

Susan C. Campisi, Khadija N. Humayun, Yaqub Wasan, Sajid Soofi, Muhammad Islam, Wendy Lou, Ashley Vandermorris, Olle Söder and Zulfiqar A. Bhutta*

Self-assessed puberty is reliable in a low-income setting in rural Pakistan

<https://doi.org/10.1515/jpem-2020-0246>

Received May 6, 2020; accepted June 29, 2020; published online August 20, 2020

Abstract

Objectives: Staging sexual maturation is an integral component of adolescent research. The Pubertal Development Scale (PDS) is commonly used as a puberty self-assessment tool because it avoids the use of images. Among the youth living in rural Pakistan, we determined the accuracy of self-reported pubertal assessments using a

modified PDS compared to the 'gold standard' of physically assessed Tanner stages by a physician.

Methods: The strength of agreement between self-assessed puberty using a modified PDS and the 'gold' standard of physician-assessed Tanner stages was reported using weighted kappa (κ_w) for girls ($n = 723$) of 9.0–14.9 years of age or boys ($n = 662$) of 10.0–15.9 years of age living in the rural District of Matiari.

Results: Agreement between the gold standard and self-assessment for puberty was substantial, with a κ_w of 0.73 (95% confidence interval [CI]: 0.67; 0.79) for girls and a κ_w of 0.61 (95% CI: 0.55; 0.66) for boys. Substantial agreement was observed for both boys and girls classified as thinness but only for girls with a normal body mass index. Those who were classified as severely thin had moderate agreement. The prevalence of overestimation was 18.5% (95% CI: 15.9–21.5) for girls and 2.7% (95% CI: 1.7–4.3) for boys, while the prevalence of underestimation estimation was 8.0% (95% CI: 6.2–10.2) for girls and 29.0% (95% CI: 25.8–32.6) for boys.

Conclusions: Most girls and boys assessed their pubertal development with substantial agreement with physician assessment. Girls were better able to assess their puberty, but they were more likely to overestimate. Agreement for boys was also substantial, but they were more likely to underestimate their pubertal development. In this rural Pakistan population, the PDS seems to be a promising tool for self-assessed puberty.

Keywords: puberty development scale; self-assessment; sexual maturity rating.

The transition period between childhood and adulthood encompasses adolescence and the biological process of sexual maturation called puberty. The timing of puberty is widely used to assess the impact of exposures like nutrition and environment on the health of a youth [1]. Although pubertal assessment can be highly sensitive, generally, it is accepted to be an important factor for the correct diagnosis of certain health conditions during adolescence. Additionally, the disruption of typical pubertal timing is associated with a higher risk of many disease outcomes later in life including cardiovascular and metabolic disorders, as well as compromised adult height and bone density in both men and women [2–4]. These adverse outcomes

*Corresponding author: Zulfiqar A. Bhutta, Centre for Global Child Health, Hospital for Sick Children, Peter Gilgan Centre for Research and Learning (PGCRL), 686 Bay Street, 11th Floor, Suite 11.9805, Toronto, ON, M5G 0A4, Canada; Faculty of Medicine, Department of Nutritional Sciences, University of Toronto, Medical Sciences Building, 5th Floor, Room 5253A, 1 King's Circle College, Toronto, ON, M5S 1A8, Canada; Centre of Excellence in Women and Child Health, Aga Khan University Stadium Road, PO Box 3500, Karachi, 74800, Pakistan; and Dalla Lana School of Public Health University of Toronto Health Sciences, Building 155 College Street, 6th floor Toronto, ON, M5T 3M7, Canada, Phone: 416.813.7654, ext. 301774, E-mail: zulfiqar.bhutta@aku.edu

Susan C. Campisi, Centre for Global Child Health, Hospital for Sick Children, Peter Gilgan Centre for Research and Learning (PGCRL), Toronto, ON, Canada; Faculty of Medicine, Department of Nutritional Sciences, University of Toronto, Toronto, ON, Canada, E-mail: susan.campisi@sickkids.ca. <https://orcid.org/0000-0003-1072-9228>

Khadija N. Humayun, Yaqub Wasan and Sajid Soofi, Centre of Excellence in Women and Child Health, Aga Khan University, Karachi, Pakistan, E-mail: khadija.humayun@aku.edu (K.N. Humayun), yaqub.wasan@aku.edu (Y. Wasan), sajid.sooofi@aku.edu (S. Soofi)

Muhammad Islam, Centre for Global Child Health, Hospital for Sick Children, Peter Gilgan Centre for Research and Learning (PGCRL), Toronto, ON, Canada, E-mail: muhammad.islam@sickkids.ca

Wendy Lou, Dalla Lana School of Public Health University of Toronto Health Sciences Building, Toronto, ON, Canada, E-mail: wendy.lou@utoronto.ca

Ashley Vandermorris, Centre for Global Child Health, Hospital for Sick Children, Peter Gilgan Centre for Research and Learning (PGCRL), Toronto, ON, Canada; Division of Adolescent Medicine, Hospital for Sick Children, Toronto, ON, Canada, E-mail: ashley.vandermorris@sickkids.ca

Olle Söder, Pediatric Endocrinology Division, Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden, E-mail: Olle.Soder@ki.se

underscore the need to determine variations in pubertal milestone timing through accurate puberty measurements.

Pubertal timing refers to the age of attainment of secondary sexual characteristics. Sexual maturity rating, or Tanner stages, classify pubertal development on a five-point graded scale for breast and pubic hair development for girls and genital and pubic hair development for boys [5, 6]. The 'gold standard' for determining Tanner stages is the physical examination by a health-care professional. However, physical assessments are resource dependent and may be impractical for large cohort studies. Furthermore, owing to the sensitive nature of the physical assessment, children and parents may be less likely to consent to this method of examination over self-assessment. Alternatively, self-assessment tools have been used in the assessment of puberty development. Standardized tools for self-assessment do not exist, but the most common tool includes a series of photographs or line drawings representing the five Tanner stages. When it is not appropriate to show images to participants, the Pubertal Developmental Scale (PDS) [7] has been used extensively—a questionnaire-based measure of pubertal events that avoids the sensitive aspects of a physical exam or visually explicit sexual maturation diagrams.

This study aims to determine the accuracy of self-reported pubertal assessments using a modified PDS compared to the 'gold standard' of physically assessed Tanner stages by a physician among the youth living in rural Pakistan.

Methods

A cross-sectional study among girls (n = 723) of 9.0–14.9 years of age and boys (n = 662) of 10.0–15.9 years of age living in the rural District of Matiari, Pakistan, was undertaken between January 2019 and February 2020. Matiari has a population of about 770,000, which is representative of rural conditions in Pakistan [8]. Anthropometric measures, sexual maturation, and other participant characteristics were assessed. Details regarding the Nash-wo-Numa (growth and development) study protocol are published in the study by Campisi et al. [9]. Ethics approval has been granted by the Ethics Review Committee at the Aga Khan University, Karachi, Pakistan, 5251-WCH-ERC-18, and Research Ethics Board at SickKids Hospital, Toronto, Canada: 1000060684 (trial registration number: NCT03647553).

The puberty assessment module was composed of two components: (a) self-assessed puberty staging based on a modified PDS and (b) physically assessed Tanner staging by same-sex study physicians [5–7]. Owing to the sensitive nature of puberty assessment, additional parental consent and participant assent were obtained specifically for this module of the Nash-wo-Numa study.

Self-assessment of pubertal milestones

The PDS is a self-reported questionnaire without the use of images for children and adolescents to report the events related to the

development of secondary sexual characteristics [7, 10]. It encompasses a set of questions about acne, growth and axillary hair. Additionally, facial hair and voice changes are included for males and breast development and menstruation for females. The PDS is not intended to provide a one-to-one translation to the five Tanner Stages [11] as it measures pubertal events that are not considered by Tanner ratings like voice break in males and menarche. In the current study, modification to the PDS included removal of the questions about the pubertal growth spurt and acne. Supporting these modifications is a systematic review that observed a weak association of acne with timing of puberty especially among different ethnic groups [12] and two studies that reported poor correlations between growth spurt and PDS scores [13, 14]. All responses were dichotomous (yes/no) and categorized onto a three-point scale outlined by the UK Royal College of Physicians [15].

Self-assessment was obtained through an interview by the same-sex physician as literacy is typically low as indicated by low levels of education completion in this study. The interviewer-administered PDS has been implemented previously in similar research settings [14, 16]. Additionally, before administering the PDS, physicians explained puberty and the related body changes using a standard script.

Physician assessment of pubertal milestones

Physical assessments using the Tanner criteria have been deemed the 'gold standard' for assessment of sexual development [17]. Considering local cultural norms, Tanner stage assessment was carried out by a same-sex study physician while each participant was lightly clothed in a private clinical examination room. This methodology was adopted from a previous study on school-age children in an urban slum setting in Karachi, Pakistan [18]. Originally limited to a visual assessment for breasts, the current consensus indicates the addition of palpation to distinguish breast tissue from adipose tissue, in order to avoid the overestimations of breast staging in overweight girls [19]. Accordingly, in such cases, the current study employed palpation in breast assessment.

Tanner stage assessment training for study physicians was conducted prior to the start of the study and at regular intervals for the duration of the study. It included hands-on training conducted by a pediatric endocrinologist (K.N.H.) and an online module with an evaluation component designed at the Hospital for Sick Children proven to improve accuracy, confidence, and comfort in pubertal examinations [20, 21]. Owing to the limited capacity of this study, the physical Tanner stage assessment was carried out by the same physician who administered the PDS. One male conducted all male assessments, and one female physician conducted all female assessments.

Statistical analysis

Responses to the PDS were converted into three phases of puberty based on the criteria outlined by the Royal College of Paediatrics and Child Health in the UK [15]. Physician-assessed Tanner stages were converted to the three puberty phases as follows: *prepuberty*, Tanner stage 1; *in puberty*, Tanner stage 2 + 3; and *completing puberty*, Tanner stage 4 + 5. Body mass index (BMI) and height were converted to BMI z-scores (BMIZ) and height-for-age z-scores (HAZ) using World Health Organization Growth References for Children and Adolescents [22].

Standard cutoffs for overweight/obese ($\text{BMIZ} > 2$), normal weight ($-2 < \text{BMIZ} \leq 2$), thinness ($-3 < \text{BMIZ} \leq -2$), severe thinness ($\text{BMIZ} \leq -3$), normal height ($\text{HAZ} > -1$), mildly stunted ($-2 < \text{HAZ} \leq -1$), and stunted ($\text{HAZ} \leq -2$) were used to determine individual participants' status and the percentage of youth meeting cutoff values [23].

To determine the strength of agreement between self-assessed puberty and physician-assessed puberty, we report weighted kappa (κ_w) along with percent agreement as per the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) by Kottner et al. [24]. The linear weighting scheme by Warrens [25] was applied in the calculation of κ_w . Agreement analyses were conducted for all participants and subgroups as per BMI categories and HAZ categories. As there were very few (boys, $n = 8$; girls, $n = 3$) obese or overweight participants, the agreement analysis was not conducted on this BMI category. The strength of agreement criteria for the κ_w statistic was determined using Landis [26] benchmarks wherein $\kappa_w < 0.01$ is considered poor; $\kappa_w = 0.01\text{--}0.20$ is considered slight; $\kappa_w = 0.21\text{--}0.40$ is considered fair; $\kappa_w = 0.41\text{--}0.60$ is considered moderate; $\kappa_w = 0.61\text{--}0.80$ is considered substantial; and $\kappa_w > 0.8$ is considered almost perfect. The proportion of underestimation and the proportion of overestimation when comparing the self-assessed PDS against physician-assessed puberty for the overall sample and BMI category and stunting status are also reported. Participant characteristics (mean, 95% confidence interval [CI], or n , %) stratified by agreement level (correctly assessed, overestimated, or underestimated) for girls and boys were calculated. We determined mean values to be significantly different if the 95% confidence intervals did not overlap. All analysis was performed in Stata version 14.2.

Results

This analysis represents 1,385 children and adolescents (52.2% females). Participant characteristics are presented in Table 1. The overall κ_w for girls was substantial at 0.73 (95% CI: 0.67; 0.79) and for boys was substantial at 0.61 (95% CI: 0.55; 0.66). Bodyweight classified as normal (girls only) or thinness had a substantial agreement. Conversely, those who were classified as severely thin had moderate agreement (Figure 1). As there were too few participants classified as overweight or obese, agreement was not calculated for this bodyweight classification. Overall, 3×3 puberty phase tables as well as 3×3 tables for each BMI category and stunting status by sex are presented in Supplementary Tables 1–5.

Girls were more likely to overestimate, while boys were more likely to underestimate their pubertal development. The prevalence for overestimation was 18.5% (95% CI: 15.9–21.5) for girls and 2.7% (95% CI: 1.7–4.3) for boys. Conversely, the prevalence for underestimation estimation was 8.0% (95% CI: 6.2–10.2) for girls and 29.0% (95% CI: 25.8–32.6) for boys. The only participant characteristics that differed by the level of agreement (overestimation or underestimation) were age, weight, and education level. The mean age of girls was significantly greater at 12.53

years (95% CI: 12.35; 12.81) among those who overestimated their pubertal development than among those who correctly assessed their pubertal development at 11.96 years (95% CI: 11.8; 12.11), whereas the mean age of 13.76 years (95% CI: 13.59; 13.93) for boys who underestimate their pubertal development was significantly greater than those who correctly assessed their pubertal development, with a mean age of 12.44 years (95% CI: 12.28; 12.59) (Supplementary Table 6). For girls, the mean weight (kg) was significantly lower at 26.85 kg (95% CI: 25.71; 28.98) among those who underestimated their pubertal development than among those who correctly estimated (31.41 kg; 95% CI: 31.61; 32.22) or overestimated (32.98 kg; 95% CI: 32.14; 33.81) their pubertal development. There were no differences in HAZ, BMI, BMIZ, and mid-upper arm circumference among those who correctly assessed, overestimated, or underestimated their pubertal development for boys and girls. Participant characteristics by level of agreement (correctly assessed, overestimated, or underestimated) are found in Supplementary Table 6 and 7.

Discussion

The results of the current study indicate that the majority of girls and boys living in Matiari can assess their pubertal development with substantial agreement to clinician assessment when using a questionnaire tool without images like the PDS. Although not significant owing to the small number of participants in the severely thin category, self-assessment among such participants had the lowest agreement with physician-assessed puberty regardless of sex (Figure 1). While this may reflect the inability to detect breast tissue with low bodyfat in girls, these results cannot be similarly explained in boys. Not surprisingly, stunting did not impact the ability to self-assess puberty. Stunting did not impact the ability to self-assess puberty. Overall, female self-assessment was more accurate among Matiari participants.

Few studies have reported the agreement between youth-assessed puberty using the PDS and clinician-assessed puberty [13, 14, 27, 28]. Furthermore, comparing the current study to those in the literature is problematic owing to the use of different agreement statistics and to the use of five-point over three-point puberty development scales. Two recent studies reported agreement between youth-assessed puberty using the PDS and clinician-assessed puberty use a three-point scale for puberty similar to the current study. One study conducted in the USA reported a moderate strength of agreement ($\kappa = 0.57$; 95% CI: 0.49–0.66), but this kappa coefficient was unweighted and not stratified by sex thereby limiting the sex-wise

Table 1: Participant characteristics, by sex.

Variable	Frequency, n (%) or mean (SD)	
	Girls (n = 723)	Boys (n = 662)
Age, mean (SD)	12.0 (1.7)	12.8 (1.7)
Weight, kg	31.3 (8.6)	34.6 (10.0)
Height, cm	138.7 (11.5)	145.7 (13.4)
HAZ	-1.7 (1.1)	-1.3 (1.1)
HAZ categories		
Normal height (HAZ > -2)	454 (62.8)	483 (73.0)
Stunted (HAZ ≤ -2)	269 (37.2)	179 (27.0)
BMI, kg/m ²	16.0 (2.5)	16.0 (2.3)
BMIZ	-1.2 (1.1)	-1.4 (1.2)
BMIZ categories		
Overweight/Obese (BMIZ > 2)	3 (0.4)	8 (1.2)
Normal weight (-2 ≤ BMIZ ≤ 2)	572 (79.1)	447 (67.5)
Thinness (-3 ≤ BMIZ < -2)	113 (15.6)	144 (21.8)
Severe thinness (BMIZ < -3)	35 (4.9)	63 (9.5)
Stunted and thinness	77 (10.7)	77 (11.6)
MUAC, cm	19.3 (3.4)	19.5 (2.9)
MUAC cutoff		
≥16 cm	665 (92.0)	607 (91.7)
<16 cm	58 (8.0)	55 (8.3)
Currently attending school	374 (51.8)	495 (74.8)
Completed level of formal education		
None	291 (40.3)	137 (20.7)
Grade 1–5	368 (51.0)	377 (57.0)
Grade 6–11	63 (8.7)	148 (22.4)

SD, standard deviation; HAZ, height-for-age z-score; BMI, body mass index; BMIZ, BMI z-score; MUAC, mid-upper arm circumference.

All z-scores were calculated using the World Health Organization Growth References for Children and Adolescents [22].

comparison to the current study [27]. Conversely, a Brazilian study reported high levels of agreement but used intraclass correlation coefficients of 0.82 for boys (95% CI: 0.70–0.90) and of 0.81 for girls (95% CI: 0.70–0.89) [13]. Older PDS agreement studies map the PDS onto the five-point Tanner stages. These studies report a lower strength of agreement,

which can be related to the differences between using a five-point scale over a three-point scale. A study from China reported moderate strength of agreement ($\kappa_w = 0.57$; 95% CI: 0.44, 0.71) for girls and ($\kappa_w = 0.58$; 95% CI: 0.47, 0.69) for boys [28] despite providing the youth with Tanner stages (using images) before completing their PDS assessment. A study from Soweto, South Africa, also mapped the PDS onto the five Tanner stages and reported a fair agreement ($\kappa = 0.34$), but this kappa coefficient was unweighted and only for girls [14]. The authors suggested that possible reasons for the low agreement included were a lack of awareness of pubertal development and low literacy in Soweto. Possible reasons for the greater strength of agreement between the current study and the aforementioned studies include mapping the PDS onto a three-category puberty scale, the use of dichotomous responses, and the explanation of puberty by study physicians before the PDS.

A recent systematic review reporting the agreement between self-reported to clinician-based physical assessment of Tanner Stages observed greater agreement when more details were provided to the youth [29]. In the current study, participants were provided additional pubertal information through the mini-education physician script/dialogue which may have contributed to a better agreement. Also consistent with this recent systematic review was the finding related to lower strength agreement among boys. Irrespective of the puberty self-assessment tool, boys seem to be less reliable than girls in accurately assessing pubertal development. One possible explanation is that changes in boys like voice change and testicular development are more subtle, whereas for girls, breast development and menarche are more easy to assess.

Inaccurate self-assessments seemed to be impacted by age and by weight among girls and boys. It is possible that these observations are the result of bias observed in the

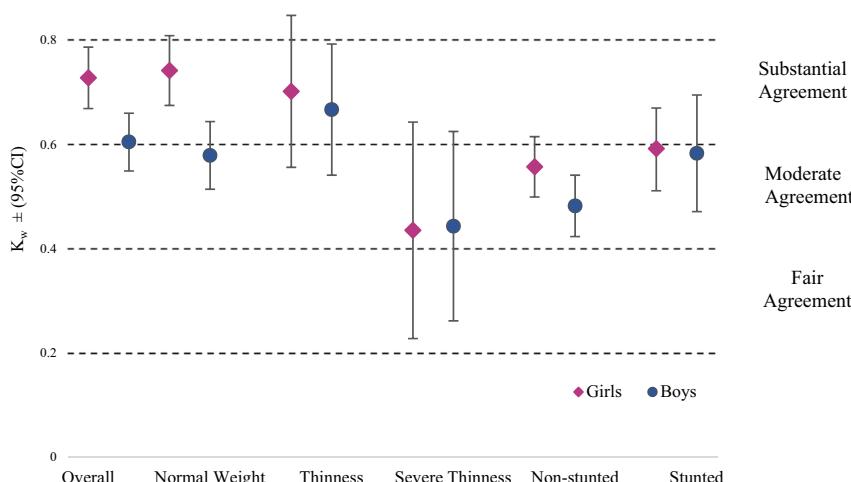


Figure 1: Strength of agreement between self-assessed puberty using the PDS and physician-assessed puberty for bodyweight categories, by sex. PDS = Pubertal Development Scale; CI = confidence interval.

reporting of socially desirable norms but need further validation in this population. The bias to report 'socially desirable' norms has been reported in adolescent literature for the misreporting of self-assessed body weight and height [30].

One of the most important strengths of the current study pertains to its large sample size from an under-represented rural low- and middle-income countries (LMIC) population. Previous studies examining the agreement between the PDS and clinician-assessed Tanner stages contained sample sizes of <300 participants. The current study also used one physician per sex, thereby avoiding inter-rater variability, thus eliminating potential bias due to the rater. Additionally, physicians underwent periodic retraining and evaluation using the online assessment module [20, 21]. It is possible that low school attendance and low literacy rates among participants in the current study could have contributed to a lack of knowledge around sexual development and negatively impacted the ability of the participants to self-assess pubertal development, although this seems to have been largely mitigated by the study design including puberty education before the PDS puberty assessments. Because the PDS does not translate directly into Tanner stages, this limitation should be considered in light of various clinical situations and research questions. A limitation of the current study is the low number of overweight and obese participants, for which accuracy of the modified PDS could not be calculated. Validation of the modified PDS in similar settings with participants whose BMI is higher is necessary.

With more than 2,000 citations, the PDS is an extremely popular tool used to assess puberty. Surprisingly, there is limited evidence in the literature assessing the agreement between youth-assessed puberty using the PDS and clinician-assessed puberty using Tanner stages mapped onto a five-point or three-point scale. In the current study, we found the PDS to be a promising tool when exact correspondence to Tanner stages is not necessary and when accompanied by a short puberty education script. In LMIC settings, administering a modified PDS could inform general clinical practice and research, where the invasive physical exam is difficult to administer owing to cultural sensitivity, lack of same-sex physicians, or prohibitive costs.

Acknowledgments: We appreciate the ongoing assistance and support provided by Agha Shakeel, Fariha Shaheen, Imran Ahmed, Arjumand Rizvi, and the entire team at the Aga Khan University. Appreciation is due to the youth of Matiari and their families for agreeing to participate in this study to advance adolescent research in rural Pakistan.

Research funding: None declared.

Author contributions: S.C.C. conceptualized and designed the study, developed the statistical analysis plan, drafted the initial manuscript, and reviewed and revised the manuscript. K.N.H. assisted with formulating culturally acceptable accurate puberty assessment procedures, conducted the physician training for puberty assessments, and reviewed and revised the manuscript. W.L. developed the statistical analysis plan and reviewed and revised the manuscript. M.I. performed the statistical analysis and reviewed and revised the manuscript. A.V. and O.S. assisted with interpretation of the results and reviewed and revised the manuscript. Z.A.B. provided senior supervision for the study, conceptualized and designed the study, assisted with interpretation of the results, and reviewed and revised the manuscript. All authors approved the final manuscript to be submitted.

Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

Ethical approval: Ethics approval has been granted by the Ethics Review Committee at the Aga Khan University, Karachi, Pakistan, 5251-WCH-ERC-18, and Research Ethics Board at SickKids Hospital, Toronto, Canada: 1000060684 (trial registration number NCT03647553).

Informed consent: Informed consent was obtained from all individuals included in this study.

References

1. Palmert MR, Dunkel L. Clinical practice. Delayed puberty. *N Engl J Med* 2012;366:443–53.
2. Day FR, Elks CE, Murray A, Ong KK, Perry JR. Puberty timing associated with diabetes, cardiovascular disease and also diverse health outcomes in men and women: the UK Biobank study. *Sci Rep* 2015;5:11208.
3. Limony Y, Koziel S, Friger M. Age of onset of a normally timed pubertal growth spurt affects the final height of children. *Pediatr Res* 2015;78:351–5.
4. Zhu J, Chan Y-M. Adult consequences of self-limited delayed puberty. *Pediatrics* 2017;139:e20163177. <https://doi.org/10.1542/peds.2016-3177>.
5. Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. *Arch Dis Child* 1969;44:291–303.
6. Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. *Arch Dis Child* 1970;45:13–23.
7. Petersen AC, Crockett L, Richards M, Boxer A. A self-report measure of pubertal status: reliability, validity, and initial norms. *J Youth Adolesc* 1988;17:117–33.
8. Pakistan Bureau of Statistics. 2017 Population census Islamabad. Islamabad: Pakistan Bureau of Statistics, Ministry of Statistics, Islamabad, Islamic Republic of Pakistan; 2018. Available from: <http://www.pbscensus.gov.pk/>.

9. Campisi SC, Wasan Y, Soofi S, Monga S, Korczak DJ, Lou W, et al. Nash-wo-Numa (childhood growth & development) study protocol: factors that impact linear growth in children 9 to 15 years of age in Matiari, Pakistan. *BMJ open* 2019;9:e028343.
10. Carskadon MA, Acebo C. A self-administered rating scale for pubertal development. *J Adolesc Health* 1993;14:190–5.
11. Shirtcliff EA, Dahl RE, Pollak SD. Pubertal development: correspondence between hormonal and physical development. *Child Dev* 2009;80:327–37.
12. Tan JKL, Bhate K. A global perspective on the epidemiology of acne. *Br J Dermatol* 2015;172:3–12.
13. Pompeia S, Zanini GAV, Freitas RS, Inacio LMC, Silva FCD, Souza GR, et al. Adapted version of the pubertal development scale for use in Brazil. *Rev Saude Publica* 2019;53:56.
14. Norris SA, Richter LM. Are there short cuts to pubertal assessments? Self-reported and assessed group differences in pubertal development in African adolescents. *J Adolesc Health* 2008;42:259–65.
15. Royal College of Paediatrics and Child Health. Fact sheet:UK 2–18 years growth chart. In: RCPCH. London, England: Royal College of Paediatrics and Child Health; 2012.
16. Schmitz KE, Hovell MF, Nichols JF, Irvin VL, Keating K, Simon GM, et al. A validation study of early adolescents' pubertal self-assessments. *J Early Adolesc* 2016;24:357–84.
17. Herman-Giddens ME, Bourdon CJ, Dowshen SA, Reiter EO. Assessment of sexual maturity stages in girls and boys. 2011. <http://site.ebrary.com/id/10477942>.
18. Campisi SC, Humayun KN, Rizvi A, Lou W, Söder O, Bhutta ZA. Later puberty onset among chronically undernourished adolescents living in a Karachi slum, Pakistan. *Acta Paediatr* 2020;109:1019–25.
19. Bonat S, Pathomvanich A, Keil MF, Field AE, Yanovski JA. Self-assessment of pubertal stage in overweight children. *Pediatrics* 2002;110:743–7.
20. Ens A, Janzen K, Palmert MR. Development of an online learning module to improve pediatric residents' confidence and knowledge of the pubertal examination. *J Adolesc Health* 2017;60:292–8.
21. Puberty assessment learning module Toronto. Available from: <http://www.sickkids.ca/Endocrinology/Resources/index.html>.
22. de Onis M, Onyango AW, Borghi E, Siyam A, Nishida C, Siekmann J. Development of a WHO growth reference for school-aged children and adolescents. *Bull World Health Organ* 2007;85:660–7.
23. WHO Multicentre Growth Reference Study Grou. WHO Child Growth Standards based on length/height, weight and age. *Acta Paediatr Suppl* 2006;450:76–85.
24. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *Int J Nurs Stud* 2011;48:661–71.
25. Warrens MJ. Weighted kappas for 3 × 3 tables. *J Probab Stat* 2013;2013:1–9.
26. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
27. Koopman-Verhoeff ME, Gredvig-Ardito C, Barker DH, Saletin JM, Carskadon MA. Classifying pubertal development using child and parent report: comparing the pubertal development scales to Tanner staging. *J Adolesc Health* 2020. <https://doi.org/10.1016/j.jadohealth.2019.11.308>.
28. Chan NP, Sung RY, Nelson EA, So HK, Tse YK, Kong AP. Measurement of pubertal status with a Chinese self-report pubertal development scale. *Matern Child Health J* 2010;14:466–73.
29. Campisi SC, Marchand J, Siddiqui FJ, Islam M, Bhutta ZA, Palmert MR. Can we rely on adolescents to self-assess puberty stage? A systematic review and meta-analysis. *J Clin Endocrinol Metab* 2020. <https://doi.org/10.1210/clinem/dgaa135>.
30. Gil J, Mora T. The determinants of misreporting weight and height: the role of social norms. *Econ Hum Biol* 2011;9:78–91.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/jpem-2020-0246>).