Child Development, March/April 2009, Volume 80, Number 2, Pages 327–337

Pubertal Development: Correspondence Between Hormonal and Physical Development

Elizabeth A. Shirtcliff
University of New Orleans

Ronald E. Dahl
University of Pittsburgh School of Medicine

Seth D. Pollak
University of Wisconsin-Madison

Puberty is advanced by sex hormones, yet it is not clear how it is best measured. The interrelation of multiple indices of puberty was examined, including the Pubertal Development Scale (PDS), a picture-based interview about puberty (PBIP), and a physical exam. These physical pubertal measures were then associated with basal hormones responsible for advancing puberty. Participants included 160 early adolescents (82 boys). Puberty indices were highly correlated with each other. The physical exam stages correlated well with boys' and girls' testosterone and dehydroepiandrosterone and less so with girls' estradiol. The PDS and PBIP were similarly related to basal hormones. Self-report may be adequate when precise agreement is unnecessary. Multiple measures of puberty are viable options, each with respective strengths.

Adolescence constitutes a transition between childhood and adulthood whose onset includes pubertal maturation. Puberty has important implications for the development of regulatory competence and many aspects of physical, emotional, cognitive, and social development, including decision making and mental health (Steinberg et al., 2006). For these reasons, biobehavioral researchers increasingly seek to examine measures of puberty to clarify studies of emotion-related neural circuitry (Nelson, Leibenluft, McClure, & Pine, 2005; Sisk & Foster, 2004), psychopathology (Angold & Worthman, 1993; Cyranowski, Frank, Young, & Shear, 2000), cognition (Steinberg, 2005), and behavioral changes (Carskadon, Acebo, Jenni, Dahl, & Spear, 2004; Steinberg, 2000). Yet, it is not clear how to best evaluate pubertal development (Brooks-Gunn, Warren, Rosso, & Gargiulo, 1987). Here, we compare several measures of pubertal maturation, including hormonal indices. The main hormones responsible for advancing secondary sexual characteristics were captured by measuring testosterone and dehydroepiandrosterone (DHEA), two androgens that facilitate masculine development, and estradiol, an estrogen that facilitates feminine development. We evaluated agreement between physical exam and different methods of self-report, the associations between hormones and the physical exam, and the extent to which self-report methods led to parallel relations with hormonal measures as did the physical exam.

Physical Measures of Puberty

Nearly five decades ago, Tanner (1962) described five stages of puberty, ranging from 1 (*no development*) to 5 (*adult development*). These stages capture visible secondary sexual characteristics such as breast/genital development and pubic hair growth. Since its introduction, the gold standard for measuring pubertal status has been a physical exam conducted by a clinician employing Tanner's

This project was supported by grants from the National Institute of Mental Health to Elizabeth Shirtcliff (K01 MH077687), Ronald Dahl (R24 MH67346), and Seth Pollak (R01 MH61285, R01 MH068858). Infrastructure support was provided by P30-HD03352 (M. Seltzer, Center Director). Special thanks are due to the Pediatric Nurse Practitioners for their diligence and commitment to this research project: Marie Heilegenstein, R.N.; Lois Hoornstra, R.N.; and Kim Squires, R.N. We appreciate the research assistance of Patrick Bauer, Aaron Cohn, Johnna Dorshorst, Jamie Hanson, Chastity Jensen, and Abby Noack. Invaluable guidance was provided by Lorah D. Dorn and Elizabeth J. Susman throughout the study, including the use of the script for the interview measure. Finally, we thank the adolescents and their families who made this research possible.

Correspondence concerning this article should be addressed to Elizabeth A. Shirtcliff, Department of Psychology, University of New Orleans, 2000 Lakeshore Drive, New Orleans, LA 70148. Electronic mail may be sent to birdie.shirtcliff@uno.edu.

methods (Dorn, Dahl, Woodward, & Biro, 2006). Yet, researchers often find it difficult to integrate physical exams into nonclinical settings, and the Tanner stages measure only one dimension of development—external signs of physical development. To address these problems, the Pubertal Development Scale (PDS) asks adolescents to answer less invasive questions about puberty without mapping directly onto Tanner stages (Petersen, Crockett, Richards, & Boxer, 1988). The kappa (κ) concordance between the physical exam and PDS is only .24, however (Brooks-Gunn et al., 1987).

An alternative self-report method maps directly onto Tanner stages. Adolescents examine photographs or line drawings of models at each Tanner stage and indicate which image they most closely resemble (Morris & Udry, 1980). Although it is easy to administer, there is only moderate agreement between the physical exam and various versions of self-reported Tanner stage, with an average κ around .50 (see review by Coleman & Coleman, 2002; as well as more recent work by Desmangles, Lappe, Lipaczewski, & Haynatzki, 2006; Hergenroeder, Hill, Wong, Sangi-Haghpeykar, & Taylor, 1999; Schmitz et al., 2004). A few frequently cited studies, however, report excellent agreement (κs above .70; Boas, Falsetti, Murphy, & Orenstein, 1995; Carskadon et al., 1980; Norris & Richter, 2005). Agreement between self-reported PDS and self-reported Tanner stage was also moderate, κ = .50 (Bond et al., 2006). In addition to examining the agreement between multiple puberty measures, we then also explored whether certain factors predicted which adolescents had low agreement between puberty measures.

Hormonal Measures of Puberty

Although steroid hormones advance pubertal maturation, there is an imperfect match of hormones with physical measures of puberty for several reasons. Hormone levels change across the day, there are individual differences in hormone concentrations necessary to advance puberty, and there is overlap in hormone levels across each pubertal stage (Dawes et al., 1999). Nevertheless, knowledge about underlying hormonal processes provides information about pubertal maturation not available from overt physical measures alone.

The earliest peripheral sign of puberty occurs when androgens begin to be released gradually from the adrenal gland (Palmert et al., 2001). DHEA and other adrenal androgens cause pubic hair growth, body odor, acne, and prepubertal growth (Havelock, Auchus, & Rainey, 2004; Lucky, Biro, Simbartl, Morrison, & Sorg, 1997). Adrenal androgens increase twofold in boys from when they show no pubertal development to when they reach adultlike development (Biro, Lucky, Huster, & Morrison, 1995). Puberty shows moderate correlations with DHEA in both sexes (Shirtcliff, Zahn-Waxler, Klimes-Dougan, & Slattery, 2007), although another study failed to detect a relation in boys or girls (Maskarinec et al., 2005).

Testosterone, the primary androgen released from the gonads (Rubinow & Schmidt, 1996), causes genital development in males (Hiort, 2002). Boys with delayed puberty show rapidly advancing pubertal maturation when administered testosterone (Finkelstein et al., 1999; Geller, Rogol, & Knitter, 1983). Testosterone is approximately 45 times higher by adulthood as compared to prepubertal development in boys (Biro et al., 1995), but the rise is smaller in girls (Legro, Lin, Demers, & Lloyd, 2000). Puberty and testosterone are highly correlated in boys, but no clear association is evident in girls (Granger et al., 2003; Maskarinec et al., 2005).

In contrast, estradiol is the primary estrogen released from the gonads and other peripheral tissues (Fernandez-Garcia et al., 2002). Estradiol causes breast development, encourages female-typical fat distributions and long bone fusion during growth spurts, and helps stimulate ovulation and menstruation (Frank, 2003; MacGillivray, Morishima, Conte, Grumbach, & Smith, 1998). Girls with delayed puberty show advancing pubertal maturation when administered estradiol (Finkelstein et al., 1999; Rosenfield et al., 2005). Estradiol is 4–9 times higher in late adolescent girls as compared to childhood (Ikegami et al., 2001). Although more hormonal signals are involved, measuring these three hormones (testosterone, DHEA, and estradiol) should provide converging information about gonadal and adrenal hormonal signals of puberty, two distinct components of maturation in early adolescents. The goal of the current study was to examine how well different physical pubertal stages corresponded with each other and, in addition, how these measures captured basal hormones.

Method

Participants

Participants were 82 boys and 78 girls recruited from the community through an existing laboratory registry and local advertisements. Adolescents ranged from 9 through 14 years of age

(M = 11.2 years), capturing early adolescence when pubertal stage is most variable. Exclusion criteria included use of allergy or asthma medication. Participants were from diverse backgrounds; Hollingshead scores spanned the full socioeconomic gradient (M = 41.9, SD = 15.6, range = 5–66). Forty-eight percent of participants were White, 26% were Black, and 26% were Asian, Hispanic, Mixed, or unspecified. Body mass index (BMI) was 21.8 on average (SD = 9.3), with 21% of the sample above the 95th percentile of BMI for age (overweight) and 2% below the 5th percentile (underweight).

Procedures

Adolescents and their parent(s) provided informed assent and consent, respectively. Adolescents completed the self-report measures, and then a pediatric nurse practitioner (PNP) conducted the physical exam. Participants provided saliva throughout the laboratory day and were sent home with supplies for additional saliva collection. Procedures were approved by the University of Wisconsin Institutional Review Board.

Measures

PDS. Adolescents completed the five PDS questions about physical development, scored from 1 (*no*) to 4 (*development seems complete*; Petersen et al., 1988). Reliability of the PDS was high (α = .77 for boys, α = .81 for girls). Few (3%) adolescents had missing PDS scores. We developed a coding system to convert the PDS to a 5-point scale in order to parallel the physical exam Tanner stages (available upon request). Although interrelated, puberty is not a single event. Therefore, our coding system differentially captured gonadal and adrenal hormonal signals of physical development. In girls, growth spurt, breast development, and menarche are associated with gonadal hormonal signals. In boys, growth spurt, deepening of voice, and facial hair growth are associated with gonadal hormones. For both sexes, pubic/body hair and skin changes are associated with adrenal hormones.

Picture-Based Interview About Puberty (PBIP). In a comfortable room, a research assistant spoke with adolescents about ''changes that happen when you grow up'' with the assistance of a script and photographs (Dorn & Susman, 2002). Following this discussion, researchers left the room while adolescents reported their assessment of pubertal stage. Female researchers interviewed girls and male researchers interviewed half of the boys. There was no difference in accuracy of staging based on the sex of the interviewer, $p > .29$.

Physical exam. Adolescents were given the option to wear a hospital gown or loose clothing for the exam. Height and weight were measured and later used to calculate BMI (age-corrected using Centers for Disease Control and Prevention CDC guidelines). Experienced PNPs were trained to conduct exams for research purposes by the second author. PNPs inspected breast development with brief palpation for girls and visually examined pubic hair. An orchidometer measured testicular size in boys (Genentech, 1997) along with visual inspection of genitals and pubic hair. Interobserver reliability (N = 10 exams, 6.3%) was good, κ = .88. Thirteen percent of participants refused the exam, but assented to self-report measures. Those who refused the exam did not differ in age, race, BMI, or stage; there was a trend for boys (N = 14) to refuse more than girls (N = 6), $\chi^2(1) = 3.2$, $p = .07$.

Hormonal measures. Adolescents provided eight saliva samples across the laboratory day, beginning immediately after providing informed consent (M = 9:38, SD = 1:27 hr) through bedtime (M = 21:04, SD = 1:32 hr). Saliva was collected by passive drool (Shirtcliff, Granger, Schwartz, & Curran, 2001). At the time of each sample, participants completed a short diary, which has been used previously (Granger et al., 2003). Samples were immediately frozen at −80°C until they were aliquotted to minimize freeze/thaw cycles. Participants were also sent home with supplies for home collection. To capture the full day, participants collected six samples on each of 4 days at prespecified times between waking (M = 7:45, SD = 1:19 hr) and bedtime (M = 21:17, SD = 1:13 hr), prior to mealtimes. To ensure compliance, cryuvials were stored in time-locked caps (Aardex, Zug, Switzerland), which recorded collection time and date. Samples were stored in home freezers until all samples were collected, then the batch was shipped overnight on ice and stored at −80°C (Dabbs, 1991). All 32 samples were assayed for testosterone and DHEA. One morning sample from each day of saliva collection was assayed for estradiol in girls (M = 9:39, SD = 0:51 hr). Salivary estradiol is not valid in boys (Shirtcliff et al., 2000).

Hormone determination. Enzymeimmunoassays were completed by Madison Biodiagnostics (Madison, WI) using Salimetrics' kits (State College, PA). Samples were measured in duplicate; duplicates that varied by more than 7% were repeat tested. For DHEA, the range of sensitivity was from 5 to 1,000 pg/ml. The average intra-assay coefficient of

variation (CV) was 5.6% and the average interassay CV was 8.2%. For testosterone, the range of sensitivity was from 1 to 600 pg/ml. Average intra- and interassay CVs were 4.6% and 8.3%, respectively. For estradiol, the range of sensitivity was 1–32 pg/ml. Average intra- and interassay CVs were 7.1% and 7.5%, respectively.

A basal hormone measure was calculated using hierarchical linear modeling, which separates within-the-day and day-to-day variation in hormone levels (N = 3,704) from individual basal levels (N = 160), thereby allowing removal of the effects of several important control variables and accurate aggregation across repeated measures of each hormone to a single basal level. At the within-the-day level, we controlled for linear and quadratic time since waking (in minutes) and time of day to account for the individual's intrinsic and extrinsic rhythm. These values were allowed to vary so that each individual had his or her own rhythm removed from the basal estimate. An empirical Bayes estimate of each log-transformed hormone was extracted after accounting for additional control variables at both within-the-day and day-to-day levels (e.g., flow rate, response to awakening, location, medication usage, exercise, emotion, who the child was with). Basal DHEA comprised 73.8% of the total variation in DHEA, $p < .0001$. Basal testosterone comprised 80.0% of the variation in testosterone, $p < .0001$. For estradiol, an average across the five samples/individual was calculated as analyses revealed no day-to-day predictors of estradiol (including controls above as well as menstrual cycle day-count, cycle regularity, menarcheal status). Basal estradiol comprised 36% of the variation in estradiol, $p < .0001$. Less variation in estradiol was basal than the other hormones, perhaps due to the reduced number of samples.

Statistical analyses. Kappas and percent accuracy described precise agreement between the three puberty measures. Pearson correlations examined whether measures were associated, without necessitating precision. To examine which adolescents were inaccurate informants, we calculated the discrepancy between the exam and the PDS and PBIP, respectively, using a difference score. We assessed whether gender, age, stage, BMI, or race influenced accuracy of adolescents' self-report using linear regression. Race was coded as White, Black, or Other. Structural equation modeling (SEM) simultaneously examined how the physical exam was associated with steroid hormones, with separate models for boys and girls (because estradiol was measured in girls only). The physical exam was first modeled with basal hormones, removing nonsignificant coefficients. Poor model fit was indicated by significant χ^2 values, comparative fit index (CFI) less than .95, or root mean square error of approximation (RMSEA) greater than .10. Next, parallel models were fit substituting the respective self-report measures. To test whether models were parallel to the exam, we fixed coefficients to be identical to the physical exam and examined the reduction in model fit compared to when coefficients were unconstrained. If the indices of practical fit were too high/low or the χ^2 was significant (indicating models were not parallel), we removed constraints on coefficients which resulted in the greatest model improvements.

Results and Discussion

How Did Self-Report PDS Map Onto the Physical Exam?

Correlations between the physical exam and the PDS are presented in Table 1. The concordance between the physical exam and the PDS gonadal stage was modest, $\kappa = .36$, $\chi^2(16) = 93.0$, $p < .0001$ (Table 2). Accuracy was defined as self-report of the same stage as the physical exam. Fifty-two percent of adolescents' gonadal scores were accurate (54% boys, 47% girls), whereas 18% overestimated (15% boys, 27% girls) and 30% underestimated stage (31% boys, 27% girls) compared to the exam. The concordance between the physical exam and the PDS adrenal stage was also modest, $\kappa = .36$, $\chi^2(16) = 90.6$, $p < .0001$ (Table 3). Fifty percent of adolescents were accurate (60% boys, 44% girls), whereas 29% underestimated (26% boys, 34% girls) and 21% overestimated pubic hair (14% boys, 23% girls).

How Did the PBIP Map Onto the Physical Exam?

The concordance between the physical exam and PBIP breast/genital stage was modest, $\kappa = .36$, $\chi^2(16) = 120.9$, $p < .0001$ (Tables 1 and 2). Forty-nine percent of adolescents reported the same breast/genital stage as the exam (41% boys, 57% girls), whereas 26% overestimated (35% boys, 17% girls) and 25% underestimated stage (24% boys, 17% girls). The parallel concordance for pubic hair was good, $\kappa = .43$, $\chi^2(16) = 137.2$, $p < .0001$ (Table 3). Fifty-six percent of adolescents reported the same pubic stage as the exam (54% boys, 58% girls), whereas 24% overestimated (26% boys, 21% girls)

Table 1
Intercorrelations of the Physical Exam, Picture-Based Interview About Puberty (PBIP), and Pubertal Development Scale (PDS) With Girls Above (in Italics) and Boys Below (in Bold)

	Physical exam		PBIP		PDS	
	Breast/genital	Pubic hair	Breast/genital	Pubic hair	Gonadal	Adrenal
Physical exam						
Breast/genital		*.85*	*.83*	*.76*	*.65*	*.65*
Pubic hair	**.93**		*.75*	*.88*	*.69*	*.71*
PBIP						
Breast/genital	**.60**	**.60**		*.79*	*.77*	*.72*
Pubic hair	**.69**	**.71**	**.71**		*.73*	*.81*
PDS						
Gonadal	**.65**	**.63**	**.59**	**.68**		*.72*
Adrenal	**.63**	**.68**	**.70**	**.70**	**.65**	
Mean (*SD*) stage						
Boys	**2.4 (1.3)**	**2.3 (1.3)**	**2.7 (1.2)**	**2.3 (1.2)**	**2.0 (1.1)**	**2.3 (1.2)**
Girls	*2.9 (1.5)*	*2.8 (1.5)*	*2.9 (1.2)*	*2.7 (1.4)*	*2.5 (1.3)*	*2.9 (1.3)*

Note. All intercorrelations have *p*s < .001. There are no mean differences in staging between the physical exam and PBIP, *p*s > .2, or PDS, *p*s > .06.

Table 2
Concordance of the Physical Exam Breast/Genital Stage, Pubertal Development Scale (PDS), and the Picture-Based Interview About Puberty (PBIP)

	Physical exam breast/genital stage[a]					
A. PDS gonadal score	I	II	III	IV	V	Total[b]
I.	61.1	28.1	10.3	5.0	5.3	36
II.	19.4	53.1	20.7	15.0	5.3	34
III.	16.7	18.8	51.7	30.0	21.1	37
IV.	2.8		6.9	30.0	26.3	14
V.			10.3	20.0	42.1	15
Total (*N*)[b]	36	32	29	20	19	136
B. PBIP	Physical exam breast/genital stage[a]					
I: No development	54.1	18.2	3.4	4.8		28
II: Breast bud	24.3	48.5	20.7	14.3		34
II: Testes started to grow						
III: Breast tissue beyond areola	18.9	30.3	48.3	19.0	4.8	36
III: Penis growth in length						
IV: Areola second mound on breast	2.7	3.0	17.2	61.9	61.9	33
IV: Penis growth in width and length						
V: Adultlike development			10.3		33.3	10
Total (*N*)[b]	37	33	29	21	21	141
C. PBIP	PDS gonadal score[a]					
I.	51.3	18.4	4.3			29
II.	33.3	44.7	13.0	13.3	5.9	39
III.	12.8	28.9	41.3	26.7	11.8	41
IV.	2.6	7.9	32.6	53.3	52.9	36
V.			8.7	6.7	29.4	10
Total (*N*)[b]	39	38	46	15	17	155

[a]Column percentages. [b]Number of participants.

Table 3
Concordance of the Physical Exam Pubic Hair Stage, Pubertal Development Scale (PDS), and the Picture-Based Interview About Puberty (PBIP)

	Physical exam pubic hair stage[a]					
A. Self-report (PDS)	I	II	III	IV	V	Total[b]
I.	77.8	25.8	18.2	3.8		48
II.	17.8	51.6	27.3	23.1	16.7	38
III.	4.4	19.4	31.8	26.9	25	25
IV.		3.2	18.2	34.6	33.3	18
V.			4.5	11.5	25	7
Total (*N*)[b]	45	31	22	26	12	136
B. PBIP	Physical exam pubic hair stage[a]					
I: No development	73.9	25	9.1	3.7		45
II: Sparse wispy strands	19.6	50	22.7	3.7		31
III: Darker, courser hair	6.5	21.9	36.4	14.8	14.3	24
IV: Course hair along most of pubis		3.1	31.8	55.6	35.7	28
V: Adultlike development, hair extends to upper thighs				22.2	50.0	13
Total (*N*)[b]	46	32	22	27	14	141
C. PBIP	PDS adrenal score[a]					
I.	66.1	26.8	3.6			49
II.	26.8	36.6	14.3	9.1		36
III.	7.1	24.4	32.1	18.2	12.5	28
IV.		12.2	39.3	63.6	12.5	31
V.			10.7	9.1	75.0	11
Total (*N*)[b]	56	41	28	22	8	155

[a]Column percentages. [b]Number of participants.

and 20% underestimated stage (19% boys, 21% girls).

How Did Self-Report PDS Map Onto the PBIP?

Correlations between the two self-report measures are reported in Table 1. The concordance between the PDS gonadal stage and the breast/genital PBIP stage was low, $\kappa = .29$, $\chi^2(16) = 98.4$, $p < .0001$ (Table 2). Forty-five percent of adolescents reported the same breast/genital PBIP stage as the PDS gonadal score (37% boys, 52% girls). The parallel concordance for pubic hair was moderate, $\kappa = .37$, $\chi^2(16) = 152.1$, $p < .0001$ (Table 3). Fifty-two percent of adolescents reported the same pubic stage on the PBIP as the PDS (47% boys, 57% girls).

The physical exam is well suited for a wide range of behavioral endocrinology-oriented questions or when an objective measure of physical development is desirable (Dorn et al., 2006). Though precise agreement was modest, the two self-report measures were correlated with the physical exam, suggesting they mutually capture underlying pubertal processes. If precision is not necessary, adolescents are relatively good observers. We should note that 13% of adolescents refused the physical exam, but none refused the PBIP. Researchers conducting an exam might consider supplementation with a self-report measure to capture this subset.

Which Adolescents Were Inaccurate Informants?

Using linear regression, where the discrepancy between the physical exam and PDS was the outcome, we found that neither sex nor BMI influenced accuracy, *p*s > .09. Stage (based on the exam) qualified an effect of age ($\beta = .32$, $p < .003$ for age; $\beta = -.68$, $p < .001$ for stage). In general, adolescents overestimated pubertal maturation when they were at lower stages of development relative to their peers and underestimated development when they were at higher stages than their peers. As expected, this distortion of staging was age specific. For example, 11- and 12-year-olds were accurate at Stage 3 but overestimated Stage 2 and underestimated Stages 4+; 13-; and 14-year-olds were

accurate at Stage 4, but tended to overestimate Stages 2 and 3 and underestimate Stage 5. This may reflect the desirability of adolescents to appear like the developmental stage that is most typical for their age. White adolescents overestimated stage more often than non-Caucasian adolescents ($\beta = .21$, $p < .02$).

Analyses of the discrepancies between the physical exam and PBIP yielded similar findings. Sex and BMI did not influence accuracy, ps > .15. Stage qualified the effect of age, ($\beta = .32$, $p < .003$ for age; $\beta = -.68$, $p < .001$ for stage), such that adolescents sometimes overestimated development when at lower stages and underestimated development when at higher stages. Again, adolescents tended to report stages that were most typical of their age. White or Black adolescents overestimated stage more often than other adolescents ($\beta = .19$, $p = .03$).

Young adolescents may not be able to self-report an exact stage—particularly if they are maturing earlier or later than their peers. Measurement problems may primarily affect studies that employ a cut-score to describe adolescents as pre- or postpubertal rather than as a continuous process; this may be especially problematic in research designed to isolate early and late maturing adolescents.

Was the Physical Exam Stage Associated With Hormones?

Table 4 presents hormone values across Tanner stages for boys and girls. Boys' basal testosterone and DHEA were predicted by the physical exam, with one exception. Genital development was not associated with DHEA; dropping this coefficient did not change model fit, $\chi^2(1) = 0.5$, $p = .46$, CFI > .999, RMSEA < .0001 (Figure 1A). For girls, breast development was not associated with testosterone or DHEA, and pubic hair was not associated with estradiol (Figure 2A). These three coefficients were dropped without reducing the goodness of fit, $\chi^2(3) = 1.75$, $p = .63$, CFI > .999, RMSEA < .0001. In sum, the physical exam captured basal testosterone and DHEA well in both sexes, though estradiol was modestly related to the exam.

We were surprised that breast development explained such a small amount of variability in estradiol and await replication in a different (and perhaps older) sample or in which more of the variability was basal. Pubic hair was particularly good at capturing basal hormones in both sexes. The endocrine signaling of pubic hair generally begins earlier (between ages 6 and 9) and may be more established than breast/genital development in this early age range. When interested in comparing boys and girls, assessments of pubic hair may be emphasized, as they performed well in both sexes.

Was the Self-Report PDS Associated With Hormones Like the Physical Exam?

An identical model substituting the PDS stages for boys demonstrated marginal model fit, $\chi^2(4) = 9.15$, $p = .06$, CFI = .97, RMSEA = .13, but no single coefficient differed from the physical

Table 4
Hormone Levels (and Standard Errors) in Boys and Girls Across the Physical Exam Stages

	Boys		Girls		
	Testosterone	DHEA	Testosterone	DHEA	Basal estradiol[a]
Physical exam breast/genital stage					
I	15.90 (1.49)	34.77 (4.93)	18.66 (2.56)	34.59 (6.05)	4.36 (1.62)
II	20.25 (2.90)	42.42 (8.23)	19.41 (2.01)	47.09 (8.25)	3.60 (0.44)
III	29.87 (3.44)	66.51 (13.99)	26.93 (2.27)	76.82 (11.97)	3.81 (0.65)
IV	54.57 (8.13)	111.54 (22.07)	29.85 (2.23)	93.39 (13.97)	5.96 (1.57)
V[b]	50.87 (6.17)	51.67 (9.09)	32.12 (4.58)	123.03 (19.95)	3.81 (.57)
Physical exam pubic hair stage					
I	16.25 (1.44)	34.52 (5.06)	16.86 (1.90)	35.88 (8.48)	4.61 (1.51)
II	23.90 (3.57)	46.77 (8.23)	23.19 (2.21)	48.94 (5.92)	3.02 (0.34)
III	29.38 (4.15)	68.42 (15.42)	27.06 (3.16)	84.92 (13.56)	4.10 (0.78)
IV	53.23 (6.43)	102.10 (20.25)	29.41 (1.73)	102.51 (14.58)	5.58 (1.25)
V[b]	59.80 (3.78)	57.60 (15.52	34.26 (5.99)	123.51 (22.18)	3.85 (0.77)

Note. Average hormone levels (not basal) are presented to aid in comparison across studies. DHEA = dehydroepiandrosterone.
[a]Estradiol was measured in girls only. [b]Five boys were genital stage V; three were pubic stage V.

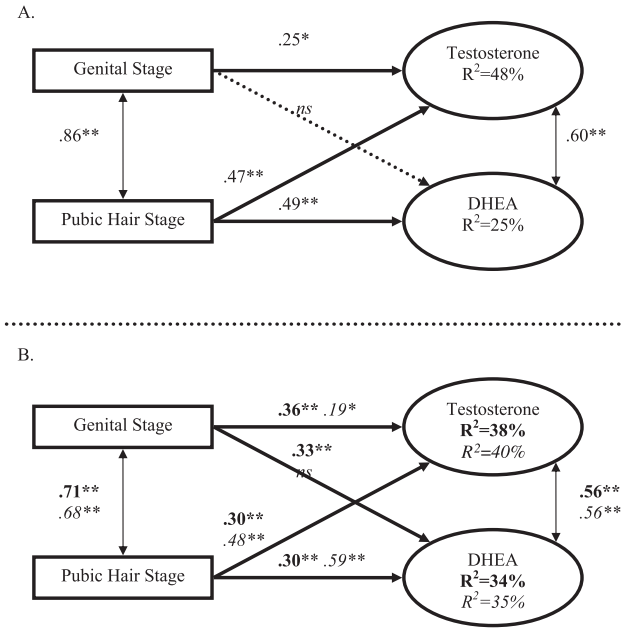

Figure 1. Structural equation model for boys.
Note. (A) Standardized β coefficients when testosterone and dehydroepiandrosterone (DHEA) are predicted by the physical exam genital and pubic hair stage. (B) Model parallel to (A) with testosterone and DHEA predicted by the Pubertal Development Scale stages (in italics) and Picture-Based Interview About Puberty stages (in bold).
*$p < .05$. **$p < .01$.

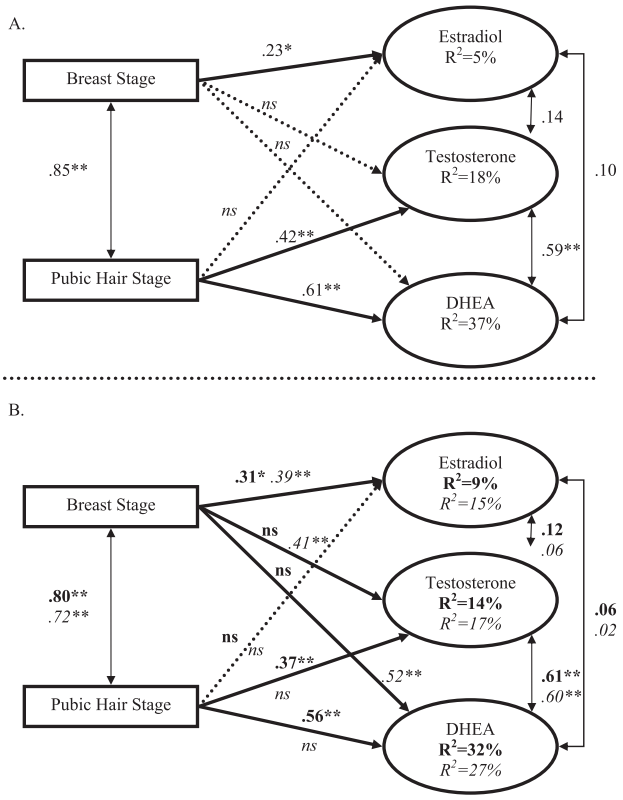

Figure 2. Structural equation model for girls.
Note. (A) Standardized β coefficients when testosterone, dehydroepiandrosterone (DHEA), and estradiol are predicted by the physical exam breast and pubic hair stage. (B) Model parallel to (A) with basal hormones predicted by the Pubertal Development Scale (in italics) and Picture-Based Interview About Puberty (in bold).
*$p < .05$. **$p < .01$.

exam model, *p*s > .15. Like the exam, gonadal development measured using the PDS did not predict DHEA, $p = .58$, but all other coefficients were significant. The PDS basically led to parallel relations with boys' basal hormones, as did the physical exam (Figure 1B).

For girls, an identical model substituting PDS stages for the exam resulted in marginal model fit, $\chi^2(6) = 10.7$, $p = .10$, CFI = .96, RMSEA = .10. Three coefficients were substantially different from the physical exam, accounting for most of the model misspecification, $\chi^2(3) = 8.4$, $p = .04$, CFI = .96, RMSEA = .15. Unlike the physical exam, the PDS gonadal score was associated with testosterone and DHEA, and the PDS adrenal score was not as highly related to DHEA as in the physical exam model. Constraining the remaining coefficients to be parallel to the physical exam did not reduce model fit, $\chi^2(3) = 4.16$, $p = .25$, CFI = .99, RMSEA = .07, indicating that the PDS gonadal score was more broadly related to basal hormones than the physical exam, whereas the adrenal score was less predictive of girls' basal DHEA (Figure 2B).

Given that the PDS is a common measure and is easily employed in a variety of settings (e.g., schools, screening mailers), it should be welcome news that this self-report measure captured basal hormones in parallel with the physical exam in boys; in girls, the gonadal score performed slightly better than the exam. That girls' adrenal score was not associated with basal hormones is perplexing because the adrenal score included items, such as body/pubic hair growth and skin changes, that are related to hormones like DHEA (Grumbach, 2002).

Because the basal hormones were complicated, all SEM analyses were recalculated using a simple average. These models fit similarly—the RMSEA was on average .05 different, the χ^2 never exceeded 2.3, and the percent of variance in each hormone explained by pubertal status differed on average by only 3.7%.

Was the PBIP Associated With Hormones Like the Physical Exam?

An identical model that substituted PBIP stages for the physical exam for boys demonstrated poor model fit, $\chi^2(4) = 13.4$, $p < .01$, CFI = .94, RMSEA = .17, indicating that PBIP was not parallel with the physical exam (Figure 1B). We allowed genital development to be related to DHEA and found it significantly predicted DHEA, $p = .009$. When the other three coefficients were constrained to be identical to the exam model, goodness of fit was excellent, $\chi^2(3) = 0.87$, $p = .8$, CFI > .999, RMSEA < .0001, suggesting that PBIP led to parallel relations with basal hormones as did the physical exam and, in addition, that the PBIP genital stage predicted DHEA better than the physical exam.

For girls, an identical model that substituted PBIP for the physical exam fit well, $\chi^2(6) = 3.96$, $p = .7$, CFI > .999, RMSEA < .0001, indicating that PBIP captured basal hormones in a similar manner as the physical exam (Figure 2B). In sum, the PDS and PBIP were related to basal hormones in parallel or occasionally better than the physical exam.

That the PBIP mapped onto basal hormones in parallel to the physical exam (or slightly better for boys' prediction of basal DHEA) is an additional advantage of the PBIP for potentially addressing hormone-related research questions. Use of the PBIP is most viable (a) when high correlations with an objective measure like the physical exam are sought after, (b) when the Tanner metric is desirable, and (c) when basal hormones (particularly in boys) are outcomes of interest or are proximally associated with outcomes of interest. Nevertheless, even the best measure of external pubertal status captured less than half of the variability in basal hormones. Directly measuring hormones is often feasible.

Although we were agnostic about which measure would be optimal, we were surprised that self-reported PDS and PBIP scores were occasionally better correlates with basal hormones than the exam. This may be due to the unique perspectives of clinicians and adolescents. Although clinicians have a range of knowledge comparing one adolescent to another, rarely do they observe the same adolescent across time. In contrast, adolescents have little experience with other individuals, yet they have daily insights into their own pubertal changes. The adolescents' perspective may be optimal for noticing changes in their bodies across months and years. Basal hormones likewise capture a gradual, continuous developmental process. Adolescents may generally be more attuned to the confluence of this internal process with external developmental changes. Choosing measures that encompass the subjective adolescent experiences may be suitable for many biopsychosocial research questions.

Limitations

Several limitations should be considered. First, although our study is ethnically and socioeconomically diverse, the sample size limited the extent to which we could explore individual differences, such as mechanisms behind racial differences in accuracy of self-report. Second, other hormones involved in pubertal maturation (e.g., DHEA-sulfate, androstenedione, and progesterone) could yield different associations with exam and self-report measures. Third, estradiol varies across the menstrual cycle. This limitation is noticeable because estradiol was weakly related to pubertal development. Girls were not recruited to come to the laboratory during a particular phase of their menstrual cycle because (a) estradiol begins to cycle years before girls' first menstruation (menarche), so this would reduce cycle effects in menarcheal but not premenarcheal girls; (b) after menarche, cycles are often irregular (48% of our girls), so it would be difficult to schedule by day-count; and (c) days in which girls were likely accurate (i.e., during menstruation) are when estradiol is at its nadir and least likely to differentiate early from late puberty (Dawes et al., 1999). More frequent or systematic estradiol measurement may reveal stronger associations with puberty than we found.

Conclusions

Here we reported different ways to measure pubertal development in early adolescence. Our broad goal was to understand the association between these measures so that researchers can be informed about which measure(s) best addresses particular research questions. Because puberty encompasses a suite of changes and is not a single process, different measures may best capture different aspects of adolescent development. The answer about which measure(s) is best may depend on which aspects of puberty are of interest for a particular study or research question. Inclusion of pubertal measures provides essential information about developmental changes in adolescence.

References

Angold, A., & Worthman, C. W. (1993). Puberty onset of gender differences in rates of depression: A developmental, epidemiologic and neuroendocrine perspective. *Journal of Affective Disorders*, *29*, 145–158.

Biro, F. M., Lucky, A. W., Huster, G. A., & Morrison, J. A. (1995). Pubertal staging in boys. *Journal of Pediatrics*, *127*, 100–102.

Boas, S. R., Falsetti, D., Murphy, T. D., & Orenstein, D. M. (1995). Validity of self-assessment of sexual maturation in adolescent male patients with cystic fibrosis. *Journal of Adolescent Health*, *17*, 42–45.

Bond, L., Clements, J., Bertalli, N., Evans-Whipp, T., McMorris, B. J., Patton, G. C., et al. (2006). A comparison of self-reported puberty using the Pubertal Development Scale and the Sexual Maturation Scale in a school-based epidemiologic survey. *Journal of Adolescence*, *29*, 709–720.

Brooks-Gunn, J., Warren, M. P., Rosso, J., & Gargiulo, J. (1987). Validity of self-report measures of girls' pubertal status. *Child Development*, *58*, 829–841.

Carskadon, M. A., Acebo, C., Jenni, O. G., Dahl, R. E., & Spear, L. P. (2004). Regulation of adolescent sleep: Implications for behavior. In R. E. Dahl & L. P. Spear (Eds.), *Adolescent brain development: Vulnerabilities and opportunities* (pp. 276–291). New York: New York Academy of Sciences.

Carskadon, M. A., Harvey, K., Duke, P., Anders, T. F., Litt, I. F., & Dement, W. C. (1980). Pubertal changes in daytime sleepiness. *Sleep*, *2*, 453–460.

Coleman, L., & Coleman, J. (2002). The measurement of puberty: A review. *Journal of Adolescence*, *25*, 535–550.

Cyranowski, J. M., Frank, E., Young, E., & Shear, M. K. (2000). Adolescent onset of the gender difference in lifetime rates of major depression: A theoretical model. *Archives of General Psychiatry*, *57*, 21–27.

Dabbs, J. M., Jr. (1991). Salivary testosterone measurements: Collecting, storing, and mailing saliva samples. *Physiology and Behavior*, *49*, 815–817.

Dawes, M. A., Dorn, L. D., Moss, H. B., Yao, J. K., Kirisci, L., Ammerman, R. T., et al. (1999). Hormonal and behavioral homeostasis in boys at risk for substance abuse. *Drug and Alcohol Dependence*, *55*, 165–176.

Desmangles, J. C., Lappe, J. M., Lipaczewski, G., & Haynatzki, G. (2006). Accuracy of pubertal Tanner staging self-reporting. *Journal of Pediatric Endocrinology and Metabolism*, *19*, 213–221.

Dorn, L. D., Dahl, R. E., Woodward, H. R., & Biro, F. (2006). Defining the boundaries of early adolescence: A user's guide to assessing pubertal status and pubertal timing in research with adolescents. *Applied Developmental Science*, *10*, 30–56.

Dorn, L. D., & Susman, E. J. (2002). *Puberty script: Assessment of physical development in boys and girls*. Cincinnati, OH: Cincinnati Children's Hospital Medical Center.

Fernandez-Garcia, B., Lucia, A., Hoyos, J., Chicharro, J. L., Rodriguez-Alonso, M., Bandres, F., et al. (2002). The response of sexual and stress hormones of male procyclists during continuous intense competition. *International Journal of Sports Medicine*, *23*, 555–560.

Finkelstein, J. W., D'Arcangelo, M. R., Susman, E. J., Chinchilli, V. M., Kunselman, S. J., Schwab, J., et al. (1999). Self-assessment of physical sexual maturation in boys and girls with delayed puberty. *Journal of Adolescent Health*, *25*, 379–381.

Frank, G. R. (2003). Role of estrogen and androgen in pubertal skeletal physiology. *Medical and Pediatric Oncology*, *41*, 217–221.

Geller, B., Rogol, A., & Knitter, E. (1983). Preliminary data on the dexamethasone suppression test in children with major depressive disorder. *American Journal of Psychiatry*, *140*, 620–622.

Genentech. (1997). *Assessment of pubertal stages* (G70577–R1#LF0050). San Francisco: Author.

Granger, D. A., Shirtcliff, E. A., Zahn-Waxler, C., Usher, B., Klimes-Dougan, B., & Hastings, P. (2003). Salivary testosterone diurnal variation and psychopathology in adolescent males and females: Individual differences and developmental effects. *Development and Psychopathology*, *15*, 431–449.

Grumbach, M. M. (2002). The neuroendocrinology of human puberty revisited. *Hormone Research*, *57*(Suppl. 2), 2–14.

Havelock, J. C., Auchus, R. J., & Rainey, W. E. (2004). The rise in adrenal androgen biosynthesis: Adrenarche. *Seminars in Reproductive Medicine*, *22*, 337–347.

Hergenroeder, A. C., Hill, R. B., Wong, W. W., Sangi-Haghpeykar, H., & Taylor, W. (1999). Validity of self-assessment of pubertal maturation in African American and European American adolescents. *Journal of Adolescent Health*, *24*, 201–205.

Hiort, O. (2002). Androgens and puberty. *Best Practice and Research Clinical Endocrinology and Metabolism*, *16*, 31–41.

Ikegami, S., Moriwake, T., Tanaka, H., Inoue, M., Kubo, T., Suzuki, S., et al. (2001). An ultrasensitive assay revealed age-related changes in serum oestradiol at low concentrations in both sexes from infancy to puberty. *Clinical Endocrinology*, *55*, 789–795.

Legro, R. S., Lin, H. M., Demers, L. M., & Lloyd, T. (2000). Rapid maturation of the reproductive axis during perimenarche independent of body composition. *Journal of Clinical Endocrinology and Metabolism*, *85*, 1021–1025.

Lucky, A. W., Biro, F. M., Simbartl, L. A., Morrison, J. A., & Sorg, N. W. (1997). Predictors of severity of acne vulgaris in young adolescent girls: Results of a five-year longitudinal study. *Journal of Pediatrics*, *130*, 30–39.

MacGillivray, M. H., Morishima, A., Conte, F., Grumbach, M., & Smith, E. P. (1998). Pediatric endocrinology update: An overview. The essential roles of estrogens in pubertal growth, epiphyseal fusion and bone turnover: Lessons from mutations in the genes for aromatase and the estrogen receptor. *Hormone Research*, *49*(Suppl. 1), 2–8.

Maskarinec, G., Morimoto, Y., Novotny, R., Nordt, F. J., Stanczyk, F. Z., & Franke, A. A. (2005). Urinary sex steroid excretion levels during a soy intervention among young girls: A pilot study. *Nutrition and Cancer*, *52*, 22–28.

Morris, N. M., & Udry, R. J. (1980). Validation of a self-administered instrument to assess stage of adolescent development. *Journal of Youth and Adolescence*, *9*, 271–280.

Nelson, E. E., Leibenluft, E., McClure, E. B., & Pine, D. S. (2005). The social re-orientation of adolescence: A neuroscience perspective on the process and its relation to psychopathology. *Psychological Medicine*, *35*, 163–174.

Norris, S. A., & Richter, L. M. (2005). Usefulness and reliability of Tanner pubertal self-rating to urban Black adolescents in South Africa. *Journal of Research on Adolescence*, *15*, 609–624.

Palmert, M. R., Hayden, D. L., Mansfield, M. J., Crigler, J. F., Jr., Crowley, W. F., Jr., Chandler, D. W., et al. (2001). The longitudinal study of adrenal maturation during gonadal suppression: Evidence that adrenarche is a gradual process. *Journal of Clinical Endocrinology and Metabolism*, *86*, 4536–4542.

Petersen, A., Crockett, L., Richards, M., & Boxer, A. (1988). A self-report measure of pubertal status: Reliability, validity, and initial norms. *Journal of Youth and Adolescence*, *17*, 117–133.

Rosenfield, R. L., Devine, N., Hunold, J. J., Mauras, N., Moshang, T., Jr., & Root, A. W. (2005). Salutary effects of combining early very low-dose systemic estradiol with growth hormone therapy in girls with Turner syndrome. *Journal of Clinical Endocrinology and Metabolism*, *90*, 6424–6430.

Rubinow, D. R., & Schmidt, P. J. (1996). Androgens, brain, and behavior. *American Journal of Psychiatry*, *153*, 974–984.

Schmitz, K. E., Hovell, M. F., Nichols, J. F., Irvin, V. L., Keating, K., Simon, G. M., et al. (2004). A validation study of early adolescents' pubertal self-assessments. *Journal of Early Adolescence*, *24*, 357–384.

Shirtcliff, E. A., Granger, D. A., Schwartz, E., & Curran, M. J. (2001). Use of salivary biomarkers in biobehavioral research: Cotton-based sample collection methods can interfere with salivary immunoassay results. *Psychoneuroendocrinology*, *26*, 165–173.

Shirtcliff, E. A., Granger, D. A., Schwartz, E. B., Curran, M. J., Booth, A., & Overman, W. H. (2000). Assessing estradiol in biobehavioral studies using saliva and blood spots: Simple radioimmunoassay protocols, reliability, and comparative validity. *Hormones and Behavior*, *38*, 137–147.

Shirtcliff, E., Zahn-Waxler, C., Klimes-Dougan, B., & Slattery, M. (2007). Salivary dehydroepiandrosterone responsiveness to social challenge in adolescents with internalizing problems. *Journal of Child Psychology and Psychiatry*, *48*, 580–591.

Sisk, C. L., & Foster, D. L. (2004). The neural basis of puberty and adolescence. *Nature Neuroscience*, *7*(10), 1040–1047.

Steinberg, L. (2000). Gallagher lecture. The family at adolescence: Transition and transformation. *Journal of Adolescent Health*, *27*, 170–178.

Steinberg, L. (2005). Cognitive and affective development in adolescence. *Trends in Cognitive Sciences*, *9*, 69–74.

Steinberg, L., Dahl, R., Keating, D., Kupfer, D., Masten, A. S., & Pine, D. (2006). The study of developmental psychopathology in adolescence: Integrating affective neuroscience with the study of context. In D. Cicchetti & D. Cohen (Eds.), *Handbook of developmental psychopathology* (2nd ed., pp. 710–741). New York: Wiley.

Tanner, J. M. (1962). *Growth at adolescence*. Springfield, IL: Thomas.