Adults' perceptions of children's age: A developmental approach

Jessie M. Bridgewater [*], Tuppett M. Yates

University of California, Department of Psychology, 900 University Ave., Riverside, CA 92521, United States of America

ARTICLE INFO

ABSTRACT

This study examined adults' perceptions of children's age (APCA) using images drawn from a longitudinal investigation of children ($N = 245$; 49 % assigned female at birth, 88.6 % non-white) who were photographed at 4, 6, 8, 10 and 12 years of age and then rated by 400 college students (65.7 % assigned female at birth; 94.7 % non-white). Multilevel modeling demonstrated that APCA increased with children's chronological age. However, raters overestimated boys' ages more than girls' ages in early development (i.e., ~ ages 4–5.5 years), whereas they overestimated girls' ages more than boys' ages in later development (i.e., ~ ages 5.5–12 years). Raters also evaluated children they perceived as having darker skin tone as younger than children they perceived as having lighter skin tone until ~ age 6, after which they evaluated children they perceived as having darker skin tone as older than children with lighter skin tone. Regarding children's ethnicity/race, raters perceived Black children as younger than non-Black children until ~ age 6, after which they perceived Black children as older than non-Black children. Rater assigned sex and ethnicity/race also predicted age ratings. For example, female raters over-estimated children's ages more than male raters from ~ ages 4–9 years, after which male raters overestimated children's ages more than female raters. The results suggest that child assigned sex, child ethnicity/race, child skin tone, rater assigned sex, and rater ethnicity/race are associated with APCA across time. Patterns of APCA and developmental shifts therein have implications for applied contexts and future research.

In 2012, 15-year-old Alexis Sumpter was arrested in Harlem for using her student subway card. The arresting officers believed she was too old to use a student card because it was only valid for those under the age of 19. Alexis was kept in custody until police were shown her birth certificate. Upon leaving the police station, Alexis had to get her wrists treated at a hospital due to being kept in handcuffs for a prolonged period of time.

(Parascandola, 2012)

All too often, Black and Brown children are misperceived as being much older than they are, frequently with dire consequences (e.g., the murder of Tamir Rice). However, despite research on adults' perceptions of other adults' age (e.g., Nielsen et al., 2015; Voegeli et al., 2021), few studies have examined adults' perceptions of children's age (APCA; see Epstein et al., 2017; Cooke & Halberstadt, 2021, Goff et al., 2014 for exceptions). Moreover, the limited research on APCA to date has been conducted outside developmental psychology, despite the need to understand if and how the form and function of APCA may change over time with varying implications for children's adaptation.

The current study filled important gaps in our understanding of APCA and child development. First, we drew on a longitudinal study of development to document college students' ratings of images from an ethnically and racially diverse sample of 245 children taken at 4, 6, 8, 10 and 12 years of age. Second, we evaluated the associations of child assigned sex at birth, child skin tone, child ethnic/racial status, rater assigned sex at birth, rater ethnic/racial status, rater child exposure, and rater child contact, as well as their interactions, on patterns of growth in APCA over time (i.e., slope) and on APCA levels in early adolescence at age 12 (i.e., intercept). Finally, given our expectation that children with more advanced pubertal status would receive higher APCA ratings (Carter & Seaton, 2025), we evaluated child pubertal status as an additional predictor of the APCA intercept at age 12. The current study extends prior work by employing a developmental approach to examine how APCA changes from the preschool period to early adolescence in a large sample of college student raters evaluating an ethnically/racially diverse sample of children over time.

Ecological approaches for understanding adults' perceptions of children's age (APCA)

The ecological approach to social perception (McArthur & Baron,

1983) draws on prominent theories (e.g., Gibson's theory of visual perception; Gibson, 2014) to provide a conceptual framework for research on age (and other) perceptions (e.g., Montepare & Zebrowitz, 1998). Three tenets of this approach are especially relevant to the current study. First, social perceptions, including age perceptions, serve social and biological adaptive functions. For example, age perceptions are relevant to social goals (e.g., driving a car) and species survival (e.g., likelihood of reproductive selection). Second, social perceptions are informed by people's movements, voices, faces, and bodies. Third, physical qualities indirectly communicate "affordances" to others regarding how they can interact with the perceived individual. Together, these principles explain how social perception informs social action, while recognizing the potential for perceptual inaccuracies. For example, "overgeneralization effects" (Zebrowitz, 1997) may bias perceptions as when a "baby faced" person is presumed to be younger in age, naïve in conation, and innocent in motivation (Berry & McArthur, 1985; Poutvaara et al., 2009). Thus, systematic inaccuracies in social perception across groups may drive biased social actions.

Although the ecological approach to social perception provides a useful framework for understanding the origins and implications of APCA, it does not consider whether age perceptions shift over development, nor whether the implications of age perceptions (and inaccuracies therein) may change over time or context. As the first comprehensive theory of contextual influences in developmental psychology, Bronfenbrenner's ecological systems theory (EST; Bronfenbrenner, 1979) is well-suited to consider APCA through the lens of development. Recent efforts to identify and understand development among ethnic/racial minority youth provide additional theoretical grounding for developmental research on APCA. For example, García Coll's integrative model for the study of developmental competencies in minority children emphasizes how "social position factors" influence development (García Coll et al., 1996). Social position factors encompass individual characteristics used by societies to hierarchically situate individuals based on assigned sex, ethnicity/race, social class, and other factors in ways that shape developmental outcomes for minority youth via structural racism and the discriminatory practices it supports. Importantly, the integrative model also incorporates "child characteristics," including chronological age and physical features (e.g., skin tone, height, weight), as important influences on children's developmental competencies (e.g., educational, socioemotional functioning). Together, ecological theories of social perception and child development support the value of studying APCA as a likely influence on adults' actions towards children.

Age perceptions

Although extant research with child samples demonstrates the salience of both self-perceptions (Nishina et al., 2018) and others' perceptions (Goff et al., 2014) for understanding child development. Scholars suggest that potential bias in APCA may underlie known disproportionalities in surveillance and arrest rates across groups as defined by children's assigned sex and/or ethnicity/race (e.g., Hall et al., 2016). However, only a handful of studies have systematically examined APCA within or across groups.

In a groundbreaking investigation, Goff et al. (2014) asked 59 college students (median age of 19, 58 % female; 89.8 % white) and 60 police officers (median age of 38; 7 % female; 73.3 % white) to make various evaluations of children in a criminal justice context. Raters were asked to provide their assessments of age and culpability as applied to images of Black, Latino, and white boys ages 10–17. They were shown 24 pictures (eight per group) along with a description of a crime the child in the picture had reportedly committed (i.e., misdemeanor or felony). Across eight images and crime descriptions, raters were asked to guess the age of the boy in the picture. College students rated Black felony child suspects as older than white felony child suspects, while police officers rated Black and Latino child suspects, but not white ones, as

older than their chronological age. This study provided valuable inspiration for both non-empirical (Davis & Marsh, 2020; Dumas & Nelson, 2016) and empirical (Cooke & Halberstadt, 2021; Epstein et al., 2017) papers on APCA. For example, in an online study of 325 adults (62 % female; 74 % white), Epstein et al. (2017) asked participants to complete one of two surveys about imagined Black or white girls regarding facets of adultification (including how much they seemed "older than their age") in each of the following age groups: 0–4, 5–9, 10–14, and 15–19. They found that Black girls were more likely to be perceived as older than their stated age beginning at age 5. In contrast, in a study of 152 parents' age perceptions (average age 33.47 years old; 52 % female; 67 % white) across 20 images of Black children and 20 images of white children ages 10–13, Cooke and Halberstadt (2021) found no evidence that Black children were perceived as older than white children. Most recently, as part of a pilot study about visual maturity (Koch et al., 2023), 44 college students (average age 20 years old; 77 % female; no ethnic/racial information provided) were asked to indicate the age of four randomly selected full-body images of 12-year-olds varying by sex (i.e., female versus male), race (i.e., Black versus white), and maturation (low versus high). On average, high maturation Black females and males, as well as white females were perceived as significantly older than their low maturation counterparts, but there was not a significant difference in APCA between high and low maturation white males.

Prior studies provide important descriptive data regarding APCA across different groups of children based on assigned sex, ethnicity/race, pubertal status, and purported child behaviors. However, significant methodological limitations continue to constrain the reach of extant APCA research. For example, the absence of longitudinal data using the same child stimuli across all age groups limits our understanding of whether APCA changes over time and/or in different ways for different groups of children. Moreover, Epstein and colleagues' (2017) use of imagined girls as opposed to actual images of girls to assess APCA raises concern given the documented influence of appearance features, such as eye size and skin tone, on age perceptions (Zebrowitz, 1997). Finally, although it is not known how many stimuli are needed to obtain reliable perception estimates, using more than eight images for each ethnic/racial group (and just four for each behavior type within group) would have strengthened prior conclusions (e.g., Goff et al., 2014; Koch et al., 2023). The current study addressed these limitations by exploring how both child and rater characteristics related to the degree of accuracy between adults' perceptions and children's chronological age over time.

APCA and child characteristics

Descriptive studies suggest that child characteristics may be associated with APCA (e.g., assigned sex, Epstein et al., 2017; ethnicity/race, Goff et al., 2014), yet the current study is unique in its systematic investigation not only of the main effect of child characteristics on APCA, but also of interactive effects among child characteristics within and across developmental time. For example, this study tested how child characteristics, such as assigned sex, ethnicity/race, and skin tone, related to APCA, as well as the degree to which these associations varied for younger versus older children. This kind of developmental analysis is important for identifying potential shifts in the salience of these characteristics for understanding APCA across time.

Prior studies have examined APCA among boys (Black, Latino, and white; Goff et al., 2014), girls (Black and white; Epstein et al., 2017), and in samples of Black and white girls and boys (Cooke & Halberstadt, 2021). However, as shown by Koch et al. (2023), it is also important to consider children's pubertal status when examining associations between assigned sex and APCA. For example, in addition to differences in APCA by both assigned sex and maturational status, Koch et al. (2023) found that participants were more likely to call the police on youth who appeared physically older and on boys compared to less visibly mature youth and girls. Beyond the criminal context, studies documenting the sexualization of girls by the media (Lamb & Koven, 2019) and various

people in their lives (e.g., family and friends; Zurbriggen et al., 2007) suggest that girls may be perceived as older than boys of the same chronological age. Indeed, because girls typically experience puberty earlier than boys (Brix et al., 2019), adults may be more likely to engage in earlier and greater age overestimation when rating female as compared to male children. Thus, this study not only evaluated APCA among girls and boys across developmental time (i.e., ages 4–12), but also evaluated pubertal status effects on predicted sex differences in early adolescence (age 12) when girls were expected to be rated as older than boys of the same chronological age.

The emphasis on child ethnic/racial status in studies of APCA is unsurprising given patterns of racialized perceptions and actions in adult samples. For example, racialized anger bias, which refers to the phenomenon of misperceiving anger in Black people substantially more than white people (Halberstadt et al., 2018), has been documented in studies of adult-rated images of other adults (Halberstadt et al., 2018) and of children (Halberstadt et al., 2022). Extant research on APCA has focused on comparisons between stimuli depicting Black and white children (Cooke & Halberstadt, 2021; Epstein et al., 2017), with the exception of Goff et al. (2014) who also included Latino boys. This pattern mirrors broader research on ethnic/racial disproportionalities, which have accorded relatively less attention to Latine groups and even less to youth who identify with multiple ethnic/racial identities (Skiba et al., 2011). Addressing these concerns, this investigation evaluated APCA using images of the same Latine, multi-ethnic/racial, Black, and white children taken at ages 4, 6, 8, 10 and 12. Moreover, amidst mounting concern regarding overreliance on ethnic/racial status given marked phenotypic (and other) heterogeneities within groups (Saperstein et al., 2016), we evaluated child skin tone effects, in addition to those associated with child ethnic/racial group status, to evaluate if and how child skin tone may be associated with apparent group differences (Roth, 2016).

Although there are no published investigations of child skin tone and APCA, research documenting associations between skin tone and differential treatment in varied structural systems supports its salience. For example, in the educational system, youth with darker skin tone are more likely to be suspended (Hannon et al., 2013) and receive lower grades compared to their peers with lighter skin tone even when assigned sex, age, immigrant status, parental education, and family income are held constant (Thompson & McDonald, 2016). Of course, consistent with known intersectionality effects (Crenshaw et al., 2015; Galliher et al., 2017), a child's skin tone must be considered alongside other individual characteristics, such as assigned sex. For example, the aforementioned findings of Hannon et al. (2013) regarding skin tone and school suspension were driven by the experiences of Black adolescent females. Thus, this study examined individual and interactive associations of child assigned sex, ethnic/race, and skin tone with APCA across developmental time.

APCA and rater characteristics

Although studies have not yet focused on how rater characteristics may be related to APCA, a wealth of facial recognition research suggests that people are better at recognizing faces that share their own gender (e.g., own-gender bias; Herlitz & Lovén, 2013), race (i.e., other-race effect; Meissner & Brigham, 2001), and age (i.e., own-age bias or other-age effect; Kuefner et al., 2008; Rhodes & Anastasi, 2012). Although these effects differ from APCA because the studies use recognition paradigms that assess if the participant recognizes the presented stimuli in a subsequent trial, whereas APCA research uses perception paradigms to assess the age a participant rates a child, prior work underscores the importance of considering rater characteristics when employing the kinds of facial stimuli used in this study.

Previous research has examined APCA among different groups of adult raters, including parents (e.g., Cooke & Halberstadt, 2021), college students (e.g., Koch et al., 2023), and police officers (e.g., Goff et al.,

2014), but no studies have examined if and how rater characteristics, such as rater assigned sex, ethnicity/race, or familiarity with children might be related to APCA. Given that most studies have used mixed sex rater samples, it is surprising that none have formally tested for rater sex differences in APCA ratings. Similarly, although most previous studies have relied on majority-white rater samples (e.g., Cooke & Halberstadt, 2021; Goff et al., 2014), even those with more diverse raters have not tested for differential patterns of APCA by rater ethnicity/race. In addition to potential main effects of rater sex and ethnicity/race on APCA ratings, it is important to consider potential interactions of rater characteristics with child chronological age and other child characteristics in relation to APCA. Building on the aforementioned facial recognition literature, we expected that raters would provide more accurate age ratings of children who shared their assigned sex and ethnicity/race.

In addition to rater assigned sex and ethnicity/race, this study considered if and how rater familiarity with children would relate to APCA. Drawing again on facial recognition research, Kuefner et al. (2008) found that college students who reported that they did not have extensive experience with infants and young children were better at recognizing adult faces compared to faces of newborns and children, whereas preschool teachers showed no significant differences in recognition accuracy across infant, child and adult faces. Even in a sample of primarily non-parenting adults, such as the college student raters in this study, familiarity with children may vary. Thus, this investigation examined if and how rater exposure to children via cohabitation with siblings, cousins, or other children under the age of 18 in the primary residence and rater contact frequency with children via different occupational exposures (e.g., nanny, classroom assistants, camp counselor) related to APCA.

The current study

This investigation employed growth curve models to understand how college student raters perceived children's age from the preschool period (age 4) through early adolescence (age 12) by examining their ratings of headshots across 245 children who were photographed at ages 4, 6, 8, 10, and 12 as part of a longitudinal study. This study addressed gaps in our understanding of APCA by achieving two primary goals. The first goal of this study was to examine unconditional growth models that documented APCA across 1000 images of these children from ages 4–12 as rated by college students. Our first hypothesis (H1) was that adult raters would perceive children's age as increasing over time (i.e., positive slope), and there would be significant variance such that some children would show steeper increases in perceived age ratings than others. Our second hypothesis (H2) was that there would be significant variation in APCA at age 12 (i.e., intercept variance). The second goal of this study was to examine a conditional growth model to evaluate differences in the slope and intercept of APCA as a function of child characteristics (i.e., assigned sex, ethnic/racial status, and skin tone), rater characteristics (i.e., assigned sex, ethnic/racial status, child exposure, and child contact), and their interactions with each other, as well as with child chronological age. Our third hypothesis (H3) was that college student raters would perceive children who were assigned female at birth as older than children who were assigned male at birth yielding steeper positive slopes over time and higher age estimates at age 12. Our fourth hypothesis (H4) was that raters would perceive Black children as older than children from all other ethnic/racial groups yielding steeper positive slopes over time and higher age estimates at age 12. Our fifth hypothesis (H5) was that raters would perceive children with darker skin tone as older than children with lighter skin tone resulting in faster perceived age gains over time and higher intercept estimates at age 12. That said, for our sixth hypothesis (H6), we expected to find a significant interaction between child assigned sex and skin tone such that females with darker skin tone would show both the steepest slope for perceived age and the highest age 12 intercept. Regarding rater

characteristics, our seventh and eighth hypotheses were that college students would provide more accurate age ratings of children who shared the rater's assigned sex (H7) and ethnicity/race (H8) as compared to child targets who did not share the raters' assigned sex and ethnicity/race. Finally, our ninth and tenth hypotheses were that college student raters who reported greater child exposure (i.e., living with more children under age 18; H9) and contact (i.e., greater frequency of contact with children under age 18, such as through work; H10) would provide more accurate age ratings over time compared to raters with less child exposure and contact.

Method

Participants

Participants were 250 caregiver-child dyads who were part of an ongoing, longitudinal study of development that began in the preschool period (M_{age_wave1} = 49.05 months, SD = 2.95). Children were diverse with regard to assigned sex (49 % female, 51 % male), ethnicity/race (45.7 % Latine, 24.9 % Multi-ethnic/racial, 18 % Black, and 11.4 % white), and economic status (32.7 % in poverty). The majority of caregivers were biological mothers (91.4 %), followed by foster/adoptive mothers (3.6 %), and grandmothers, aunts, or other kin (4.8 %). Regarding education levels, 16.4 % of caregivers did not have a high school degree, 13.9 % had a high school diploma or GED, 56.4 % had some post-secondary education (e.g., vocational training or two-year degree), and 13.3 % had a bachelor's degree or higher. At Wave 1, most caregivers were married (60.4 %) or in a committed relationship (19.6 %), and 67.3 % were engaged in full- or part-time employment. Five dyads were excluded from the original sample of 250 because the child's legal guardian did not consent for the use of the child's images at all Waves. Across the data waves at ages 4, 6, 8, 10 and 12, 227 (92.7 %) of the 245 eligible child participants completed two or more laboratory assessments.

Raters were 400 adult (i.e., 18 or older) college students recruited from introductory psychology courses at a large university in Southern California from September 2022 to February 2023. Most raters were assigned female at birth (i.e., 65.7 %) and identified as women (i.e., 63.6 %). Raters were between the ages of 18 and 35 (M = 19.643 years old, SD = 1.845 years old) and were economically diverse with 41.8 % receiving Pell Grant support for low-income students. The adult sample was ethnically/racially diverse with raters identifying as Latine (28.5 %), Asian (28.3 %), multi-ethnic/racial (23.0 %), or white (5.25 %). Regarding familiarity with children, approximately half the participants (53.1 %) reported cohabitation with at least one child under the age of 18 in their primary residence at the time of data collection, but only 2 were actively parenting. Further, 41.5 % of the raters reported very frequent child contact (e.g., working with children regularly in their job), 38.5 % reported somewhat frequent child contact (e.g., occasional interactions with children under the age of 18), and 20 % reported having no contact with minor children.

Procedures

Families were recruited via flyers advertising a "study of children's learning and development," which were distributed to community-based childcare programs in Southern California. At each Wave, dyads completed a three-hour, video-recorded laboratory assessment, which consisted of measures with the child, the caregiver, and the caregiver and child interacting. Caregivers received $25 per hour of assessment and children received an age-appropriate gift at the end of each assessment. Informed consent was obtained from the child's legal guardian at all waves and informed assent was collected from children beginning at age 8. In addition to informed consent for participation in the research study, legal guardians were given the option to provide consent for the use of their child's video recordings for future research

and educational purposes. Children's images were included as rating targets in the current study only when the legal guardian provided this additional consent (rates of decline are specified in the Data Analytic Plan section). That said, families were not advised of this specific investigation because it was not developed until well after data collection was complete.

Rater data were collected by trained research assistants in small groups in a university computer lab. To be eligible, raters had to be at least 18 years old and able to read English. Raters were excluded if they or anyone in their family had ever participated in a study on the University campus before 2022. Following informed consent, raters completed a digital assessment that included 1) a pre-rating survey measuring basic sociodemographic characteristics (e.g., assigned sex, ethnicity/race), 2) a training session to familiarize raters with the images and rating prompts for 10 randomly selected, unique child participants, 3) a formal rating session using images of 50 randomly selected, unique child participants, and 4) a post-rating survey to measure raters' exposure to children (i.e., number of children under the age of 18 residing in their primary residence), contact with children (i.e., frequency of contact with children under the age of 18), and parenting status (only 2 raters were parents). The digital assessment was created using Inquisit 6 (Millisecond software) and the full reproducible code is reported in the Supplemental Material. Given the sensitive nature of the visual stimuli, research assistants remained in the room to ensure that raters did not capture images of the stimuli on their phones. Raters received course credit for their participation and were debriefed after they completed the post-rating survey. All procedures pertaining to the initial data collection with child participants and the subsequent rating protocol with college student participants were approved by the human research review board of the participating university.

Measures

Image ratings

College students' ratings of each child image were collected using headshots taken from video-recorded assessments at each data wave. Each headshot showed the child with a neutral expression and open eyes from the shoulders up to avoid clothing- and height-related age cues. Headshots were screened three times for i) initial selection, ii) lighting, and iii) image quality to ensure that the final stimuli were as uniform as possible. Each Portable Network Graphic (PNG) was 576 pixels tall, though width varied (~ 300 to 800 pixels) to preserve the quality of the image and the child's facial proportions.

To avoid potentially biasing ratings, raters were told that they were going to be evaluating "pictures of people," rather than "children." Following the 10-image training period, raters evaluated 50 unique child images in 5 sets of 10 images each. Images were drawn at random from each assessment wave and randomized in presentation such that each participant rated 10 unique child images drawn from each of the five assessment waves and randomized over ages 4, 6, 8, 10, and 12. The algorithm ensured that no rater evaluated more than one image of a given child (i.e., if a child's image was randomly drawn at one data wave, it was excluded from that rater's rating pool at other Waves). Raters were shown each headshot on a computer with a rating scale just below the image that asked them to indicate the child's age from 1 to 20 years (with half ages), skin tone ("What is this person's skin tone?"; 1 = *very light*, 2 = *Light*, 3 = *Somewhat light*, 4 = *Medium*, 5 = *Somewhat dark*, 6 = *Dark*, 7 = *Very dark;* Keith et al., 2010), ethnic/racial status ("What is this person's ethnicity and/or race? Choose all that apply."), and image quality ("Please rate the quality of this image;" 1 = *very poor* to 7 = *excellent*). Following each batch of 10 ratings, raters received a mandatory 30-s break to minimize fatigue. Lastly, to prevent raters from spending too much time evaluating any single image, images automatically advanced after 45 s even if the ratings were not complete.

Child characteristics

Child assigned sex and ethnicity/race were reported by the participating caregiver. As described earlier, child skin tone ratings were obtained using a 7-point likert scale at each data wave. Finally, at age 12, children reported on their pubertal status using the Pubertal Development Scale (PDS; Petersen et al., 1988), which includes five items about growth in height, body hair, skin changes (e.g., pimples), breast development (girls only), voice changes (boys only), menstruation (girls only), and facial hair growth (boys only). The PDS has demonstrated strong reliability and validity in diverse samples (Stumper et al., 2020).

Rater characteristics

Rater assigned sex and ethnicity/race were assessed via self-report. College students' exposure to children was assessed using a self-reported item regarding the number of minor children living in the rater's primary residence at the time of data collection (i.e., "How many children under the age of 18 reside in your primary home [i.e., where you lived before attending this university]?"). Raters could enter whole numbers from *0* to *20* (*53.1 % living with at least one child under age 18; M* $= 0.835$ *and SD* $= 0.963$). College students' frequency of contact with minor children was assessed using one self-reported item asking, "What is your frequency of contact with children under the age of 18?" from 1 (*Not very frequent;* e.g., *I rarely, if ever, interact with children under the age of 18*) to 2 (*Somewhat frequent;* e.g., *occasional interactions with children under the age of 18*), to 3 (*Very frequent;* e.g., *I work with children regularly*).

Data analytic plan

A multilevel growth modeling approach accommodated variation in the number of raters for each image, as well as other sources of variability among children and raters (i.e., child assigned sex, skin tone, and image quality). Each college student rater provided a skin tone rating and ethnicity/race rating for each of the 50 images they rated which we used as predictors in the multilevel growth models. Child assigned sex at birth and ethnicity/race was provided by the parent in each dyad. However, child assigned sex was used as a predictor of APCA in the multilevel growth models, whereas the parent-provided ethnicity/race of the child was used to describe the demographics of the child sample and was not included as a predictor in these analyses.

Prior to analyses, all variables were evaluated for distributional assumptions to support parametric statistics (i.e., skew, kurtosis; Afifi et al., 2007). For each data wave, one MANOVA evaluated differences across ratings of perceived age, child skin tone, and image quality by child assigned sex, ethnicity/race, and their interaction, and a second MANOVA evaluated these same variables by rater assigned sex, ethnicity/race, and their interaction. Bonferroni-corrected post-hoc comparisons probed pairwise comparisons for child and rater ethnicity/race. Visual stimuli for select children were missing at ages 4 ($n = 33$, 13.5 %), 6 ($n = 47$, 19.2 %), 8 ($n = 52$, 21.2 %), 10 ($n = 49$, 20.0 %) and 12 ($n = 71$, 29.0 %) due to recording errors, visit non-completion, or caregiver non-consent for the use of the child's images at that data wave.

Multilevel growth curve models were computed in RStudio using the *lme4* package (Bates et al., 2014) with an unstructured covariance matrix and Satterthwaite degrees of freedom. The full reproducible code is available in the Supplemental Material. Study hypotheses were evaluated using a three-stage process to determine 1) the shape of growth, 2) the optimal number of random effects, and 3) significant predictors of growth parameters (i.e., slope and intercept). First, an unconditional no-growth model tested whether there were significant within- and between-person variances in APCA at age 12. The optimal shape of the growth curve was determined by considering polynomial functions assessing linear, quadratic, and cubic changes in APCA over time with the intercept set to age 12, as well as whether there were significant within- and between-person variances in these change parameters. Superior model fit was defined as a significant Likelihood Ratio Test (LRT),

as well as relatively low AIC and BIC values (Horváth & Plunkett, 2016). Second, after determining the best-fitting shape of growth, two unconditional models identified the appropriate number of random effects using LRTs to assess model fit. The *rand* function from the *lmerTest* package evaluated each random effect individually to determine if removing the random effect significantly worsened the model fit. The first model analyzed four random child effects (i.e., random slope, child assigned sex, child skin tone, and image quality) to determine which of these factors accounted for systematic variance between children's perceived age trajectories. The second model analyzed four random rater effects (i.e., random slope, child assigned sex, child skin tone, and image quality) to determine which of these same factors accounted for systematic variance between raters' ratings. Third, a final unconditional model combined significant random child effects and significant random rater effects, along with significant intercept random effects. Fourth, a conditional growth model examined the significance of three child characteristics, including assigned sex (dummy-coded; female = 1), ethnicity/race (effect coded with the smallest group *white* as the reference group), and skin tone (mean-centered ratings), as well as four rater characteristics, including assigned sex (dummy-coded; female = 1), rater ethnicity/race (effect coded with the smallest group *other* as the reference group), child exposure (mean-centered), and child contact (mean-centered). In addition to main effects, pairwise interactions among all child and rater characteristics were examined in predictions of APCA across chronological age. Finally, an exploratory within-time analysis at age 12 evaluated the extent to which the conditional model was impacted by the inclusion of puberty status.

Results

Descriptive and bivariate analyses

Each image of each child at each of the 5 data waves received at least 6 unique ratings, yielding a total of 15,702 ratings ($M_{ratings/image/wave} = 15.749–16.039$; $SD = 3.817–4.245$). The number of raters per image at each wave ranged from 8 to 28 at age 4 ($M = 16.005$, $SD = 3.942$), 6–27 at age 6 ($M = 16.039$, $SD = 3.817$), 7–30 at age 8 ($M = 16.005$, $SD = 4.045$), 6–28 at age 10 ($M = 15.886$, $SD = 3.965$), and 7–28 at age 12 ($M = 15.749$, $SD = 4.245$).

Descriptive statistics and bivariate correlations are presented by ratings ($N = 15,702$) and by child ($N = 245$) for the primary study variables (i.e., chronological age, perceived age, skin tone ratings, and image quality ratings). Multilevel growth models utilized all 15,702 ratings (i.e., long data) of all available images of each child. Fig. 1 provides a visual representation of children's average chronological age values and raters' perceived age estimates across all 5 data waves showing that college student raters overestimated children's age at all waves with the greatest inaccuracies observed at age 4, followed by steady improvement up to age 12.

Descriptive statistics, bivariate correlations, and summaries for both are provided for all data waves in the Supplemental Material. Beyond a consistent pattern wherein white children received the lightest skin tone ratings and Black children received the darkest skin tone ratings at each data wave, there were no consistent differences in mean values across study variables by either child or rater characteristics across data waves. Similarly, at the bivariate level, there was only one consistent relation such that chronological age was positively associated with perceived age within each data wave.

Multilevel growth models

As shown in Table 1 and Table 2, a series of unconditional growth models identified the appropriate shape of growth for APCA across unconditional means/no growth, linear, and quadratic models. The default optimizer was changed to *bobyqa* to facilitate model convergence given the complexity of the models (e.g., various random effects). The
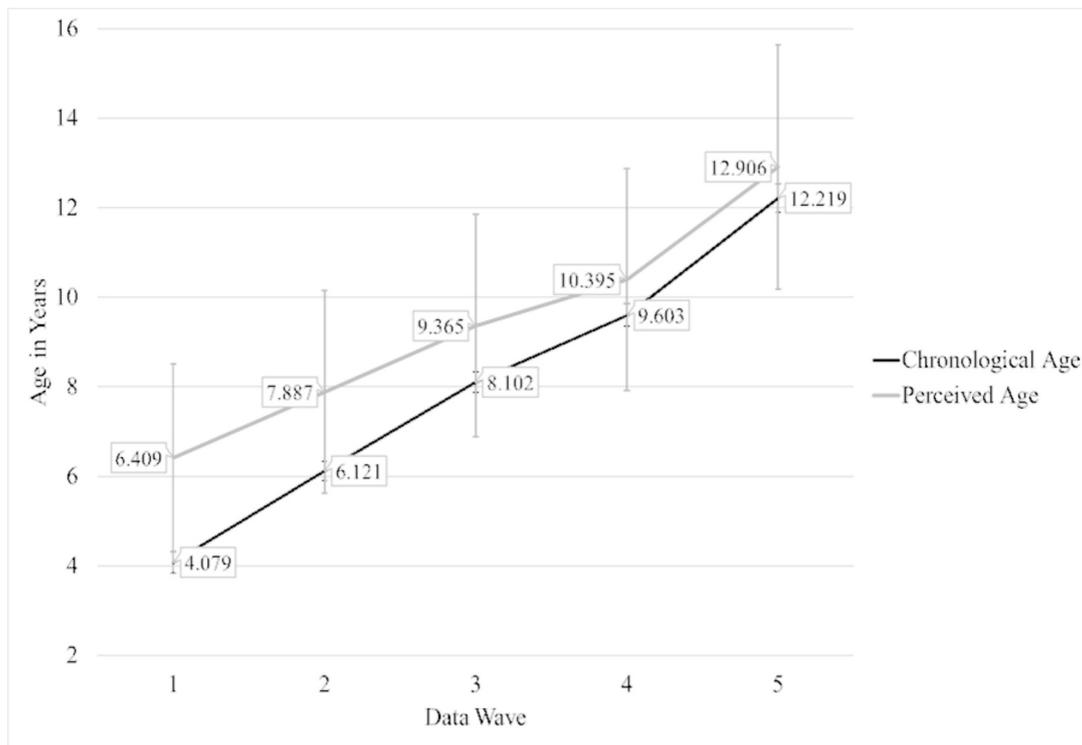
Fig. 1. Average chronological and perceived age across data waves at ages 4, 6, 8, 10, and 12 years.
Note. Vertical error bars represent the standard deviation for each average. Bar color corresponds to line color (i.e., the grey bars refer to average perceived ages).

Table 1
Unconditional growth models: fit indices.

Model	AIC	BIC	Log-likelihood	Deviance
No Growth	79,947	79,970	−39,970	79,941
Linear	69,495	69,541	−34,742	69,483
Quadratic	69,879	69,917	−34,934	69,869

Table 2
Unconditional growth models: likelihood ratio test.

Superior Model	Chi-Square	df	p
–	–	–	–
Linear	10,458	3	< 0.001
Linear	386	1	< 0.001

intraclass correlation from the unconditional means model indicated that 16.6 % of the total variation in perceived age ratings was attributable to differences between the child images.

LRTs using the *anova* function from the *lavaan* package indicated that the linear model was the best fitting unconditional model (see Table 3 and Table 4). The pseudo R^2 for the linear model indicated that 49.7 % of the within-person variance in APCA was associated with linear time. With the intercept centered at age twelve, the fixed effects estimates indicated that the average predicted perceived age rating for the age twelve images was 12.481 years old, and APCA increased by 0.776 years, annually. The random effects estimates indicated that perceived age ratings at age 12 varied by 1.364 years and the rate of change varied

Table 3
Unconditional linear growth model parameters: fixed effects.

Name	Estimate	SE	df	t	p
Intercept	12.481	0.098	220	127	<. 001
Slope	0.776	0.014	212	54	<. 001

Table 4
Unconditional linear growth model parameters: random effects.

Group	Name	Variance	SD
Images	Intercept	1.862	1.364
	Slope	0.035	0.187
Residual	–	4.601	2.145

by 0.187 years across the different images.

Two unconditional models identified the appropriate number of random effects using LRTs to assess model fit (see Table 5). The first model analyzing four random child effects indicated that child assigned sex should be removed. The second model analyzing four random rater effects indicated that image quality should be removed. The third and final model combining three significant random child effects (i.e., random slope, skin tone, and image quality) and three significant random rater effects (i.e., random slope, child assigned sex, and skin tone), as well as the two significant intercept random effects, revealed that both the slope and intercept were significant such that the rate of change in APCA increased over time and the intercept at age 12 significantly differed from zero.

A conditional linear multilevel growth model assessed the unique and interactive effects of child assigned sex (dummy coded; female = 1),

Table 5
Unconditional linear growth models: random effects analysis results.

Model	Effect	AIC	p
Child Effects	Slope	69,823	< 0.001
	Child Sex	69,416	0.710
	Skin Tone	69,444	< 0.001
	Image Quality	69,484	< 0.001
Rater Effects	Slope	68,980	< 0.001
	Child Sex	68,407	< 0.05
	Skin Tone	68,379	< 0.001
	Image Quality	68,373	0.121

child ethnicity/race (effect-coded with white as the reference group), skin tone ratings (mean-centered), rater assigned sex (dummy coded; female = 1), rater ethnicity/race (effect coded with other as the reference group), rater child exposure (mean-centered), rater child contact (mean-centered), and their interactions on APCA from ages 4 to 12 (see Table 6). The LRT indicated this conditional model fit significantly better than the unconditional model.

There were three significant findings regarding child assigned sex. Although the first finding indicated that female sex predicted higher APCA, the second finding revealed that this effect was qualified by a significant interaction with chronological age (i.e., linear slope). As shown in Fig. 2, college student raters perceived boys as being older than girls from ~ ages 4–5.5, after which girls were perceived as older than boys. The third finding was a significant interaction between child assigned sex and Latine rater ethnicity/race such that the overestimation of girls' ages compared to boys' ages was more pronounced among non-Latine ($b = 0.212$, $p = 0.104$) raters as compared to Latine raters ($b = 0.089$, $p = 0.526$), though neither simple slope attained significance. *Note.* Simple slopes are displayed. Child sex was dummy coded as 0 = male and 1 = female.

There were two significant findings regarding child ethnicity/race. First, a significant interaction with Latine rater ethnicity/race indicated that Latine raters perceived non-Latine children as younger ($b = -0.134$, $p = 0.193$), whereas non-Latine raters perceived non-Latine children as older ($b = -0.208$, $p = 0.040$). Second, as shown in Fig. 3, a significant interaction with chronological age suggested that raters perceived Black children as younger than non-Black children until ~ age 6, after which they perceived Black children as older. Finally, consistent with this Black child ethnicity/race effect, a significant interaction between college students' child skin tone ratings and their concurrent perceptions of chronological age indicated that college students rated children they perceived as having darker skin tone as younger than children they perceived as having lighter skin tone until ~ age 6, after which they rated children they perceived as having darker skin tone as older (see Fig. 4).

Regarding rater characteristics, there was a significant interaction between rater assigned sex and child chronological age. As shown in Fig. 5, female raters overestimated children's ages more than male raters from ~ ages 4–9, after which male raters overestimated children's ages more than female raters. Regarding rater ethnicity/race, there was a significant interaction between Asian rater ethnicity/race and child skin tone ratings such that Asian raters were more likely to rate children they perceived as having lighter skin tone as younger and children they perceived as having darker skin tone as older ($b = 0.074$, $p = 0.024$), but this pattern was not significant for non-Asian raters ($b = 0.025$, $p = 0.298$). Finally, there were no significant main or interactive effects of rater child exposure or rater child contact in the prediction of APCA.

Within-time pubertal analysis at age 12

A within-time analysis at age 12 explored the extent to which the model was impacted by the inclusion of child pubertal status. Table 7 depicts the parameter estimates and 95 % bootstrapped confidence intervals (CI) across 10,000 resamples for each predictor. A fully saturated path model wherein perceived age ratings were regressed onto child assigned sex, mean-centered skin tone ratings, mean-centered pubertal status, and their interactions revealed a significant main effect of pubertal status. Specifically, college students rated children who reported that they were further along in their pubertal development as older than children with lower self-reported pubertal development scores. Although there were no significant main effects of child assigned sex or skin tone ratings, there was a significant interaction such that college students perceived girls with lighter skin tone as significantly older than those with darker skin tone, but there was no impact of skin tone on APCA for boys (see Fig. 6).

Post-hoc within-time regression at age 4

Intrigued by the dramatic age overestimation of nearly 2 years at the age 4 data wave, we conducted a post-hoc, within-time regression to examine child assigned sex, child skin tone, and their interaction as potential correlates of APCA at Wave 1 (age 4). Table 8 depicts the parameter estimates and 95 % bootstrapped CIs across 10,000 resamples for each predictor. Skin tone emerged as the only significant predictor such that college students perceived children with darker skin tone as older than children who were perceived to have lighter skin tone.

Discussion

This study advanced our understanding of APCA by achieving two primary goals. First, unconditional growth models identified patterns of APCA from the preschool period through early adolescence. Second, a conditional growth model evaluated differences in the slope and intercept of APCA as a function of child characteristics, rater characteristics, and their interactions, both with one another and with child chronological age. Together, ten hypotheses evaluated 1) APCA patterns across childhood (H1,H2), 2) APCA and child characteristics (H3-H6), and 3) APCA and rater characteristics (H7-H10). Regarding APCA patterns across childhood, college student raters perceived children as being older than their chronological age at every data wave, and this overestimation was most pronounced in early development. Regarding child characteristics, child assigned sex, ethnicity/race, and skin tone were associated with APCA. Regarding rater characteristics, rater assigned sex, ethnicity/race, child exposure, and child contact were not significantly associated with APCA. However, several significant interactions further qualified these patterns. Together, these findings advance the neophyte literature on APCA by documenting its expression from the preschool period into early adolescence as related to both child and rater characteristics.

APCA patterns across childhood (H1, H2)

As predicted, there was significant variation in APCA over time (i.e., slope; H1) and at age 12 (i.e., intercept; H2). On average, college student raters perceived children as being older than their chronological age at every data wave. However, this overestimation effect was most pronounced in early development with an average 2-year inflation of age 4 images, followed by a gradual improvement in accuracy over time such that age estimates at age 12 were inflated by only ~8 months. The appearance of age overestimation in this study is somewhat in line with prior studies of adults' age ratings of child images. For example, Goff et al. (2014) found that adults (i.e., college students and police officers) overestimated children's ages across photos of Black, Latino, and white males ages 10–17. However, since the pattern of findings was similar across ages, these authors collapsed their data to yield unitary age difference scores (i.e., perceived age minus chronological age) across ethnic/racial groups and by crime types. In this series of studies, overestimations of children's age across ethnic/racial groups and by crime types ranged from ~2–4 years. Interestingly, whereas the current study suggests adults may be more accurate when estimating the ages of older children, Goff et al. (2014) found a comparatively larger overestimation effect when collapsing perceived age estimates across a broader and older range of child participants (i.e., 10–17). This accuracy difference may reflect the current focus on a community sample of child participants, whereas Goff et al. (2014) presented raters with a sample of purported perpetrators for whom age overestimation may have been more pronounced.

Due to the use of differing age ranges and child populations in prior studies, it is difficult to directly compare the current findings with previous research. For example, Cooke and Halberstadt (2021) examined parents' age ratings of images of Black and white girls and boys, but similarly grouped across ages 10–13, yielding an average perceived age

Table 6

Conditional linear growth model of APCA on child characteristics, rater characteristics, and their interactions.

	Fixed Effect	*B*	*β*	*SE*	*df*	*t*	*p*
	Intercept	12.467	< 0.001	0.287	633.644	43.479	< 0.001
	Child Age (linear slope)	**0.850**	**0.731**	**0.039**	**563.013**	**21.548**	**< 0.001**
Child Female Sex	**Child Female Sex**	**0.583**	**0.090**	**0.227**	**388.919**	**2.562**	**0.011**
	Child Female Sex*Child Age	**0.091**	**0.079**	**0.027**	**203.044**	**3.364**	**0.001**
	Child Female Sex*Child Ethnicity/race (Latine)	0.080	0.012	0.192	241.946	0.416	0.678
	Child Female Sex*Child Ethnicity/race (Multi)	0.118	0.016	0.226	241.754	0.522	0.602
	Child Female Sex*Child Ethnicity/race (Black)	−0.099	−0.012	0.281	250.659	−0.353	0.724
	Child Female Sex*Child Skin Tone	−0.046	−0.016	0.045	296.756	−1.028	0.305
	Child Female Sex*Rater Female Sex	−0.023	−0.005	0.076	394.760	−0.298	0.766
	Child Female Sex*Rater Ethnicity/race (Latine)	**−0.123**	**−0.018**	**0.059**	**395.498**	**−2.092**	**0.037**
	Child Female Sex*Rater Ethnicity/race (Asian)	0.036	0.005	0.059	398.516	0.610	0.542
	Child Female Sex*Rater Ethnicity/race (Multi)	−0.048	−0.007	0.063	391.162	−0.766	0.444
	Child Female Sex*Rater Child Exposure	−0.032	−0.007	0.042	400.793	−0.757	0.449
	Child Female Sex*Rater Child Contact	0.039	0.006	0.055	401.581	0.705	0.482
	Child Ethnicity/race (Latine)	0.058	0.012	0.188	410.188	0.306	0.760
	Child Ethnicity/race (Latine)*Age	0.011	0.012	0.021	208.148	0.520	0.604
	Child Ethnicity/race (Latine)*Child Skin Tone	0.014	0.004	0.035	314.709	0.407	0.684
	Child Ethnicity/race (Latine)*Rater Female Sex	−0.079	−0.024	0.048	14,497.846	−1.656	0.098
	Child Ethnicity/race (Latine)*Rater Ethnicity/race (Latine)	**0.074**	**0.012**	**0.037**	**14,489.588**	**1.993**	**0.046**
	Child Ethnicity/race (Latine)*Rater Ethnicity/race (Asian)	0.047	0.008	0.037	14,517.824	1.271	0.204

	Fixed Effect	*B*	*β*	*SE*	*df*	*t*	*p*
Child Ethnicity/race	Child Ethnicity/race (Latine)*Rater Ethnicity/race (Multi)	0.029	0.004	0.040	14,484.320	0.742	0.458
	Child Ethnicity/race (Latine)*Rater Child Exposure	−0.030	−0.007	0.027	14,408.869	−1.096	0.273
	Child Ethnicity/race (Latine)*Rater Child Contact	−0.003	−0.001	0.035	14,464.823	−0.100	0.921
	Child Ethnicity/race (Multi)	−0.124	−0.022	0.225	411.437	−0.550	0.583
	Child Ethnicity/race (Multi)*Child Age	0.020	0.019	0.025	206.291	0.814	0.417
	Child Ethnicity/race (Multi)*Child Skin Tone	0.019	0.006	0.040	290.421	0.478	0.633
	Child Ethnicity/race (Multi)*Rater Female Sex	0.069	0.018	0.057	14,463.804	1.218	0.223
	Child Ethnicity/race (Multi)*Rater Ethnicity/race (Latine)	−0.020	−0.003	0.044	14,430.913	−0.461	0.645
	Child Ethnicity/race (Multi)*Rater Ethnicity/race (Asian)	0.011	0.001	0.044	14,472.158	0.240	0.811
	Child Ethnicity/race (Multi)*Rater Ethnicity/race (Multi)	−0.075	−0.009	0.047	14,410.329	−1.581	0.114
	Child Ethnicity/race (Multi)*Rater Child Exposure	−0.006	−0.001	0.032	14,469.497	−0.193	0.847
	Child Ethnicity/race (Multi)*Rater Child Contact	0.034	0.005	0.041	14,486.445	0.826	0.409
	Child Ethnicity/race (Black)	0.018	0.003	0.292	736.519	0.063	0.950
	Child Ethnicity/race (Black)*Child Age	**−0.088**	**−0.073**	**0.032**	**330.189**	**−2.787**	**0.006**
	Child Ethnicity/race (Black)*Child Skin Tone	−0.039	−0.013	0.052	521.837	−0.749	0.454
	Child Ethnicity/race (Black)*Rater Female Sex	−0.072	−0.017	0.093	14,229.437	−0.778	0.436
	Child Ethnicity/race (Black)*Rater Ethnicity/race (Latine)	−0.043	−0.005	0.072	14,309.731	−0.590	0.555
	Child Ethnicity/race (Black)*Rater Ethnicity/race (Asian)	−0.002	< 0.001	0.072	14,176.021	−0.028	0.978
	Child Ethnicity/race (Black)*Rater Ethnicity/race (Multi)	−0.027	−0.003	0.079	14,331.227	−0.349	0.727
	Child Ethnicity/race (Black)*Rater Child Exposure	0.036	0.006	0.053	14,206.524	0.680	0.496
	Child Ethnicity/race (Black)*Rater Child Contact	−0.063	−0.008	0.068	14,362.997	−0.929	0.353
Child Skin Tone	Child Skin Tone	0.064	0.032	0.063	1688.360	1.023	0.306
	Child Skin Tone*Child Age	**0.019**	**0.049**	**0.007**	**5588.433**	**2.841**	**0.005**

	Fixed Effect	*B*	*β*	*SE*	*df*	*t*	*p*
	Child Skin Tone*Rater Child Exposure	−0.011	−0.005	0.017	1082.938	−0.627	0.531
	Child Skin Tone*Rater Child Contact	0.030	0.011	0.022	1233.991	1.341	0.180
Rater Sex	Rater Female Sex	−0.132	−0.019	0.170	404.655	−0.777	0.438
	Rater Female Sex*Child Age	**−0.089**	**−0.121**	**0.024**	**392.993**	**−3.729**	**< 0.001**
	Rater Female Sex*Child Skin Tone	0.046	0.032	0.031	1179.659	1.484	0.138
	Rater Female Sex*Rater Ethnicity/race (Latine)	−0.268	−0.080	0.219	392.205	−1.220	0.223
	Rater Female Sex*Rater Ethnicity/race (Asian)	0.228	0.070	0.210	393.953	1.084	0.279
	Rater Female Sex*Rater Ethnicity/race (Multi)	0.173	0.052	0.216	392.320	0.800	0.424
	Rater Female Sex*Rater Child Exposure	0.048	0.019	0.173	391.754	0.280	0.779
	Rater Female Sex*Rater Child Contact	−0.166	−0.055	0.201	395.644	−0.825	0.410
Rater Ethnicity/race	Rater Ethnicity/race (Latine)	0.097	0.021	0.311	452.440	0.311	0.756
	Rater Ethnicity/race (Latine)*Child Age	−0.028	−0.030	0.019	393.034	−1.491	0.137
	Rater Ethnicity/race (Latine)*Child Skin Tone	−0.035	−0.012	0.025	1228.545	−1.443	0.149
	Rater Ethnicity/race (Latine)*Rater Child Exposure	0.021	0.005	0.119	392.608	0.178	0.859
	Rater Ethnicity/race (Latine)*Rater Child Contact	0.110	0.018	0.159	394.317	0.687	0.492
	Rater Ethnicity/race (Asian)	−0.516	−0.110	0.307	454.816	−1.680	0.094
	Rater Ethnicity/race (Asian)*Child Age	−0.032	−0.035	0.019	393.145	−1.750	0.081
	Rater Ethnicity/race (Asian)*Skin Tone	**0.048**	**0.017**	**0.024**	**1167.686**	**2.001**	**0.046**
	Rater Ethnicity/race (Asian)*Rater Child Exposure	−0.044	−0.009	0.125	392.378	−0.348	0.728
	Rater Ethnicity/race (Asian)*Rater Child Contact	0.105	0.017	0.155	395.221	0.678	0.498
	Rater Ethnicity/race (Multi)	−0.106	−0.021	0.329	451.324	−0.323	0.747
	Rater Ethnicity/race (Multi)*Child Age	0.017	0.017	0.020	392.292	0.833	0.406
	Rater Ethnicity/race (Multi)*Child Skin Tone	−0.010	−0.003	0.026	1332.021	−0.398	0.690
	Fixed Effect	*B*	*β*	*SE*	*df*	*t*	*p*

(continued on next page)

Table 6 (*continued*)

	Fixed Effect	B	β	SE	df	t	p
	Rater Ethnicity/race (Multi)*Rater Child Exposure	0.020	0.004	0.144	392.049	0.139	0.889
	Rater Ethnicity/race (Multi)*Rater Child Contact	−0.049	−0.008	0.158	394.427	−0.308	0.758
Rater Child Exposure & Contact	Rater Child Exposure	0.157	0.046	0.247	440.326	0.638	0.524
	Rater Child Exposure*Child Age	0.018	0.026	0.013	398.570	1.307	0.192
	Rater Child Contact	0.178	0.041	0.297	449.499	0.600	0.549
	Rater Child Contact*Child Age	0.009	0.010	0.017	395.388	0.501	0.617
	Rater Child Contact*Rater Child Exposure	−0.071	−0.014	0.103	394.921	−0.695	0.488

Note. Boldface text denotes significant pathway.
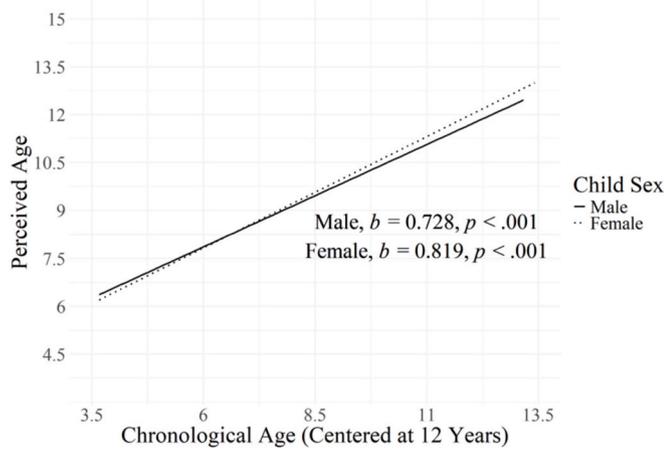
Fig. 2. Interaction between child chronological age (linear slope) and child assigned sex.

of 10.6 years old, which is lower than the average perceived age of 11.569 years obtained in the current study when collapsing across ages 10–12. That said, a recent study by Koch et al. (2023) provides a suitable comparison to some of the current findings because they only used images of 12-year-olds. Although the average perceived age of their participants ranged from 10.68 for low maturation white males to 14.27 for high maturation white females, the overall mean perceived age across conditions was 12.13 years old, which is comparable to the average perceived age of 12.91 years old in the current study.

APCA and child characteristics (H3-H6)

Child assigned sex at birth and ethnicity/race were related to APCA patterns over time. Regarding assigned sex, our hypothesis that girls would be perceived as older than boys at all data waves (i.e., ages 4–12; H3) was only partially confirmed because boys were perceived as older than girls during early development (i.e., ~ ages 4–5.5), after which girls were perceived as older. The significance of female assigned sex as a predictor of APCA, particularly at later ages, mirrors well-established patterns of earlier maturation among girls than boys (Brix et al., 2019). Consistent with this interpretation, our within-time analysis at age 12 showed that adults perceived children who reported being further along in their pubertal development at age 12 as older than their less mature peers. That said, given the developmental significance of pubertal timing and tempo (Beltz et al., 2014), future studies with time-varying measures of puberty at all waves will be best-suited to support a comprehensive investigation of child assigned sex, pubertal development, and APCA.

Regarding ethnicity/race, our hypothesis that adults would perceive Black children as older than children from all other ethnic/racial groups over time (H4) was supported starting ~ age 6, but not during the preschool period. This finding is inconsistent with Cooke and Halberstadt (2021) study of older children, which showed that parents' age ratings did not differ significantly between groups of Black and white children (ages 10–13). Further work will be needed to determine whether these inconsistent findings reflect different child characteristics (e.g., children of different ages) and/or rater characteristics (e.g., college students versus parents). Our hypothesis that college students would rate children whom they perceived as having darker skin tone as older than children whom they perceived as having lighter skin tone (H5) was partially supported. Prior to ~ age 6, raters underestimated the
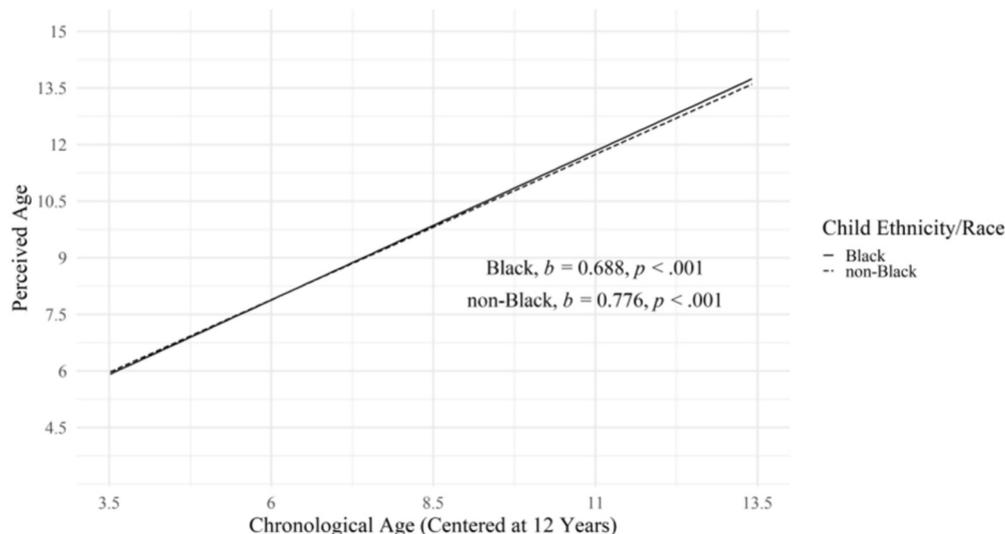

Fig. 3. Interaction between child chronological age (i.e., linear slope) and black child ethnicity/race effect.
Note. Simple slopes are displayed. Child ethnicity/race was effect-coded as Black = 1, multi-ethnic/racial, Latine = 0, and white = −1.
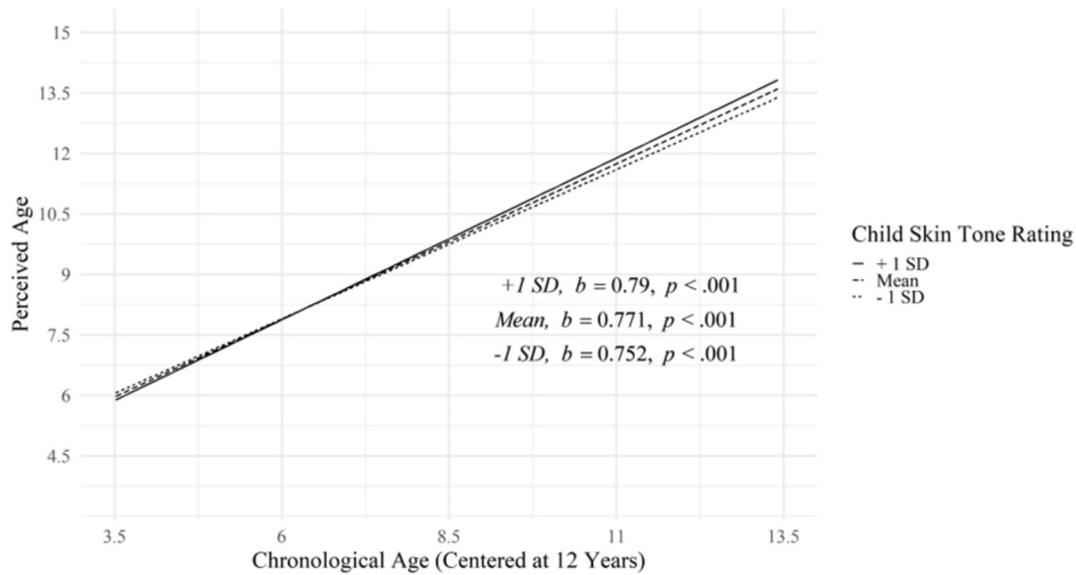
Fig. 4. Interaction between child chronological age (i.e., linear slope) and child skin tone ratings.
Note. Simple slopes are displayed for skin tone ratings at +1 standard deviation [*SD*] above the mean, at the mean, and at −1 *SD* below the mean. Higher skin tone ratings indicate darker perceived skin tone.
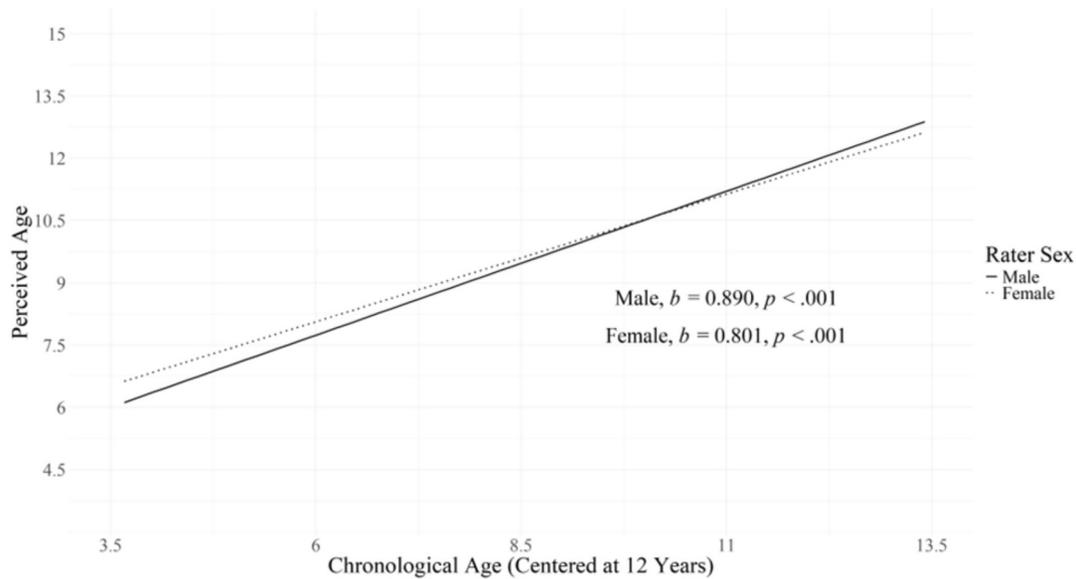

Fig. 5. Interaction between child chronological age (i.e., linear slope) and rater assigned sex.
Note. Simple slopes are displayed. Rater sex was dummy coded as 0 = male and 1 = female.

age of children with darker skin tone, but the hypothesized pattern of overestimation was evident for older children with darker skin tone.

Interestingly, the post-hoc, within-time analysis at age 4 revealed that the robust age overestimation of nearly 2 years during the preschool period was solely predicted by skin tone ratings, but in an opposite pattern, such that 4-year-old children who were rated as having darker skin tone were perceived as older than children with lighter skin tone. It is important to note that these findings are based on a much simpler regression model than the multilevel models used in the primary analyses. However, despite this discrepancy in model complexity, these post-hoc findings do parallel qualitative research suggesting that adults perceive Black children (ages 2–16 years) as older (and less innocent) than white children (Powell & Coles, 2021).

Considering interactive patterns, our hypothesis that age overestimation effects would be greatest for girls with darker skin tone was

not supported (H6). However, the within-time analysis at age 12 did reveal a significant interaction between child assigned sex and skin tone such that female sex predicted older age estimates, but only for girls with lighter skin tone ratings. Although this pattern is consistent with Koch et al. (2023) who found the greatest age overestimation among white high maturation females, our findings should be interpreted with caution because the same raters reported on both APCA and perceived skin tone at the same time, which limits causal inference.

APCA and rater characteristics (H7-H10)

Overall, our hypotheses regarding rater characteristics were not supported, but there were several unexpected findings. Our predictions that raters would provide more accurate age ratings for children who shared their assigned sex (H7) and ethnicity/race (H8) were not

Table 7

Within-time regression including pubertal status as a predictor of age 12 perceived age.

Predictors	B	b (Bootstrapped SE)	p (95 % CI)
Female Sex	0.036	0.116 (0.246)	0.637 (−0.340, 0.618)
Skin Tone	0.078	0.094 (0.116)	0.420 (−0.136, 0.322)
Pubertal Status	0.604	1.573 (0.300)	< 0.001 (0.964, 2.134)
Female Sex*Skin Tone	−0.240	−0.424 (0.187)	0.023 (−0.776, −0.040)
Female Sex*Pubertal Status	−0.191	−0.655 (0.417)	0.116 (−1.466, 0.157)
Female Sex*Skin Tone*Pubertal Status	0.021	0.050 (0.182)	0.784 (−0.311, 0.411)

supported. This contradicts previous literature suggesting that people are better at recognizing faces that share their own gender (e.g., own-gender bias; Herlitz & Lovén, 2013) and race (i.e., other-race effect; Meissner & Brigham, 2001). These discrepant findings could be attributed to the fact that prior research protocols were recognition-based (i. e., does the participant recognize the target from an array?), whereas APCA is perception-based (i.e., how does the participant evaluate or perceive the target?). It is also important to note that the differing ethnic/racial groups found in our child sample (i.e., Latine, Black, multiethnic/racial, and white) as compared to our college student rater sample (i.e., Latine, Asian, multi-ethnic/racial, and other) prevented us from properly evaluating same ethnic/racial biases in this study (i.e., H8).

Our final hypotheses focused on raters' familiarity with children such that we expected raters reporting greater child exposure (H9) and higher frequency child contact (H10) would provide more accurate age ratings over time compared to raters reporting less child exposure and contact. The obtained findings did not support these hypotheses, which contradicts prior studies (e.g., Kuefner et al., 2008). However, as with prior studies examining recognition patterns based on assigned sex or

ethnicity/race, the limited work on adults' familiarity with children has been recognition- rather than perception-based.

Importantly, there were four significant findings involving rater assigned sex and rater ethnicity/race that we did not anticipate. However, as the first study to consider relations between rater characteristics and APCA over time, these findings should be interpreted with caution pending replication. First, there was a significant interaction between rater assigned sex and child chronological age such that female raters overestimated children's ages more than male raters from ∼ ages 4–9, after which male raters overestimated children's ages more than female raters. Second, there was a significant interaction between the Latine rater ethnicity/race effect and child assigned sex such that Latine raters were more likely to overestimate girls' ages as compared to boys' ages than were non-Latine raters, though neither group-specific simple slope attained significance. Third, Latine raters perceived non-Latine children as younger while non-Latine raters perceived non-Latine children as older. Fourth, compared to non-Asian raters, Asian raters were more likely to rate children they perceived as having lighter skin tone as younger while rating children they perceived as having darker skin tone as older.

Summary

Descriptively, the obtained results suggest that child assigned sex, child ethnicity/race, child skin tone, and both rater assigned sex and rater ethnicity/race are related to APCA across time with interesting

Table 8

Post-hoc within-time regression of child assigned sex at and child skin tone as predictors of APCA at age 4.

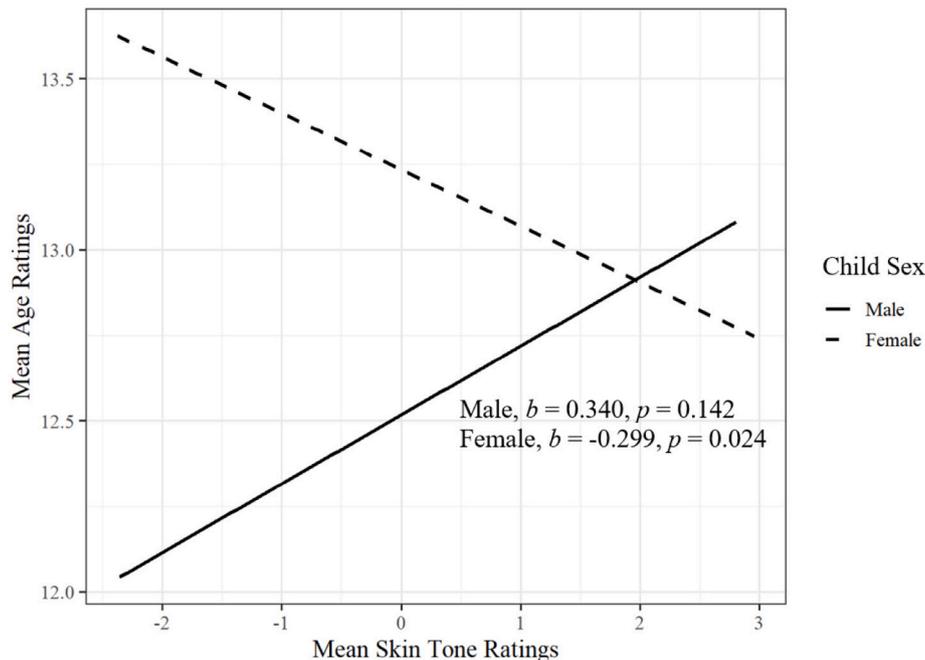| Predictors	B	b (Bootstrapped SE)	p (95 % CI)
Child Female Sex	−0.045	−0.096 (0.144)	0.505 (−0.375, 0.186)
Skin Tone	0.234	0.174 (0.069)	0.012 (0.034, 0.307)
Child Female Sex*Skin Tone	−0.107	−0.133 (0.098)	0.251 (−0.305, 0.085)

Fig. 6. Interaction between child skin tone ratings and child assigned sex at age 12.
Note. Simple slopes are displayed. Child sex was dummy coded as 0 = male and 1 = female.

developmental shifts in the direction of these predicted relations, particularly ~ age 6. For example, boys were perceived as older than girls during early development (i.e., ~ ages 4–5.5), after which girls were perceived as older. These findings suggest that the transition to formal schooling (i.e., ~ age 6) may be a salient turning point for APCA. Additionally, rater characteristics, particularly in combination with child features, were significantly related to APCA, though must be interpreted with caution given their concurrent evaluation.

Strengths and limitations

The current study featured several strengths. First, longitudinal data afforded images of the same children to support the examination of APCA over a period of nine years. This design dramatically extends extant studies spanning four (Cooke & Halberstadt, 2021) and seven years (Goff et al., 2014), particularly given that both prior studies collapsed ratings across age, rather than examining child chronological age as a meaningful factor in its own right. Furthermore, our ability to provide images of the same children at different ages reduced numerous confounds affecting prior studies that featured variability in both the age and the identity of the target child. Second, with 1000 child stimuli, this study offered the largest dataset of any APCA investigation to date (previously, the largest data set was 40 stimuli; Cooke & Halberstadt, 2021). Third, both the child targets and college student raters were diverse in terms of assigned sex at birth and ethnicity/race. Regarding child targets, the current study extends beyond prior investigations of single-sex child targets (Goff et al., 2014) or of only Black-white ethnic/racial differences (e.g., Cooke & Halberstadt, 2021). Our capacity (and effort) to consider rater characteristics in this study is similarly valuable. Prior APCA-related studies have utilized samples with little variation in raters' assigned sex (e.g., 77 % female in Koch et al., 2023; 7 % female in Goff et al., 2014) and/or ethnicity/race (e.g., 74 % white in Epstein et al., 2017; 89.8 % white in Goff et al., 2014). Importantly, the current sample of raters was also diverse with respect to child exposure (i.e., 53.1 % reported living with at least one minor child) and contact (i.e., 41.5 % reported very frequent contact with minor children). To our knowledge, this is the first study to evaluate if and how rater characteristics relate to APCA in their own right, over time, and in interaction with child characteristics. Fourth, the multilevel modeling in this study yielded more accurate parameter estimates by appropriately accounting for variation between both the child targets and the college student raters (Smith & DeFrates-Densch, 2016). Fifth, this study is the first to evaluate APCA as related to both categorical groupings of child ethnicity/race and a continuous metric of child skin tone, as well as to their interactions with child chronological age and assigned sex. Finally, by adopting a sociodevelopmental perspective on APCA, this study posed and answered questions that prior studies have not been able to consider, including how APCA changes over time and what accounts for these changes.

Despite these notable strengths, however, several limitations warrant consideration when interpreting the current findings. First, although diverse with respect to assigned sex, ethnicity/race, socioeconomic status, child exposure, and child contact, these college student raters were fairly homogeneous in terms of parental status (i.e., only 2 of the 400 raters indicated they were parents), age (i.e., most were in their early twenties), and education (i.e., all were college students). Some studies have examined raters who are authority figures in children's lives, such as police officers (Goff et al., 2014) and parents (Cooke & Halberstadt, 2021), but more work is needed to understand APCA among additional adult rater groups, such as teachers, childcare providers, school security personnel, and other adult professionals who work directly with children (e.g., pediatricians). It would be especially informative to incorporate raters who work in K – 12 school settings (e. g., teachers, school security personnel) because decisions made by teachers and other school staff greatly affect a child's developmental trajectory as evinced by the notorious school-to-prison pipeline (Barnes

& Motz, 2018). Although school personnel are likely to know the age of students, teachers, administrators, and security staff may nevertheless act on biased age perceptions when interpreting and responding to a student's behavior in the moment. Indeed, Zebrowitz (1997) found that even when parents were shown the ages of children next to their yearbook image, they still assigned more cognitively demanding chores to children with more mature looking faces as compared to "baby faced" children. The same trend emerged when adults made punishment judgements such that "baby faced" children received less severe punishments for the same offenses. In a unique experimental study by Carter et al. (2018), elementary school teachers were randomly presented with vignettes about children's externalizing or academic behaviors that included drawings of Black or white 10-year-old girls in varying stages of pubertal development (i.e., early, on-time, or late). After the stimulus presentation, teachers estimated each girl's future academic and social functioning. Despite knowing the chronological age of the girls in the vignettes, teachers expected worse academic and social functioning for early-maturing girls. Moreover, compared to their early-maturing white peers, teachers expected early-maturing Black girls to have more problems interacting and relating to others, as well as increased difficulty with acquiring and using information in the academic context. These findings show that explicit knowledge of a child's chronological age does not eliminate the potential for problematic social expectations and actions driven by biased social perceptions to undermine children's educational adaptation. Interestingly, no research to date, including the current study, has examined APCA among adults who have the potential to actually interact with (and directly impact) the children depicted in the images they are rating. Future studies using teachers' ratings of students from the same school district, but in different grades or schools, or police officers' ratings of children who live in a neighborhood that is part of the area they patrol may address some of these limitations. Moving forward, it will be important to obtain ratings from various stakeholders in children's lives to understand the perceptions of adults who work with children in general, and the implications of APCA on the children with whom those adults work directly.

Second, because the child images in this study were extracted from video footage, they could not be standardized for image quality and child apparel, among other factors. At the same time, these variations may have strengthened this study to the degree that they rendered the target images more ecologically valid and reflective of real-life situations. Indeed, even the current standardization of images to headshots of children with neutral expressions may have impacted APCA patterns because we rarely view other people as static entities from the shoulders up. Images showing the entire child have been used in prior studies of perceived maturity (e.g., Johnson & Collins, 1988; Koch et al., 2023) and would likely help to refine our understanding of APCA. Likewise, video recordings of children would provide more comprehensive, realistic, and dynamic information as compared to static images. Although there is some support for the validity of image-based ratings (e.g., attractiveness ratings obtained from static images are largely equivalent to those obtained from video clips; e.g., Kościński, 2013; Rhodes & Anastasi, 2012), future research on APCA will benefit from careful replications including perceptual stimuli that are static versus dynamic (e.g., still images versus video clips), partial versus complete (e.g., headshots versus full-body images), and standardized versus naturally occurring (e.g., holding facial expression, apparel, or other factors constant versus allowing for random variation).

Third, ratings of child ethnicity/race and skin tone were entered as predictors of APCA in the full model but are more appropriately conceptualized as APCA correlates because they were collected from the same informant and at the same time as the APCA ratings. Thus, raters' perceptions of child ethnicity/race and skin tone may have been influenced by their perceptions of the child's age in ways that could not be examined here. Future research should utilize multiple informants of child ethnicity/race and skin tone to support causal predictions of APCA.

Finally, due to limited pubertal status data, the impact of puberty

could not be assessed in the current growth models. A supplemental, within-time analysis at age 12 revealed that the obtained sex effect resulting in girls being perceived as older than boys became non-significant when pubertal status was held constant, while the interaction between sex and skin tone became significant. Specifically, at age 12, adults rated girls with lighter skin tone as older than girls with darker skin tone, but skin tone was not related to APCA for boys. Future research should incorporate pubertal status and tempo as time-varying predictors in growth models to assess their relations with APCA across time.

Conclusion and implications

The current study provides new information regarding the phenomenology of APCA from the preschool period to early adolescence, as well as preliminary data regarding how child characteristics, rater characteristics, and their interactions relate to APCA across development. Future research should extend the current study by examining 1) APCA across different periods of development, 2) peer perceptions, and 3) the extent to which APCA impacts child development, particularly in educational and community settings.

The current study examined APCA from ages 4 to 12. However, the supplemental, within-time analysis at age 12 highlighted the potential significance of pubertal development for understanding APCA in later development. If replicated with longitudinal data extending further into adolescence, this finding has important implications for what might be contributing to differences in APCA and, by extension, for resultant differences in how children are treated by adults. For example, if puberty contributes to age misperceptions, it may be important to incorporate child development education into existing training and/or continuing education materials for professionals who interact with children as part of their job (e.g., law enforcement, school security personnel), especially later in adolescence when adults may misperceive children as legal adults with particularly serious negative consequences.

There is an ongoing need for research regarding peer perceptions of children's age. Although we were not able to find studies wherein children provided perceived age ratings of their peers, a wealth of literature regarding the socioemotional implications of pubertal timing suggests that early (perceived) pubertal timing, which is typically associated with children looking older than their chronological age, predicts later peer victimization (Reynolds & Juvonen, 2011; Skoog & Kapetanovic, 2022) and substance use (Bucci et al., 2021). Likewise, extensive research documents the negative consequences of early maturation for girls with respect to increased mental health problems and risk behaviors. For example, in a study of 822 girls, earlier breast development was associated with greater mental health symptoms for Latina and white girls but was protective for Asian girls (Deardorff et al., 2021). Regarding risk behaviors, a study of over 2000 girls showed that early puberty was associated with stronger associations between girls own delinquent behaviors and their best friend's deviance, suggesting increased susceptibility to negative peer influences among early-maturing girls (Mrug et al., 2014). As such, future research will benefit from studies examining how, not only adult, but also peer perceptions of children's age may interface with pubertal development (and other child characteristics, such as assigned sex and ethnicity/race) to influence socioemotional outcomes across adolescence.

Finally, perhaps the most challenging and important next step for future APCA research will be to examine the extent to which APCA predicts later child outcomes in applied contexts (e.g., schools). As noted earlier, research using raters who actually interact with the child targets under study and including measures of child adaptive outcomes will help to clarify the practical significance of APCA. Likewise, it is important to note that obtained patterns of statistical significance in this study may not equate to practical significance. For example, a significant interaction between child ethnicity/race and rater ethnicity/race (Latine vs. non-Latine) accounted for slight differences in APCA (i.e., less

than half a year), which may be of modest practical significance. That said, the profound age overestimation seen at age 4 in this study, coupled with a post-hoc analysis revealing child skin tone was the sole significant correlate of this overestimation warrants further examination. If replicated, this finding has important implications for how children of color may enter the school-to-prison pipeline. Indeed, the fact that this age group is especially vulnerable to age overestimation may explain disproportionalities in school disciplinary actions, such that, for example, Black preschool children are 3.6 times more likely to be suspended than white preschool children (United States Department of Education, 2016).

We know that adults' decisions directly impact children. In schools, adults decide which children get detention and which children get suspended. On the streets, adults decide who is a threat and who is not. To the degree that APCA influences these decisions, we must replicate and extend the current findings in future research. After all, since it is adults who have these biased perceptions, it is up to adults to better understand them and to mitigate their potential impact on child development.

Preregistration

The analyses presented here were preregistered. The preregistration is available at the following URL: https://aspredicted.org/sk3xz.pdf

CRediT authorship contribution statement

Jessie M. Bridgewater: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Tuppett M. Yates:** Writing – review & editing, Supervision, Funding acquisition.

Funding sources

Declaration of competing interest

We have no conflict of interest associated with this publication.

Acknowledgements

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.appdev.2025.101896.

Data availability

Data will be made available on request.

References

Afifi, A. A., Kotlerman, J. B., Ettner, S. L., & Cowan, M. (2007). Methods for improving regression analysis for skewed continuous or counted responses. *Annual Review of Public Health, 28*, 95–111.

Barnes, J. C., & Motz, R. T. (2018). Reducing racial inequalities in adulthood arrest by reducing inequalities in school discipline: Evidence from the school-to-prison pipeline. *Developmental Psychology, 54*(12), 2328–2340. https://doi.org/10.1037/dev0000613

Bates D, Maechler M, Bolker B, Walker S (2014). lme4: Linear Mixed-Effects Models Using Eigen and S4. R package version 1.1-7, URL http://CRAN.R-project.org/package=lme4.

Beltz, A. M., Corley, R. P., Bricker, J. B., Wadsworth, S. J., & Berenbaum, S. A. (2014). Modeling pubertal timing and tempo and examining links to behavior problems. *Developmental Psychology, 50*(12), 2715–2726. https://doi.org/10.1037/a0038096

Berry, D. S., & McArthur, L. Z. (1985). Some components and consequences of a babyface. *Journal of Personality and Social Psychology, 48*(2), 312–323. https://doi.org/10.1037/0022-3514.48.2.312

Brix, N., Ernst, A., Lauridsen, L. L. B., Parner, E., Støvring, H., Olsen, J., … Ramlau-Hansen, C. H. (2019). Timing of puberty in boys and girls: A population-based study. *Paediatric and Perinatal Epidemiology, 33*(1), 70–78. https://doi.org/10.1111/ppe.12507

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design.* Harvard University Press.

Bucci, R., Staff, J, Maggs, J. L., & Dorn, L. D. (2021). Pubertal timing and adolescent alcohol use: The mediating role of parental and peer influences. *Child Development, 92*(5), e1017–e1037. https://doi.org/10.1111/cdev.13569

Carter, R., Mustafaa, F. N., & Leath, S. (2018). Teachers' expectations of girls' classroom performance and behavior: Effects of girls' race and pubertal timing. *The Journal of Early Adolescence, 38*(7), 885–907. https://doi.org/10.1177/0272431617699947

Carter, R., & Seaton, E. K. (2025). Rethinking pubertal research: Embracing intersectionality. *Child Development Perspectives, 19*(2), 80–91. https://doi.org/10.1111/cdep.12527

Cooke, A. N., & Halberstadt, A. G. (2021). Adultification, anger bias, and adults' different perceptions of black and white children. *Cognition and Emotion, 35*(7), 1416–1422. https://doi.org/10.1080/02699931.2021.1950127

Crenshaw, K., Ocen, P., & Nanda, J. (2015). *Black girls matter: Pushed out, overpoliced and underprotected.* Faculty Scholarship. https://scholarship.law.columbia.edu/faculty_scholarship/3227.

Davis, J., & Marsh, N. (2020). Boys to men: The cost of 'adultification' in safeguarding responses to black boys. *Critical and Radical Social Work, 8*(2), 255–259. https://doi.org/10.1332/204986020X15945756023543

Deardorff, J., Marceau, K., Johnson, M., Reeves, J. W., Biro, F. M., Kubo, A., … Hiatt, R. A. (2021). Girls' pubertal timing and tempo and mental health: A longitudinal examination in an ethnically diverse sample. *The Journal of Adolescent Health, 68*(6), 1197–1203. https://doi.org/10.1016/j.jadohealth.2021.01.020

Dumas, M. J., & Nelson, J. D. (2016). (Re)imagining black boyhood: Toward a critical framework for educational research. *Harvard Educational Review, 86*(1), 27–47. https://doi.org/10.17763/0017-8055.86.1.27

Epstein, R., Blake, J., & González, T. (2017). *Girlhood interrupted: The erasure of black girls' childhood (SSRN scholarly paper ID 3000695).* Social Science Research Network. https://doi.org/10.2139/ssrn.3000695

Galliher, R. V., McLean, K. C., & Syed, M. (2017). An integrated developmental model for studying identity content in context. *Developmental Psychology, 53*(11), 2011–2022. https://doi.org/10.1037/dev0000299

García Coll, C., Lamberty, G., Jenkins, R., McAdoo, H. P., Crnic, K., Wasik, B. H., & Vázquez García, H. (1996). An integrative model for the study of developmental competencies in minority children. *Child development, 67*(5), 1891–1914.

Gibson, J. J. (2014). *The ecological approach to visual perception.* Psychology Press. https://doi.org/10.4324/9781315740218

Goff, P. A., Jackson, M. C., Di Leone, B. A. L., Culotta, C. M., & DiTomasso, N. A. (2014). The essence of innocence: Consequences of dehumanizing black children. *Journal of Personality and Social Psychology, 106*(4), 526–545. https://doi.org/10.1037/a0035663

Halberstadt, A. G., Castro, V. L., Chu, Q., Lozada, F. T., & Sims, C. M. (2018). Preservice teachers' racialized emotion recognition, anger bias, and hostility attributions. *Contemporary Educational Psychology, 54*, 125–138. https://doi.org/10.1016/j.cedpsych.2018.06.004

Halberstadt, A. G., Cooke, A. N., Garner, P. W., Hughes, S. A., Oertwig, D., & Neupert, S. D. (2022). Racialized emotion recognition accuracy and anger bias of children's faces. *Emotion, 22*(3), 403–417. https://doi.org/10.1037/emo0000756

Hall, A. V., Hall, E. V., & Perry, J. L. (2016). Black and blue: Exploring racial bias and law enforcement in the killings of unarmed black male civilians. *American Psychologist, 71*(3), 175.

Hannon, L., DeFina, R., & Bruch, S. (2013). The relationship between skin tone and school suspension for African Americans. *Race and Social Problems, 5*(4), 281–295. https://doi.org/10.1007/s12552-013-9104-z

Herlitz, A., & Lovén, J. (2013). Sex differences and the own-gender bias in face recognition: A meta-analytic review. *Visual Cognition, 21*(9–10), 1306–1336. https://doi.org/10.1080/13506285.2013.823140

Horváth, K., & Plunkett, K. (2016). Frequent daytime naps predict vocabulary growth in early childhood. *Journal of Child Psychology and Psychiatry, 57*(9), 1008–1017. https://doi.org/10.1111/jcpp.12583

Johnson, B. M., & Collins, W. A. (1988). Perceived Maturity as a Function of Appearance Cues in Early Adolescence: Ratings by Unacquainted Adults, Parents, and Teachers. *The Journal of Early Adolescence, 8*(4), 357–372. https://doi.org/10.1177/0272431688084004

Keith, V. M., Lincoln, K. D., Taylor, R. J., & Jackson, J. S. (2010). Discriminatory experiences and depressive symptoms among African American women: Do skin tone and mastery matter? *Sex roles, 62*(1), 48–59. https://doi.org/10.1007/s11199-009-9706-5

Koch, M. K., Burd, K. A., & Mendle, J. (2023). Same crime, same time? Differences in visual maturity affect opinions of adolescent culpability. *Applied Cognitive Psychology*. https://doi.org/10.1002/acp.4092

Kościński, K. (2013). Perception of facial attractiveness from static and dynamic stimuli. *Perception, 42*(2), 163–175. https://doi.org/10.1068/p7378

Kuefner, D., Macchi Cassia, V., Picozzi, M., & Bricolo, E. (2008). Do all kids look alike? Evidence for an other-age effect in adults. *Journal of Experimental Psychology: Human Perception and Performance, 34*(4), 811. https://doi.org/10.1037/0096-1523.34.4.811

Lamb, S., & Koven, J. (2019). Sexualization of girls: Addressing criticism of the APA report, presenting new evidence. *SAGE Open, 9*(4), Article 2158244019881024. https://doi.org/10.1177/2158244019881024

McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review, 90*(3), 215–238. https://doi.org/10.1037/0033-295X.90.3.215

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*(1), 3–35. https://doi.org/10.1037/1076-8971.7.1.3

Montepare, J. M., & Zebrowitz, L. A. (1998). Person perception comes of age: The salience and significance of age in social judgments. *Advances in Experimental Social Psychology, 30*, 93–161. https://doi.org/10.1016/S0065-2601(08)60383-4

Mrug, S., Elliott, M. N., Davies, S., Tortolero, S. R., Cuccaro, P., & Schuster, M. A. (2014). Early puberty, negative peer influence, and problem behaviors in adolescent girls. *Pediatrics, 133*(1), 7–14. https://doi.org/10.1542/peds.2013-0628

Nielsen, B. R., Linneberg, A., Christensen, K., & Schwarz, P. (2015). Perceived age is associated with bone status in women aged 25-93 years. *Age (Dordrecht, Netherlands), 37*(6), 106. https://doi.org/10.1007/s11357-015-9842-5

Nishina, A., Bellmore, A., Witkow, M. R., Nylund-Gibson, K., & Graham, S. (2018). Mismatches in self-reported and meta-perceived ethnic identification across the high school years. *Journal of Youth and Adolescence, 47*(1), 51–63. https://doi.org/10.1007/s10964-017-0726-0

Parascandola, R. (2012, August 4). *Harlem teen says she was cuffed, held in MetroCard flap.* New York Daily News. https://www.nydailynews.com/2012/08/04/harlem-teen-says-she-was-cuffed-held-in-metrocard-flap/.

Petersen, A. C., 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*(2), 117–133. https://doi.org/10.1007/BF01537962

Poutvaara, P., Jordahl, H., & Berggren, N. (2009). Faces of politicians: Babyfacedness predicts inferred competence but not electoral success. *Journal of Experimental Social Psychology, 45*(5), 1132–1135. https://doi.org/10.1016/j.jesp.2009.06.007

Powell, T., & Coles, J. A. (2021). 'We still here': Black mothers' personal narratives of sense making and resisting antiblackness and the suspensions of their black children. *Race Ethnicity and Education, 24*(1), 76–95. https://doi.org/10.1080/13613324.2020.1718076

Reynolds, B. M., & Juvonen, J. (2011). The role of early maturation, perceived popularity, and rumors in the emergence of internalizing symptoms among adolescent girls. *Journal of Youth and Adolescence, 40*(11), 1407–1422. https://doi.org/10.1007/s10964-010-9619-1

Rhodes, M. G., & Anastasi, J. S. (2012). The own-age bias in face recognition: A meta-analytic and theoretical review. *Psychological Bulletin, 138*(1), 146–174. https://doi.org/10.1037/a0025750

Roth, W. D. (2016). The multiple dimensions of race. *Ethnic and Racial Studies, 39*(8), 1310–1338. https://doi.org/10.1080/01419870.2016.1140793

Saperstein, A., Kizer, J. M., & Penner, A. M. (2016). Making the most of multiple measures: Disentangling the effects of different dimensions of race in survey research. *American Behavioral Scientist, 60*(4), 519–537. https://doi.org/10.1177/0002764215613399

Skiba, R. J., Horner, R. H., Chung, C.-G., Rausch, M. K., May, S. L., & Tobin, T. (2011). Race is not neutral: A national investigation of African American and Latino disproportionality in school discipline. *School Psychology Review, 40*(1), 85–107. https://doi.org/10.1080/02796015.2011.12087730

Skoog, T., & Kapetanovic, S. (2022). The role of pubertal timing in the development of peer victimization and offending from early- to mid-adolescence. *The Journal of Early Adolescence, 42*(1), 5–32. https://doi.org/10.1177/02724316211002265

Smith, M. C., & DeFrates-Densch, N. (2016). *Challenges and innovations in educational psychology teaching and learning.* IAP.

Stumper, A., Graham, A. A., Abramson, L. Y., & Alloy, L. B. (2020). Pubertal synchrony and depressive symptoms: Differences by race and sex. *Journal of Youth and Adolescence, 49*(11), 2275–2284. https://doi.org/10.1007/s10964-020-01314-x

Thompson, M. S., & McDonald, S. (2016). Race, skin tone, and educational achievement. *Sociological Perspectives, 59*(1), 91–111. https://doi.org/10.1177/0731121415580026

United States Department of Education. (2016). 2013-14 civil rights data collection: A first look. https://www2.ed.gov/about/offices/list/ocr/docs/CRDC2013-14-first-look.pdf.

Voegeli, R., Schoop, R., Prestat-Marquis, E., Rawlings, A. V., Shackelford, T. K., & Fink, B. (2021). Differences between perceived age and chronological age in women: A multi-ethnic and multi-centre study. *International Journal of Cosmetic Science, 43*(5), 547–560. https://doi.org/10.1111/ics.12727

Zebrowitz, L. A. (1997). *Reading faces: Window to the soul?* Routledge. https://doi.org/10.4324/9780429493188

Zurbriggen, E. L., Collins, R. L., Lamb, S., Roberts, T.-A., Tolman, D. L., Ward, L. M., & Blake, J. (2007). *Report of the APA task force on the sexualization of girls* (p. 72). American Psychological Association, Task Force on the Sexualization of Girls. http://www.apa.org/pi/women/programs/girls/report-full.pdf.