Items and Overall Mean

Item	Students		Parents	
	Fifth	Sixth	Fifth	Sixth
Boys				
Growth spurt	2.20 (0.90)	2.16 (0.84)	2.14 (0.78)	2.17 (0.83)
Body hair	1.98 (0.83) ^b	2.21 (0.80) ^{ab}	1.44 (0.65)	1.71 (0.77) ^a
Skin change	1.84 (0.82) ^b	2.02 (0.81) ^{ab}	1.49 (0.66)	1.58 (0.68)
Voice change	1.80 (0.81) ^b	1.90 (0.81) ^b	1.19 (0.46)	1.34 (0.61) ^a
Facial hair	1.54 (0.70) ^b	1.67 (0.72) ^b	1.15 (0.40)	1.27 (0.56)
Overall Mean				
PDS	1.87 (0.54) ^b	1.92 (0.54) ^{ab}	1.48 (0.40)	1.61 (0.51) ^a
Girls				
Growth spurt	2.28 (0.83)	2.48 (0.86) ^a	2.44 (0.80)	2.68 (0.65) ^a
Body hair	1.99 (0.81)	2.55 (0.74) ^a	1.99 (0.91)	2.58 (0.79) ^a
Skin change	2.07 (0.85)	2.42 (0.76) ^a	1.78 (0.82)	2.25 (0.78) ^a
Breast growth	2.24 (0.69)	2.62 (0.56) ^a	2.21 (0.78)	2.64 (0.58) ^a
Menarche	1.30 (0.90)	1.90 (1.38) ^a	1.30 (0.90)	1.85 (1.35) ^a
Overall Mean				
PDS	1.98 (0.56)	2.39 (0.59) ^a	1.95 (0.61)	2.40 (0.61) ^a

^aSignificantly higher than younger, same sex group ($p < 0.05$) by post-hoc t tests.

^bSignificantly higher than parents for same grade and sex ($p < 0.001$) by post-hoc t tests.

are also shown in Table 3. Thus, within grades, the boys' self ratings in the present study were consistently higher than those given by their parents [5th-grade boys, $F(1, 136) = 56.53, p < 0.001$; 6th-grade boys, $F(1, 119) = 65.31, p < 0.001$]. By contrast, the mean item scores for 5th- and 6th-grade girls' self-ratings and those of their parents did not differ [5th-grade girls, $F(1, 148) = 0.80, NS$; 6th-grade girls,

$F(1, 165) = 0.17, NS$]. The data were next examined for changes from 5th to 6th grades. For boys, the parent scores showed only a slight increase from grade 5 to grade 6 [$F(1, 255) = 5.08, p < 0.05$]; this transition was more marked for girls [$F(1, 313) = 43.29, p < 0.001$].

Correlations—Students, Parents, Teachers

Table 4 provides Spearman rank-order correlations between item scores, PDS, and Puberty Category Scores for students versus parents, students versus teachers, and parents versus teachers. The individual item correlations for the 5th-grade boys and their parents were lowest, showing least consistency from this pair of raters. Nevertheless, a small but statistically significant correlation was found for several individual items and for the overall mean puberty score in the 5th-grade boys and their parents (Spearman $r = 0.243; df = 137; p < 0.01$). All correlation coefficients for the older boys and their parents were higher than in 5th-grade boys and all achieved statistical significance ($p < 0.001$). All individual item and PDS correlations between girls and their parents also achieved statistical significance. Significant correlations were found between parent and child Puberty Category Scores (see calculations in footnote to Table 1) for all groups. For comparison with teachers, the Puberty Category Scores of parent and child were collapsed to yield 3-point scores comparable to the teacher ratings. No significant relationships were found in the 5th-grade boys, although significant relationships were found in all the other groups

Table 4. Spearman Correlations of Individual Items, Mean Scores, and Category Scores Among Students, Parents, and Teachers

Item	Boys		Item	Girls	
	5th	6th		5th	6th
Student versus Parent					
Growth spurt	0.127	0.568***	Growth spurt	0.505***	0.546***
Body hair	0.210*	0.359***	Body hair	0.496***	0.541***
Skin change	0.437***	0.491***	Skin change	0.506***	0.598***
Voice change	0.296***	0.325***	Breast growth	0.532***	0.487***
Facial hair	0.144	0.352***	Menarche	0.965***	0.915***
Mean PDS score	0.243*	0.482***	Mean PDS score	0.712***	0.804***
Category score	0.225**	0.382***	Category score	0.704***	0.824***
Student versus Teacher					
3-point category	0.097	0.238**	3-point category	0.441***	0.261***
Parent versus Teacher					
3-point category	-0.051	0.261**	3-point category	0.415***	0.302***

PDS, Pubertal Development Scale.

* $p < 0.02$, ** $p < 0.01$, *** $p < 0.001$, by Spearman correlation.

Table 5. Distribution of Puberty Category Scores

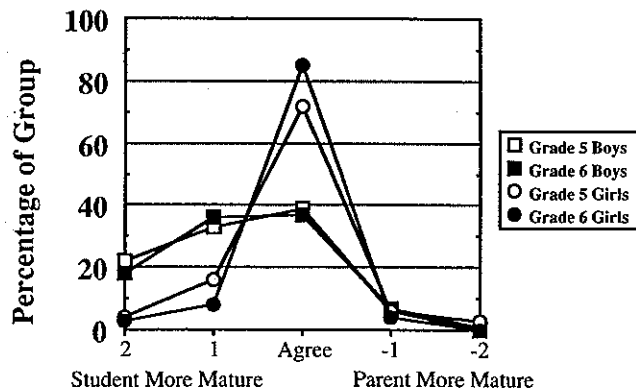
Category	Students		Parents		Teachers	
	5th	6th	5th	6th	5th	6th
Percentage of boys						
1 Prepubertal	19	11	58	41	69	70
2 Early pubertal	40	37	31	41		
3 Midpubertal	35	44	9	12	30	29
4 Late pubertal	5	8	1	5		
5 Postpubertal	1	0	0	0	1	1
Percentage of girls						
1 Prepubertal	11	1	16	3	47	39
2 Early pubertal	21	7	24	9		
3 Midpubertal	58	63	50	60	51	56
4 Late pubertal	10	29	9	27		
5 Postpubertal	0	1	0	1	2	4

among children, parents, and teachers, and were generally higher in the girls than boys.

Distribution of Puberty Category (Tanner) Scores

Table 5 provides the distribution of Puberty Category Scores for the groups, as well as the distributions for the 3-point teacher scale. The students generally rated themselves more mature than their parents rated them, particularly the boys. Figure 1 illustrates the agreement of Puberty Category Scores

Figure 1. Agreement between students and parents for Puberty Category Scale Score is shown in this figure. Percentage of each group is shown for five levels of agreement/disagreement between parent and student. If the Puberty Category Score was identical for student and parent, they were said to agree. If the Student Puberty Category Score was higher (more mature) than the parent, then the level was 1 or 2 to indicate that the student score was one or two scale values higher. If the Student Puberty Category Score was lower (less mature) than the parent, then the level was -1 or -2 to indicate that the student score was one or two scale values lower. Thus, scores to the left of "Agree" indicate that the student rated himself or herself as more mature than the parent; scores to the right indicate that parents rated the students as more mature.



between parents and children. For the girls, the agreement was quite good, whereas the boys' Puberty Category Scores were consistently more mature than their parents' ratings. The most numerous differences occurred when boys' Puberty Category Scores were 2 (early pubertal) versus 1 (prepubertal) for parents and when boys' scores were 3 (midpubertal) versus 2 for parents. Nevertheless, as mentioned previously, the correlation of Puberty Category Scores between parents and sons was statistically significant both in 5th (Spearman $r = 0.225$; $df = 144$; $p < 0.01$) and 6th (Spearman $r = 0.382$; $df = 121$; $p < 0.001$) graders.

Discussion

The validity of this self-administered rating scale for puberty in students is supported by the pretest comparing the self-ratings to physician ratings. Correlations between the student and pediatrician ratings were strong, and only 7 of the 38 adolescents rated themselves as much as one stage different from the pediatrician. Six of these discordant ratings were from boys, a sex difference consistent with findings from other studies. Although staff were present to administer the scale in the pretest, the students did not ask questions while completing the form. These results, along with the results from the larger survey, support the validity of this self-rating scale, even in the absence of an interviewer to answer questions.

The validity of the scales for survey research is further indicated by a number of consistent patterns within the results. Fifth-grade students rated themselves and were rated by parents as less mature than 6th graders. The 6th-grade girls were consistently rated more mature than boys of the same age, which is consistent with normal pubertal development (1). This pattern was true for all three raters (self, parent, teacher). Another indication of the viability of these measures comes from the significant correlations between students and parents for the PDS scores and Puberty Category Scores. Correlations were particularly high for girls, which is reasonable because the maturational scale scores in these youngsters were inherently more variable due to a greater range of developmental stages than in the less mature boys.

Results from the survey demonstrated reasonable reliability for both the student and parent versions of the puberty scales. The α coefficients are quite good for a 5-item scale and demonstrate that students and parents were able to answer the items with a reasonable degree of consistency.

Limitations of the present study include the overall response bias of the sampling procedure and the lack of an on-site researcher to answer children's questions. The children who returned their forms were rated by teachers as somewhat more socially mature, better functioning in school, and more alert and engaged than were students who failed to return forms. The failure of lower-functioning students to return survey forms may have been related to the difficulty of the questions, although they may not have returned forms for a variety of other reasons, such as an inability to get them home to parents. These issues are likely to be relevant to any off-site surveys of school children, but may be more important for certain types of research questions, such as physical development. The results of this survey, however, as well as those from another recent study (9), indicate that self-assessment scales may be quite useful in studies requiring approximations of sexual maturation. Of course, assessment of response bias is always important.

Another issue that arises using this type of self-administered rating scale has to do with missing values. As noted, a large proportion of the target sample (43% of boys and 46% of girls in 5th grade; 37% of 6th graders) was not included in the analyses because these students had "missing data" for one or more of the five physical development items. Most of these instances of missing data resulted from students marking the response that they did not know whether the event had taken place. Rewording the items may reduce the extent to which children are unable to answer. On the other hand, if we use only the three items involved in constructing the Puberty Category Score (for boys—body hair, voice change, facial hair; for girls—body hair, breast development, menarche), the percentage of missing data falls to 20% in boys and only 10% in girls. The issue of students not knowing whether such changes have taken place is greater for younger children, while older students seemed to be better able to complete the items effectively. This difficulty will always be a problem for survey studies in youngsters.

All scales, particularly for boys, are likely to be more successful with an older population who may be better able to comprehend the items and in whom maturational signs are more marked. Nevertheless,

even in older boys, changes may simply be more difficult to pinpoint than in girls, especially lacking the clearcut marker of menstruation.

Teachers' physical development ratings demonstrated the lowest correspondence to student and parent ratings. Teachers were given only three options for rating the students' development, and it is possible that adding a fourth choice to indicate "beginning" signs of physical development might improve the correspondence. Nevertheless, teacher ratings were correlated with student and parent ratings for all but 5th-grade boys.

In summary, we feel that these scales are useful measures of maturational status in settings where direct examinations, interviews, or pictorial representations are not possible.

We thank Anita Cavallo, M.D. for assessing Tanner stages on the pretest group, Kate B. Herman and Cecilia Vieira for help with survey compilation, Avi Sadeh, Sc.D. for comments on the manuscript, and the teachers, parents, and students who completed the surveys. This research was supported by NIMH grant MH45945 and a grant from Abbott Laboratories.

References

1. Tanner JM. Growth at Adolescence. Oxford: Blackwell, 1962.
2. Duke PM, Litt IF, Gross RT. Adolescents' self-assessment of sexual maturation. *Pediatrics* 1980;66:918-20.
3. Morris NN, Udry JR. Validation of a self-administered instrument to assess stage of adolescent development. *J Youth Adolesc* 1980;9:271-81.
4. Petersen AC. The Early Adolescence Study: An overview. *J Early Adoles* 1984;4:103-6.
5. Petersen AC, Crockett L, Richards M, et al. A self-report measure of pubertal status: Reliability, validity, and initial norms. *J Youth Adolesc* 1988;17:117-33.
6. Crockett LJ. Pubertal Development Scale: Pubertal categories. Unpublished manuscript 1988. (Copy available upon request from Prof. Anne Petersen, University of Minnesota Graduate School, 322 Johnston Hall, 101 Pleasant St. SE, Minneapolis, MN 55455.)
7. Carskadon MA, Acebo C. Parental reports of seasonal mood and behavior changes in children. *J Am Acad Child Adolesc Psychiatry* 1993 (in press).
8. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297-334.
9. Schlossberger NM, Turner RA, Irwin CE. Validity of self-reports of pubertal maturation in early adolescents. *J Adolesc Health* 1992;13:109-13.