

Development and psychometric validation of a child Racial Attitudes Index (RAI)

Khaya D. Clark^{1,2} · Paul Yovanoff³ · Charlotte Ursula Tate⁴

Published online: 26 January 2017
© Psychonomic Society, Inc. 2017

Abstract The Racial Attitudes Index (RAI) measures a child's racial attitudes. Designed for children aged 5–9 years, the RAI is delivered over the Internet using Audio Computer Assisted Self-Interviewing (ACASI). Unlike traditional binary forced-choice instruments, the RAI uses an expanded response format permitting a more nuanced understanding of patterns of children's racial attitudes. In addition to establishing psychometric evidence of the RAI technical adequacy, hypotheses about RAI item response patterns were tested. The racial attitudes of 336 Black and White children in grades K–3 were assessed using a forced-choice instrument (Preschool Racial Attitudes Measure II) and the RAI. Findings from this study indicate measures obtained with the RAI are technically adequate, and the measure functions invariantly across racial groups. Also, patterns of children's

racial attitudes measured with the RAI are more nuanced than those obtained using the forced-choice response format.

Keywords Racial and ethnic attitude and relations · Prejudice · Childhood development · Social cognition

Evidence from explicit attitudes research shows that young children display racial biases as early as 3 to 5 years of age (Aboud & Doyle, 1996; Bernstein, Zimmerman, Werner-Wilson, & Vosburg, 2000; Bigler & Liben, 1993; Doyle & Aboud, 1995; Gibson, Robbins, & Rochat, 2015; Jones, Parker, Joyner, & Steiner-Ulku, 1999; Levy & Hughes, 2009; Nesdale, 2000). White children, in grades kindergarten to second grade, tend to negatively evaluate Blacks while holding extremely positive views of Whites (Aboud, 1988; Baron & Banaji, 2004; Bernstein et al., 2000; Gibson et al., 2015; Katz & Kofkin, 1997; McGlothlin & Killen, 2006; Williams & Morland, 1976). Unlike White children, Black children are more heterogeneous in their biases (Aboud, 1988; Williams & Morland, 1976): they often indicate high levels of pro-White bias (Cross & Cross, 2008; Rice, Ruiz, & Padilla, 1974) or low levels of ingroup preference (Gibson et al., 2015). In general, Black children and other minority groups are theorized to display either a lack of ingroup preference or an outgroup preference for Whites because the sociopolitical context supports and reflects the marginalized status of minority groups vis-à-vis Whites (Baron & Banaji, 2009; Blumer, 1958; Bobo & Hutchings, 2003). Parallel finding occur for US Latina/o children. In a study of explicit racial attitudes, Latina/o children display an outgroup preference for Whites when compared to Latina/o (Bernat & Balch, 1979). These findings for Black and Latina/o children in the US support the principle of system justification theory that dominant views of racial status and hierarchy are internalized at a very young

✉ Khaya D. Clark
clakha@ohsu.edu

Paul Yovanoff
pyovanoff@mail.smu.edu

Charlotte Ursula Tate
ctate2@sfsu.edu

¹ Department of Family Medicine, Oregon Health and Science University, 3181 SW Sam Jackson Park Road, FM, Portland, OR 97239, USA

² Department of Sociology, Emory University, 1555 Dickey Drive, Atlanta, GA 30322, USA

³ Annette Caldwell Simmons School of Education and Human Development, Southern Methodist University, Dallas, TX 75205, USA

⁴ Department of Psychology, San Francisco State University, 1600 Holloway Avenue, Ethnic Studies and Psychology Building, San Francisco, CA 94132, USA

age (Baron & Banaji, 2009). This framework sensitizes researchers to the fact that the attitudes and racial biases of minority groups toward their own groups and other groups are relatively nuanced-based power and status differences between groups (Jost, Banaji, & Nosek, 2004; Smith & Mackie, 2002).

Explicit measures of childhood racial attitudes

Since the 1970s, two of the most commonly used explicit racial attitudes measures are the Preschool Racial Attitudes Measure II [PRAM II] (Williams & Morland, 1975) and the Katz-Zalk Projective Prejudice Test [KZPP] (Katz & Zalk, 1976). While research using these scales has contributed to our understanding of early emergence of attitudes in children, several important methodological issues have been highlighted about the response format, race-of-examiner effects, and the number of targets and perceivers presented in explicit scales. We describe these measurement issues and then discuss the development of a new scale that addresses some of these issues.

The KZPP and PRAM II both use a picture-story technique to assess racial attitudes toward Black and White individuals (Williams & Morland, 1976). An adult examiner reads mini-stories about stimulus images to a child then asks questions about each one. A typical story for both measures is: “Here are two boys. One is a smart boy. He gets A’s on all of his spelling tests. Which is the smart boy?” Children are instructed to choose one of two children. The KZPP and PRAM II structures require young children to choose just one target (e.g., the White child or the Black child). One noted limitation with forced-choice response formats for the assessment of racial attitudes is that these formats confound preference for one group with rejection of the other group. The response format does not explicitly allow children the options to say that no group has a particular attribute (“selfish”), or that both groups do. Thus, it is difficult to tell if a child’s choice of her ingroup reflects ingroup preference or outgroup derogation.

In an effort to separately assess positive and negative attitudes toward racial groups, Aboud and colleagues added a “both” category to the Multi-response Racial Attitude Measure [MRA] (Aboud, 2003; Doyle & Aboud, 1995). Research participants are instructed to apply positive or negative attributes to the ingroup, outgroup, or both groups. From a psychometric standpoint, however, there remains a problem of interpretation for negative attributes for the “both” category. As Clark and Tate (2008) argue:

“As the ‘both’ response option *includes the ingroup*, this option may not be chosen for negative attributes when a child wants to show preference for the ingroup. Thus, the child would simply choose the outgroup for the negative attributes *because it is the only option that does not include the ingroup*. Yet, if a child were allowed to

choose ‘neither’ for negative attributes, researchers would be able to more clearly interpret this result.”

Cameron et al. (2001) report usage of the MRA and its prototype measure (Doyle et al., 1988) usually do not include the “both” response in analyses, opting instead to analyze the familiar difference score between ingroup-positive and outgroup-negative responses (Clark & Tate, 2008).

Another issue with explicit measures of racial attitudes is the few studies that report on race-of-examiner effects, or how these effects are minimized through the training of research staff. In some early research with the PRAM II, researchers found race-of-examiner effects in PRAM II preschool and primary school standardization studies and at least one later KZPP study (Glover & Smith, 1997; Williams, Best, Boswell, Mattson & Graves, 1975; Williams, Best & Boswell, 1975), yet, to our knowledge, the effect is inconsistent across studies. For example, in Aboud’s review of race-of-examiner effects, she suggests that this effect might be more salient for younger children as they have a higher need for adult approval than older children (Aboud, 1988). One technique that holds promise for minimizing race-of-examiner effects and controlling for social desirability is audio computer-assisted self-interviewing [A-CASI] (Borgers, de Leeuw, & Hox, 2000; de Leeuw, Hox, & Kef, 2003). A-CASI is a technique in which a computerized voice-over reads the question to the respondent, ensuring that questions are delivered verbatim and voice inflection of the narrator is constant and consistent. Self-interviewing techniques have been shown to foster more candid reporting regarding sensitive topics in the area of racial attitudes research with adults (Krysan, 1998). Within the context of racial attitudes research with children, A-CASI administration removes the role of adult-interviewer in the administration of items, which could improve the validity of responses by allowing children to control the question and answer process on their own terms (Krumpal, 2013).

The single-target, single-perceiver structure of the PRAM II which includes drawings of Black and White people only, limits our understanding of racial attitudes in an increasingly multiracial context (Bobo & Hutchings, 2003). As noted previously, recent empirical research grounded in system justification theory indicates that minority children’s lack of ingroup preference signals a complexity in attitudinal patterns that cannot easily be examined in single-target, single perceiver instruments (Clark & Tate, 2008). One solution would be to develop an instrument that includes Asians, Blacks, Whites, and Latinos as both targets and perceivers in a single instrument. We revisit this topic in the Discussion section. What is more, and the focus of this paper, the single-target, single-perceiver instruments cannot effectively model or showcase the nuances that we have described in terms of US Black and Latina/o children that are known to exist in the literature. In fact, all the studies that demonstrate these nuanced ingroup

and outgroup attitudes rely on instruments that provide at least two types of targets (e.g., Baron & Banaji, 2009; Dunham et al., 2007). Accordingly, to effectively model the fuller range of perceiver and target dynamics, researchers should develop an instrument that includes at least two targets. Moreover, the two-target presentation should be coupled with multiple response options so that researchers can effectively disambiguate the response patterns that would indicate ingroup favoritism and outgroup negativity, for example. This latter piece (disambiguating response patterns) has yet to be fully or consistently realized in a single measurement instrument.

The purpose of this study is, therefore, to describe the development and psychometric validation of a new tool called the Racial Attitudes Index (RAI) that includes expanded response options and is delivered using A-CASI methodology. In addition, the approach provides a possible structure into including multiple-perceivers and multiple-targets in one instrument that could advance interdisciplinary understanding of racial attitudes in a multiracial context. In study one, we describe the development of the stimulus images, in study two we describe an item-reduction activity that was designed to develop the final 40-item RAI scale used in study three, and in study three, we describe the psychometric findings of the RAI compared with the PRAM II.

Study one: Photo classification activity

Method

Participants

Twenty-six children in grades one through three in Eugene, OR, USA were recruited to participate. Two elementary classrooms participated. Participants were identified as Asian ($n = 1$), Black ($n = 1$), Other ($n = 8$), and White ($n = 16$) by their parents. The other category was defined as White plus some other racial or ethnic group.

Materials

A set of 31 photos were cropped and laminated for use in this task. Eight photos were of Black girls, eight photos depicted Black boys, eight photos were of White girls, and seven¹ photos were of White boys. The primary research aim was to assess whether the individual photos were perceived as belonging in either Black or White racial groups.

¹ Eight photos were taken of White boys, but the photographer misplaced one of the photos. Therefore, we had only seven photos to use in this task.

Procedures

Research participants completed the photo classification activity in a classroom with other children, but they were instructed to refrain from watching other children as they completed the task. Participants were given an envelope that contained laminated photos (stimuli), a blank piece of paper, and paper clips. The blank piece of paper was used to cover up each child's pile(s) of photos. Research participants were asked to place the photos in baskets of people "who belong together, but don't put the photos together based on boys and girls." The children used a paper clip to affix groups together. The proctor placed the groupings into each child's envelope. A unique subject identification number was written on the outside of each envelope.

Results

The classification findings showed that overall 84% ($n = 22$) of the participants sorted the photos based on racial group (Black and White), while the remaining 16% ($n = 4$) sorted the photos based on race and gender. The percentage of participants by grades was first grade (15%), second grade (58%), and third grade (27%). Sorting the photos by racial group was not related to grade level (first, second, and third), Fisher's exact $p \geq .320$, two-tailed, and Cramer's $V = .177$, or to gender, Fisher's exact $p = .626$, two-tailed, Cramer's $V = .134$. In addition, accurate sorting of the photos into racial groups was not related to self-identified race, Fisher's exact $p = .630$, two-tailed, Cramer's $V = .146$. (Cramer's V values that are close to 1.0 indicate a strong association between the variables.)

Discussion

Based on these results, we concluded that the stimulus images were robust representations of Black and White racial groups. Therefore, all stimuli were included in the RAI instrument.

Study two: Item-reduction activity

Method

Participants

Seventy-nine children in grades 2–3 recruited from two schools in Oregon and one school in Washington participated in the item-reduction activity. Participants were assigned to complete a counterbalanced 56-item RAI (version 1 or version 2). In version 1, 57% of the participants were identified by parents as White, 14% as African American, 5% did not answer the question, and 23% were identified as the

Table 1 Version A item order

Item number	Item type	Left side	Right side	Pairs
1.	Gender	White girl	White boy	-
2.	+	Black	White	Boys
3.	-	White	Black	Girls
4.	+	Black	White	Girls
5.	Gender	White boy	White girl	-
6.	-	Black	White	Girls
7.	Gender	White boy	White girl	-
8.	+	White	Black	Girls
9.	-	Black	White	Boys
10.	Gender	Black girl	Black Boy	-
11.	+	Black	White	Boys
12.	-	White	Black	Girls
13.	Gender	White girl	White boy	-
14.	+	White	Black	Boys
15.	-	Black	White	Girls
16.	Gender	Black boy	Black girl	-
17.	-	White	Black	Boys
18.	+	White	Black	Boys
19.	Gender	Black girl	Black boy	-
20.	-	White	Black	Girls
21.	+	Black	White	Boys
22.	-	Black	White	Boys
23.	Gender	Black girl	Black boy	-
24.	+	White	Black	Boys
25.	Gender	Black girl	Black boy	-
26.	-	Black	White	Boy
27.	+	White	Black	Girls
28.	Gender	White boy	White girl	-
29.	-	White	Black	Girls
30.	+	Black	White	Boys
31.	Gender	Black boy	Black girl	-
32.	-	Black	White	Girls
33.	+	White	Black	Boys
34.	Gender	White girl	White boy	-
35.	+	Black	White	Girls
36.	-	White	Black	Boys
37.	Gender	White boy	White girl	-
38.	+	Black	White	Girls
39.	-	White	Black	Boys
40.	-	Black	White	Girls
41.	+	White	Black	Boys
42.	Gender	Black boy	Black girl	-
43.	-	White	Black	Girls
44.	+	Black	White	Boys
45.	-	White	Black	Boys
46.	+	White	Black	Girls
47.	-	Black	White	Boys
48.	Gender	Black boy	Black girl	-
49.	+	White	Black	Girls

Table 1 (continued)

Item number	Item type	Left side	Right side	Pairs
50.	-	Black	White	Boys
51.	+	Black	White	Girls
52.	-	White	Black	Boys
53.	Gender	White girl	White boy	-
54.	+	White	Black	Girls
55.	-	Black	White	Girls
56.	+	Black	White	Girls

Note. The item type column indicates the valence of the question for the racial attitude items (+ or -), or whether the item is a gender bias item. Twenty of the racial attitude items are positive (e.g., “which child is pretty?”), 20 are negative (e.g., “which child did not help the teacher?”), and the remaining 16 are gender-bias items (“Which child wants to be a fire-fighter?”). The “Left-side” and “Right-side” columns indicate on which side of the screen the image should be placed. For the racial attitude items, the “Pairs” column conveys the gender of the stimulus pairs. When an “-” appears in the column “Pairs,” this is a crossed gender-bias item, and the information is specified in the “Left-side” and “Right-side” columns

“other”² category. The other category was operationalized as children that were identified as White and some other racial or ethnic group. The composition of the sample in version 1 was second grade (14%) and third-grade (86%), and 51% were male and 49% female. In version 2, 55% of the participants were identified by parents as White, 11% as African American, 5% did not answer the question, and 29% were identified as the “other” category. The composition of the sample in version 2 was second grade (61%) and third-grade (39%), and 64% were male and 36% female.

Materials

Two counterbalanced versions of a 56-item RAI were developed and compared to assess stimulus image order variation and variation by item valence (positive and negative). In version A of the RAI, item one showed a Black child on the left side of the screen and a White child on the right side. In version B of item one, the position of the images was switched so that the Black child was on the left side and the White child was on the right side. One group of children completed version A (44%), and one group completed version B (56%). Table 1 shows the order in which the items occurred in version A. In version B, the stimuli were switched, but the question order was the same as in version A.

We hired a professional photographer to photograph children’s faces used in the stimulus images. The photographer took eight photos of each child against a gray paper background.

² In version 1, the “other” group of participants ($n = 8$) was composed of White and Latina/o children, and in version 2, the “other” group of participants ($n = 13$) was composed of White and Latina/o children and White and Asian children.

We asked the photographer to capture a range of facial expressions of each child (e.g., happy, neutral, sad), and to take mid-body to head shots so that pant and shoe style would not have to be controlled. Each child was given a white t-shirt to wear during the photo shoot. After the children were photographed, the photographer cropped the photos to standardize the appearance of the targets. The Principal Investigator reviewed all the photos and selected the pairings for the stimulus images. We created maximally similar pairs of photos by using facial expression and age range to control for attractiveness.

Item reduction measure

Before participants answered the attitude items, the computer program presented subjects with a computer mouse assessment to verify that group differences were not due to enhanced mouse-use skill and that individual participants were able to use the mouse correctly. Research participants were asked to click on three graphic faces consecutively (green, yellow, and blue). When participants chose the correct face, an applause voice-over played. For incorrect answers, a remedial voice-over stated that this was not the green, yellow, or blue face and to “please try again.” Research participants were allowed two remedial trials. If a child did not answer three out of five questions correctly, the program auto-advanced to a pause screen, and a voice-over stated, “Please raise your hand to ask for help.” All research participants successfully completed the mouse-training activity.

After completion of the computer mouse-training activity, four instructional screens provided a description of the activity to follow. These instructional screens were identical in versions A and B of the RAI. On the first instructional screen, the voice-over stated, “We have some pictures we'd like to show you and some stories that go with each one. This is not a test, so there are no right or wrong answers. Just do your best. Now, we'll show you what the questions are like and how to use the buttons.” The graphic that accompanied this voice-over is depicted in Fig. 1.

When the last part of the voice-over stated, “Now we'll show you what the questions are like and how to use the buttons,” a green highlighted circle appeared on-screen to emphasize the buttons (Fig. 2).

Item reduction activity procedure

After obtaining child assent, teachers directed students to sit down at their computer and to put on headphones. Children were instructed to not look at the person sitting next to them as they answered the questions. Since all the teachers with whom the first author spoke assured her that their students were familiar with the process of entering identification numbers, students were instructed to enter their school code and subject identification number on the log-in screen.

Item-reduction activity variables

Research participants' race, grade, and gender were treated as independent variables. These data were obtained from a demographic questionnaire that was attached to the parental consent form.

Statistical procedures

To test for whether there were differences between the two long versions of the 56-item RAI, we used a non-parametric Kolmogorov-Smirnov Z (KSZ) statistic. The KSZ statistic D is used to compare two independent distributions of categorical data.

Results

For the distribution of responses to each of the 20 racial positive items, no significant differences between the versions were obtained. One item in the negative set of items was significant ($D = 2.22, p < .001$). The question for this item was “Here are two girls. Their class is putting on a play. Who is not in the class play?” This item was excluded from subsequent analyses. For all other items, with the total sample there were no significant differences in the distribution of item responses by order or item valence, i.e., instrument version.

We repeated these analyses of version effect by race, grade, and gender. Because the overall sample by version of the instrument is small, parsing the data by race, grade, or gender further reduces the sample size. When analyzing the differences between version by race, the cell sizes for Black students is five and four for version 1 and version 2, respectively. Likewise, for grade level analyses, only five second-grade students completed version 1. Similar to the overall analyses, item responses were similar across versions irrespective of race, grade, and gender. Version differences were obtained for only four items when analyzing by race, specifically when White students responded to two negatively framed items and when children in the “other” category responded to two negatively framed items ($p < .05$ for the four items). By grade level, only for one negative item at grade three was a version effect obtained ($p < .01$). Finally, for gender, three significant differences were obtained for negatively framed items (one male $p < .05$, two female, $p < .01$).

Discussion

The findings from the item reduction activity suggest that presentation order of the image of child race and item valence did not impact significantly the 40 racial item response distributions in the overall sample. Furthermore, these findings did not change when considering participant race, grade, and gender. Though some version effects were observed, the overall numbers of statistical tests (i.e., 20 negative items, 20 positive

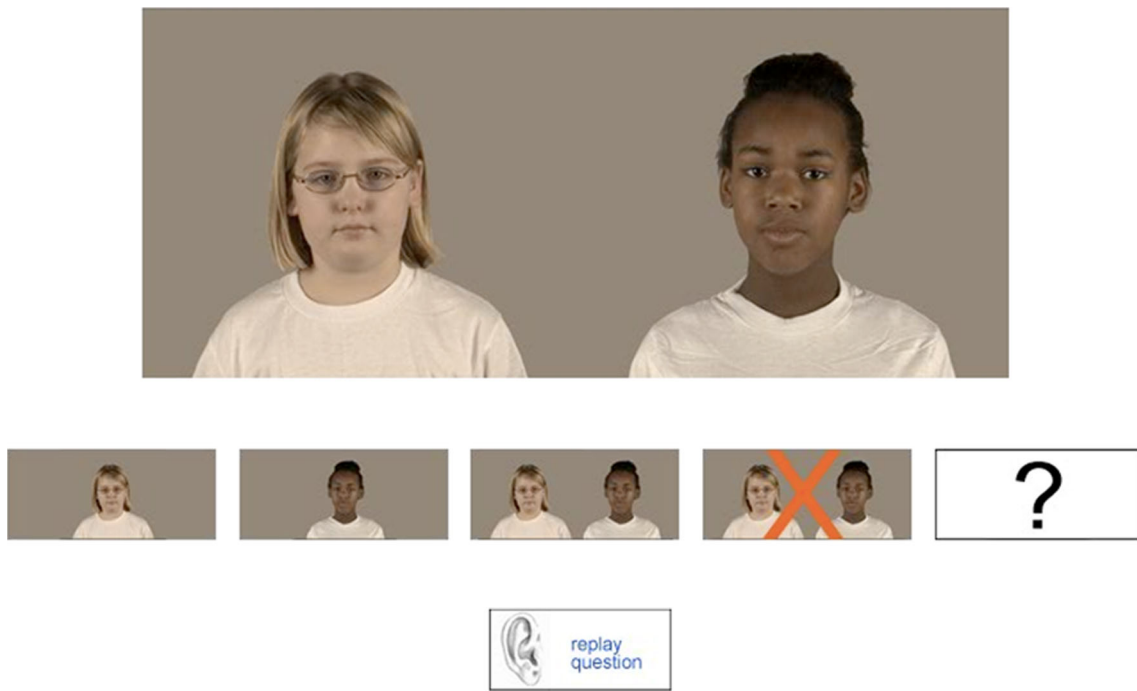


Fig. 1 Instructional screen one

items per group) completed suggest that the results may have been spurious. One possibly important finding is that all significant statistical findings, with the exception of one item, related to negatively framed items. Therefore, we concluded that image placement did not produce significant item response variation. Thus, the task for choosing the final 40 items for the RAI scale was straightforward.

Study three: Psychometric validation

The purpose of study three was to obtain psychometric evidence supporting use of the RAI. We estimated the psychometrics of the RAI using item response modeling procedures, and compared RAI item response category use to PRAM II racial attitude total scores and classifications. Psychometric

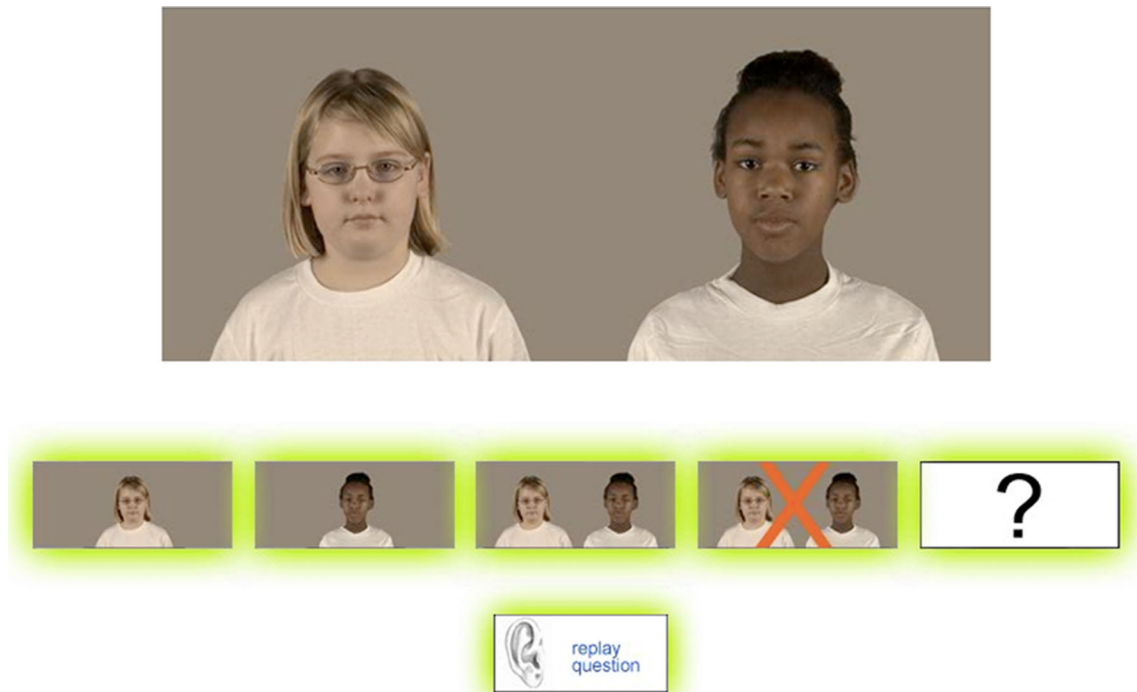


Fig. 2 Button animation for instructional screen one

modeling included testing item response model fit to the RAI response data, analysis of measurement dimensionality, and testing measurement invariance across racial groups. After establishing that the RAI was technically adequate, we tested hypotheses regarding child use of RAI response options by relating RAI item responses to PRAM II racial bias classification. As the contemporary baseline for research on young children's racial bias, the PRAM II is a solid benchmark for assessing the validity of RAI category usage. Additionally, because a substantial amount of discussion pivots around racial differences, we include "race" (Black, White) as a covariate. The purpose for including race is to isolate the sensitivity of the RAI measure to racial differences.

Two competing hypotheses were examined regarding the relation between PRAM II racial bias classification and RAI response option use. Hypothesis 1 can be described as the classic pattern. That is, based on previous empirical research and psychological theory, Hypothesis 1 states that children with high PRAM II total scores (pro-White/anti-Black) or low PRAM II total scores (pro-Black/anti-White) would use the traditional RAI response option of choosing one child, and that children who fell in the "neutral" PRAM II total score range would use the RAI categories of "both", "neither," and "I don't know." Hypothesis 2 is more current, reflecting the idea that the forced selection of the previous measures does not adequately reflect the true attitude space. Hypothesis 2 suggests that if PRAM II classified pro-White/anti-Black and pro-Black/anti-White children are given the response options of "both", "neither," and "I don't know," they would use these in addition to the single child choices. The primary purpose of study three was to test these hypotheses with analyses of obtained patterns of RAI responses conditional on PRAM bias classification.

Table 2 Participant demographics by race

	Black (n = 167)		White (n = 169)	
	n	%	n	%
Child's gender				
Male	74	44.3	86	50.9
Female	93	55.7	83	49.1
Child's grade				
K	53	31.7	35	20.7
1	32	19.2	35	20.7
2	35	21.0	45	26.6
3	47	28.1	54	32.0

Note 1. Race by Gender $\chi^2 = 1.46$, $df = 1$, n.s.

Note 2. Race by Grade $\chi^2 = 5.54$, $df = 3$, n.s.

Method

Participants

For study three, data were collected in the southeast and northwest USA with Black and White children aged 5–9 years. Each participant, recruited from elementary schools, received parental consent and provided assent to participate in the study. A demographics form was affixed to the parental consent form for obtaining information on the child's race, ethnicity, and grade. The process for recruiting schools and school districts was contingent on a school's willingness to participate, making a probability sample unfeasible, so the sampling strategy was a convenience sample (Babbie, 1995). The data collection stopping rule was based on recruiting approximately equal numbers of Black and White children. For the RAI data (Table 2), we had 167 Black and 169 White children. When obtaining PRAM II data (Table 3), we had 167 Black and 167 White children.

The sample includes 42 (12.5%) children from Oregon and 294 children from Atlanta (87.5%). Children were in grades kindergarten through grade 3. Most Oregon participants (52.4%) were in grade 3, while the Georgia participants were equally distributed across the grades K through 3. The Oregon and Georgia participants were equally distributed with respect to gender (52.4% female). All Oregon participants were White, and 43.2% of the Georgia participants were White. Most participants (87.5%) were from Georgia, and therefore the regional variation is regarded as negligible. For these reasons, analyses with respect to the racial bias do not include state as a covariate.

Institutional Review Boards at the University of Oregon, Emory University, the Atlanta Public School District, and School District 4J in Eugene, Oregon all approved the study. Children from diverse backgrounds were recruited (Black, White, and Latino); however, for the purpose of this study, we include White and Black samples only. In our sample, race is independent, statistically, from gender and grade (see Table 2).

Table 3 PRAM II racial bias classification by race

	Black (n = 167)		White (n = 167)	
	n	%	n	%
Definite pro-Black/Anti-White	19	11.4	9	5.4
Probable pro-Black/Anti-White	15	9.0	6	3.6
Neutral	82	49.1	48	28.7
Probable pro-White/Anti-Black	24	14.4	25	15.0
Definite pro-White/Anti-Black	27	16.2	79	47.3

Note 1. $\chi^2 = 41.85$, $df = 4$, $p < 0.01$

Measures

Children completed the RAI and PRAM II. Both instruments were administered using audio computer-assisted self-interview methodology [A-CASI] (Borgers et al., 2000; de Leeuw & Nicholls, 1996). In the A-CASI technique, a computerized voice-over delivers each mini-story to the respondents, ensuring that questions are delivered verbatim and that the voice inflection of the narrator is constant and consistent. The correlation of the RAI and PRAM II bias scores was used as a test of measurement validity. Also, we tested hypotheses about PRAM II bias classification and which children used the range of RAI item response options.

The PRAM II The PRAM II includes 36 color drawings of light-skinned and dark-skinned people: 24 items measure racial attitudes and 12 items measure gender bias. The PRAM II has been used with children between 3 and 12 years of age (Aboud, 1988). Twelve of the racial attitude items include positive adjectives (e.g., smart) and 12 include negative adjectives (e.g., naughty). The drawings show a range of people (younger and older) in identical positions (both standing or sitting). Traditional administration of PRAM II involves an adult examiner, with a child, in a private room, showing the participant each drawing and telling a story about it (Fig. 3).

Figure 3 depicts a PRAM II image, adapted for computerized administration, used in this study with corresponding response option boxes underneath each child. When clicked, the “replay question” box plays the story again. Participants were asked to indicate which child is responsible for a particular act or possesses a particular

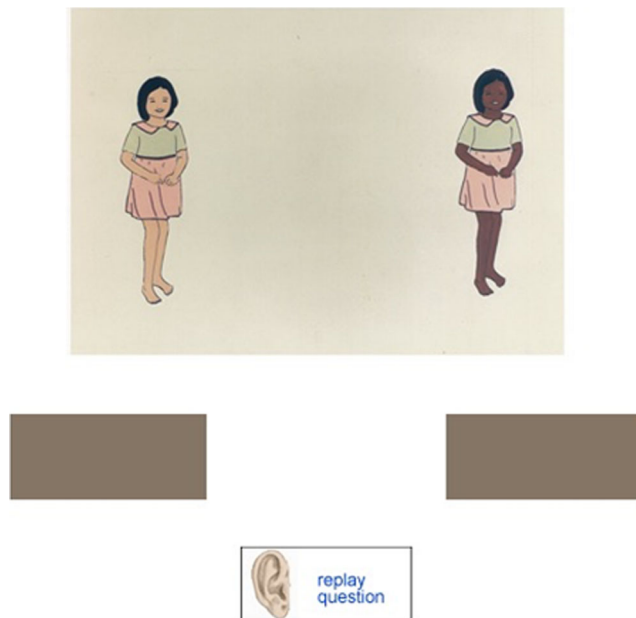


Fig. 3 PRAM II racial attitude item and response options

attribute. One study, using a teaching machine (tape recorder and image flipping device) to administer the PRAM II, yielded comparable scores to those obtained with the face-to-face administration (Best, 2005). Best, a PRAM II author, suggested that our computerized administration of the PRAM II would also yield comparable scores (D. Best, personal communication, 5 April 2005).

The PRAM II total score was used to classify children into one of five bias categories based on scale scores: definite pro-Black/anti-White (0–7), probable pro-Black/anti-White (8–9), neutral (10–14), probable pro-White/anti-Black (15–16), and definite pro-Black/anti-White (17–24). The PRAM II was scored using the scoring procedures outlined in the PRAM administration guide. Specifically, children were assigned a score of 1 if they chose White in relation to a positive adjective and 1 point was given for choosing Black in relation to a negative attribute. A zero was assigned if a child chose Black in relation to a positive item and White in relation to a negative item. The PRAM II score range is 17–24 = definite pro-White/anti-Black bias, 15–16 = probable pro-White/anti-Black bias, 10–14 = unbiased/neutral, 8–9 = probable pro-Black/anti-White, and 0–7 = definite pro-Black/anti-White. All PRAM II stimulus images were counterbalanced for order of racial stimulus images.

As a description of our sampling, we tested the association of gender, grade, and race with the PRAM II racial classification. Though gender and grade are independent of the PRAM II classification, Table 3 summarizes our sample with a cross-tabulation of PRAM II classification by race. Statistically, Black and White children were disproportionately classified. For example, 49% of the Black sample fell within the “neutral” category, whereas 29% of the White sample fell within this category. Similarly, the percentage of participants in the pro-White/anti-Black categories was quite different based on race.

The RAI The study three RAI included 24 items (12 positive and 12 negative). Each item presented either a pair of two boys or two girls, where the only difference between the two children is race. When asked to identify which child is associated with a target behavior, the response options for each of the 24 items are: (a) Black child, (b) White child, (c) both children, (d) neither child, and (e) I don’t know. Figure 4 depicts an RAI image used in a series of instructional screens to show users what the answer choices mean. Boxes underneath each image are response options. To show users answer choice meanings, animation and computer voice-over was used. For example, “If you think both boys did it, click on this button.” Participants were asked to indicate which child is responsible for a particular act or possesses a particular attribute. Not pictured here, there was a replay question option like in the PRAM II.

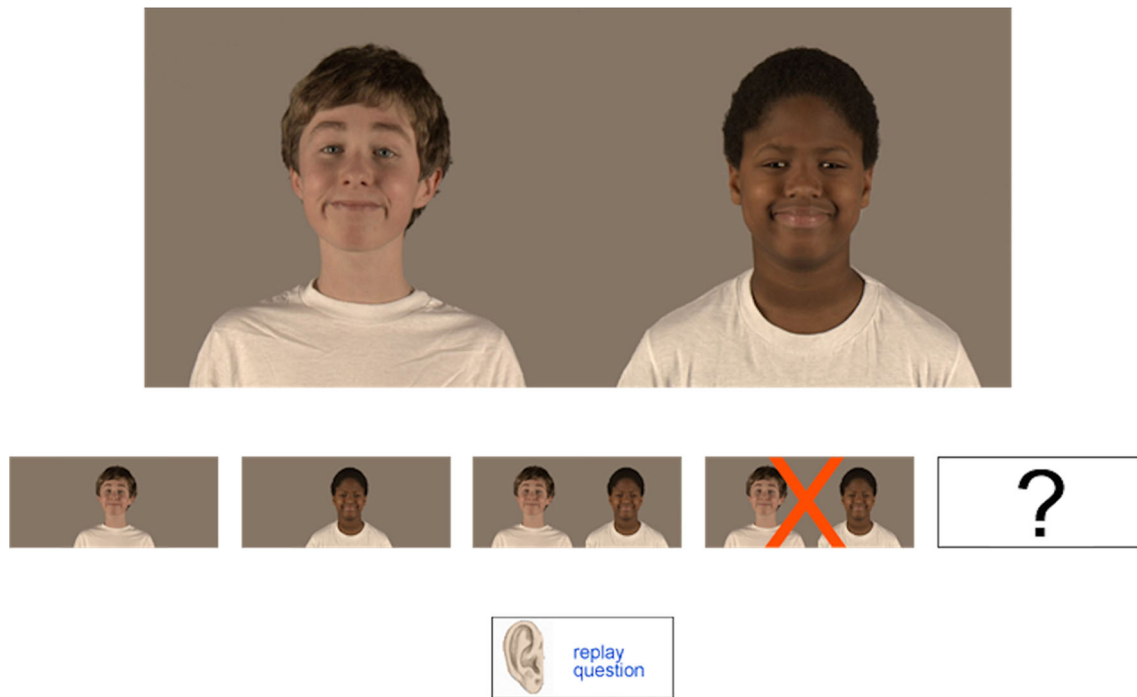


Fig. 4 RAI item and response options

RAI items and response options

Twelve RAI items included negative adjectives (racial prejudice), and 12 items used positive adjectives (racial pride). The racial attitude stimulus images were photos of Black and White children in same-sex, racially mixed pairs. All items were counterbalanced for order of racial stimulus images. Twelve photos were boy pairs and 12 photos were girl pairs.

Procedures

Each child responded to the A-CASI-based RAI and PRAM II assessments. The RAI was administered first and the PRAM was administered 2 days later. All participants completed both instruments in a 1-week period. The RAI was completed in one 20-min session. Two days later, children completed the PRAM II. Children completed both computer-based assessments in school computer labs and wore headphones to ensure privacy. A researcher entered unique identification codes for participants, to ensure de-identification while allowing electronic data files to be matched to demographic information (grade, gender, race). All research participants completed a computer-based mouse-training activity, prior to each assessment (RAI and PRAM II), to ensure that each child knew how to make answer choices by clicking on the mouse. After successfully completing the mouse-training activity, the computer-assisted self-interview (A-CASI) presented a series of instructional screens to show participants a sample question and to show how to use the response options. On the first

instructional screen, the voice-over stated, “We have some pictures that we would like to show you and some stories that go with each one. This is not a test, so there are no right or wrong answers. Just do your best. Now we will show you what the questions are like and how to use the buttons.”

Analyses

Psychometric modeling procedures were used to study construct dimensionality of the RAI, and the RAI item response process in relation to racial bias. Differential item functioning (DIF) analyses were completed to test if RAI items functioned equivalently across racial groups. The relation between the RAI and the PRAM was estimated with analysis of variance to test if the PRAM racial categories are related to the RAI score. Finally, we tested the hypotheses that the RAI item response options “both,” “neither,” and “I don’t know” would be used by children classified as biased by the PRAM II.

Item response modeling To model the RAI item response behavior, the value 0 was assigned if the child chose “Black” in response to a positive adjective or “White” to a negative adjective. Since each child had the option to indicate “neither,” “both,” or “I don’t know,” when responding to RAI items, the value 0 was interpreted as a pro-Black response. The value 2 was assigned if the child chose “White” in relation to a positive adjective and “Black” in relation to a negative adjective. The value 2 was interpreted as a pro-White response. When a “both” or “neither” response was provided,

a score of 1 was assigned. Evaluating these two response options equivalently is consistent with our theoretical hypothesis that these responses indicate a relatively “unbiased” attitude. Our modeling procedures test this hypothesis from a confirmatory rather than an exploratory perspective. The infrequently used “I don’t know” response was treated as missing and was not assigned a score. Measurement reliability was estimated with both classical and Rasch analyses. Using latent trait modeling, item response data was fitted to the Rasch unidimensional rating scale measurement model (Andrich, 1978). In addition to estimating item rating scale model parameters, we tested the assumption that the item response process was unidimensional.

Related to measurement dimensionality is the hypothesis that Rasch rating scale item parameter estimates obtained from Black and White children were equivalent after controlling for the level of child bias. Of course, Black and White children may differ in measured racial bias on the PRAM II (which they do; see Table 3), and, if so, we hypothesized the RAI would be sensitive to those differences. For children with the same level of bias, we expected the RAI item responses to be the same, regardless of race. If the item-level model parameters are different for Black and White children, after controlling for level of bias, then the measure does not function equivalently. Lack of item parameter equivalence may indicate multidimensionality.

Use of RAI response options One of our primary research questions focused on whether or not children used the response options “both,” “neither,” and “I don’t know.” We hypothesized that children who fell in the “neutral” PRAM II total score range would use these options, but our question was most related to children who displayed extreme PRAM II total scores (pro-Black/anti-White, pro-White/anti-Black). Unlike PRAM II binary options, when responding to the RAI, all children had the option to indicate “both,” “neither,” or “I don’t know,” and we hypothesized that a significant proportion of children would use them, and that the use of “both” and “neither” indicates a relatively neutral attitude.

Our null hypothesis was that those children classified as “biased” would *not* use the response options “both,” “neither,” and “I don’t know.” To test this hypothesis, we simply counted the number of “both,” “neither,” and “I don’t know” responses for each child across the 24 RAI items. Using rank order procedures (Conover & Iman, 1981), we tested statistically if the PRAM II classification groups used RAI responses at differential rates. Nonparametric analyses were used because the distributions of counts of response options used by racial bias classification tended to be highly skewed and cell sample sizes were unbalanced and small. Our procedures included tests of racial, gender, and grade effects on this relation.

Results

RAI item responses were scored 0 (pro-Black), 1 (“both,” “neither”), and 2 (pro-White). Estimates of the Rasch rating scale item parameters were obtained using Winsteps v3.74 (Linacre, 2012). For the 24 RAI items, Table 4 summarizes the Rasch rating scale model item scale location parameter, item Rasch rating scale fit statistics, and the item-measure correlations (items are sorted from high to low scale location).

Rasch rating scale model fit statistics indicate the rating scale model (see Fig. 5) fits the RAI data. The mean-square error model fit statistic has an expectation of 1, with acceptable values ranging between 0.50 and 1.5 (Linacre, 2012; Smith, 2010). For all items, the mean square fit statistic is within the range of acceptable model fit to the data. An important constraint of the rating scale model is that item category parameters remain constant across all items, with the scale location parameter being the only difference between items. Most noteworthy about the item response model shown in Fig. 5 was the use of the neutral response (i.e., “both” and “neither”). As bias increased from pro-Black to pro-White, there was an interval on the bias scale when the “both” and “neither” responses were used. This region of the bias scale is where the most probable response is a neutral response. This response applies to all RAI items.

RAI item location parameters indicate if the item is pro-Black (relatively low value) or pro-White (relatively high value) (see Figure 5). High value items required more pro-White bias for children to provide pro-White responses. For these items, it would not be unusual for a pro-White child to give a neutral or even a pro-Black response. Items with low location parameter values are interpreted as pro-Black items, for which only high pro-Black biased children would provide pro-Black responses. The item-measure score correlation was interpreted as the *item discrimination*; that is, how well the item differentiated respondents from low to high values. Ideally, item discrimination values should be at least 0.20 (Schmeiser & Welch, 2006). Only one item failed to meet this standard (see Table 4), and most items correlated 0.30 or higher with the total RAI score. For the total RAI 24-item scale, the estimated reliability coefficient alpha is 0.71, which is acceptable (Nunnally & Bernstein, 1994).

RAI dimensionality An important assumption of the Rasch model estimated for the RAI is the unidimensionality of the scale (Engelhard, 2013). First, examination of the residual correlation matrix indicated that, once the rating scale model was fitted to the data, the remaining item correlation matrix was essentially an identity matrix (on-diagonal values equal 1, off-diagonal values equal 0). The Bartlett test of sphericity and the Kaiser-Meyer-Olkin measure of sampling adequacy indicated the residual matrix was unsuitable for factor or principal component analysis (Chou & Wang, 2010).

Table 4 Rasch rating scale item location parameter estimates, model fit statistics, and item-by-measure correlations

*RAI Item (negative/positive)	Scale location	Model fit	Item by measure correlation
Who is selfish and doesn't share toys (neg)	0.53	0.99	0.38
Who got bad grade on hard test (neg)	0.52	1.30	0.24
Who washes hands and is clean (pos)	0.28	1.04	0.31
Who answered spelling question wrong (neg)	0.27	1.25	0.22
Who answered spelling question right (pos)	0.25	1.08	0.27
Who wouldn't help teacher (neg)	0.22	1.19	0.34
Who was naughty during recess (neg)	0.16	1.06	0.39
Who is helpful to teacher (pos)	0.12	1.09	0.18
Who shows new teachers where pencils are kept (pos)	0.06	0.83	0.41
Who shares toys w/other children (pos)	0.06	0.83	0.39
Who is friendly (pos)	0.03	0.76	0.40
Who helped the teacher clean up the classroom (pos)	0.01	0.88	0.32
Who is gentle with their pet (pos)	0.00	0.85	0.37
Who is naughty in the grocery store (neg)	-0.06	1.03	0.33
Who was mean and wouldn't play with new student (neg)	-0.07	1.12	0.33
Who is rough with their pets (neg)	-0.10	0.97	0.39
Who was nice and played with new student (pos)	-0.13	1.05	0.30
Who is helpful and cleans their room (pos)	-0.15	0.84	0.35
Whose story was bad (neg)	-0.25	1.03	0.34
Who is always told they are pretty (pos)	-0.28	0.85	0.35
Who is mean to other children (neg)	-0.30	1.08	0.33
Who did not help clean up (neg)	-0.32	1.13	0.29
Who doesn't have friends and is sad (neg)	-0.37	1.23	0.22
Who got reward for being quiet (pos)	-0.47	0.98	0.27

*Ordered by scale location

Using exploratory factor analysis, 1-, 2-, and 3-dimensional models were estimated and compared using information fit statistics for model comparison. Smaller values indicate relatively better fit (Dayton, 2003). Bayesian information criterion (BIC) fit statistics for the 1-, 2-, and 3-factor models were 6295.185, 6256.201, and 6252.678, respectively. Though the

BIC for the 3-factor model is smallest, these values are practically equal, suggesting that the hypothesis of the 1-factor model may be supported. One source of possible multidimensionality may be the positive or negative item frame. A confirmatory 2-factor model, testing for positive and negative factors, was fitted to the data and compared to a 1-factor model.

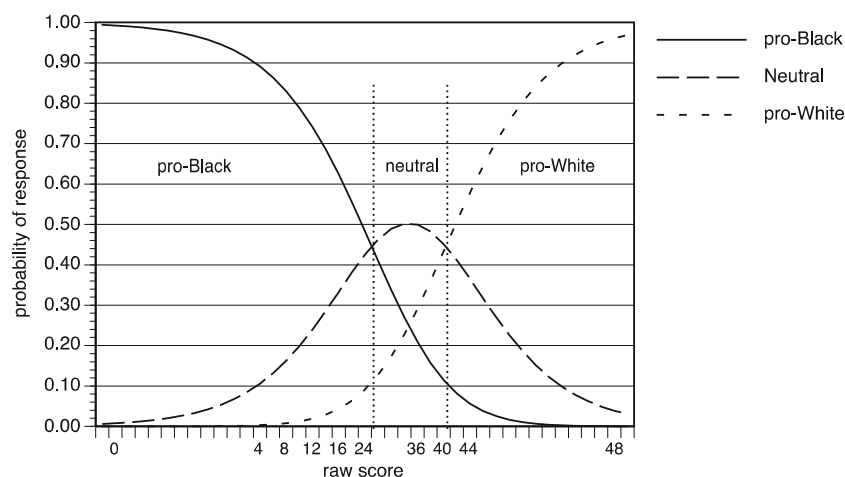


Fig. 5 Rasch rating scale model for RAI items scored 0 “pro-Black,” 1 “both or neither,” and 2 “pro-White”

The BIC fit value was 6296.703 and, when compared to the 1-factor model BIC (6295.185), the hypothesis of a 1-dimensional model was supported.

A source of measurement multidimensionality was also tested with measurement invariance modeling, specifically, differential item functioning (DIF) analyses. The assumption that RAI items function equivalently for racial identity groups is essential for group comparisons (Thissen, Rosa, & Mcleod, 2001). Rasch item parameters reported in Table 4 were assumed to be equivalent for Black and White children. Using Mantel-Haenszel DIF detection statistics (Millsap, 2011), the no-DIF model was supported for each of the 24 RAI items. RAI items do function equivalently for Black and White children in this sample. This is noteworthy because of the previous findings that called into question how White and Black children respond to extant measures of racial bias in different ways. We have demonstrated that the RAI does not reproduce asymmetric perceiver-group responses, and, instead, homes in on the pro-White, pro-Black, and neutral response concepts for all participants.

Relation between RAI and PRAM scores Though the PRAM II uses a forced-choice response format, we do hypothesize a positive relation between the PRAM II and RAI scores. The correlation between the scores was $r = 0.28$, which is lower than hypothesized. One explanation for this is the difference between the PRAM binary forced-choice response format and the RAI response format, which includes “both” and “neither.” Also, the skewed score distributions or lack of measurement reliability may be attenuating the correlation. Using analysis of variance, we compared the three PRAM II racial bias groups: (a) definite or probably pro-Black/anti-White (RAI: $M = 21.71$, $SD = 7.31$), (b) unbiased (RAI: $M = 22.94$, $SD = 6.26$), and (c) definite or probably pro-White/anti-Black (RAI: $M = 24.10$, $SD = 6.75$) with respect to their RAI total scores, which are statistically significant. A relevant and perhaps more important question pertains to how children in the three PRAM II bias categories respond to the RAI items, specifically use of the response options.

RAI response option use We hypothesized that many of the children classified by the PRAM II as biased would use the RAI categories of “both,” “neither” and “I don’t know.” To test this hypothesis, we placed participants into three groups based on the PRAM II total score classifications: definite/probable pro-Black, neutral, and definite/probable pro-White. Across the 12 positive and 12 negative RAI items, we counted the number of the 24 RAI items for which typical (pro-White/anti-Black) or atypical (pro-Black/anti-White) one-child or other-child responses were provided, the number of items for which “both” or “neither” responses were counted, and the number of for which “don’t know” responses were counted, also. Thus, there were five RAI dependent variables:

(a) number of *pro-Black/anti-White* response, (b) number of *pro-White/anti-Black* responses, (c) number of *both* responses, (d) the number of *neither* responses, and (e) the number of *don’t know* responses.

Table 5 provides frequencies of RAI response options used by PRAM II bias classification groups *conditional on race*. Within a racial category, differences between PRAM groups were tested statistically using nonparametric Kruskal-Wallis rank analyses. Generally, consistent with our hypotheses, for both Black and White children, RAI response options “both” (more frequently used) and “neither” were used by each of the three PRAM II racial bias groups. Noteworthy in Table 5 is one difference between Black and White children: for Black participants, the PRAM bias classification “Pro-White/Anti-Black” children used the RAI “pro-White/anti-Black” response options at a significantly higher rate than the other PRAM racial classes. Differences between PRAM bias groups were not obtained for the White participants (at $p < .05$). In this way, the RAI clearly identifies the known pattern that Black children can show (e.g., Baron & Banaji, 2004) in a manner that is immediately recoverable from the PRAM scores alone.

For each of the five RAI response options, we followed up with a two-way Kruskal-Wallis test of main and interaction effects (Akritas, Arnold, & Brunner, 1997), with race and PRAM II classification as the two factors. Tables 6 includes a summary of the two-way analyses. Consistent with the Kruskal-Wallis statistical tests and frequencies of RAI response use reported in Table 5, for the RAI Pro-White/Anti-Black response options, there is a difference in use between the PRAM II bias classification groups. Furthermore, the differences depend on race. The White children who are pro-White biased tend to use the “neither” response option more frequently than Black pro-White biased children. Again, this illustrates a nuance in responding that was not recoverable by the binary forced-choice format.

Analyses comparable to those used to test the effects of race and reported in Tables 5 and 6 were completed for gender and grade level. The results indicate that the relation between PRAM II bias classification and RAI item response is constant across gender and grade level. Children will use the response options “both” and “neither” irrespective of their gender or grade level. Only racial group effects were obtained. This is not surprising because the RAI items were controlled for sex, i.e., each item used either a pair of girls or a pair of boys. Again, with respect to the race, these results illustrate a nuance in responding that was not discernable by the binary forced choice format.

These results, collectively, can be related to the disproportionate rate of Black and White children classified as racially biased (Table 3). The significant PRAM II classification by race interaction reported in Table 6 and the values in Table 5 indicate the Black participants who are PRAM classified as

Table 5 RAI response option use by PRAM II bias classification by race

PRAM II Bias Classification					
Black					
	Pro-Black/ Anti-White (<i>n</i> = 34)	Neutral (<i>n</i> = 82)	Pro-White/Anti-Black (<i>n</i> = 51)	Kruskal Wallis (<i>df</i> = 2)	<i>p</i>
Pro-Black/Anti-White					
median	7	6	6	0.61	0.74
Range	0–18	0–16	0–16		
Pro-White/Anti-Black					
median	7	7	10	14.11	0.001
range	0–17	0–17	2–20		
Both					
median	5	5	3	3.96	0.14
range	0–24	0–20	0–17		
Neither					
median	1	1	0	5.45	0.07
range	0–11	0–11	0–8		
I Don't Know					
median	0	1	0	0.85	0.65
range	0–17	0–23	0–19		
White					
	Pro-Black/ Anti-White (<i>n</i> = 15)	Neutral (<i>n</i> = 48)	Pro-White/Anti-Black (<i>n</i> = 104)	Kruskal Wallis (<i>df</i> = 2)	<i>p</i>
Pro-Black/Anti-White					
median	5	7	5	2.75	0.25
range	0–18	0–13	0–15		
Pro-White/Anti-Black					
median	5	8	9	5.59	0.06
range	0–14	0–17	0–21		
Both					
median	5	4	5	0.54	0.76
range	0–14	0–17	0–23		
Neither					
median	2	1	1	0.35	0.84
min	0–10	0–11	0–20		
I Don't Know					
median	0	1	0.5	1.20	0.55
range	0–24	0–16	0–24		

pro-White biased use the RAI Pro-White/Anti-Black response options more frequently than White children who are pro-White biased. Also, Black children tended to use the “neither” response more frequently than White children.

Finally, because there is concern that the Oregon sample as relatively racially homogenous may bias these results, the analyses were replicated with the Georgia sample, only. The same statistical test results reported in Table 5 and Table 6 were obtained when the Oregon sample was excluded. Related to the possibility that the Oregon sample is not from the same population as the Georgia sample is drawn from, we focused an additional secondary analyses on comparison of the Oregon White and the Georgia White children. Most

noteworthy are the findings that the Georgia and Oregon White children do differ on the PRAM total score ($p < 0.01$) and PRAM classifications ($p < 0.01$). Specifically, Georgia White children are proportionately more Pro-White. On the RAI, there are no differences on the total raw score. On the RAI scale score, Georgia White children are more pro-White ($p < 0.05$). When counting the number of RAI response values, Georgia White children did provide more “typical” responses. The relatively small numbers of White children split across Georgia and Oregon (42 and 127, respectively) precluded invariance modeling. Assuming that the RAI functions invariantly across the two populations, with caution we can conclude that White children living in Georgia are

Table 6 RAI response option use by PRAM II bias classification and race

Source	<i>df</i>	Chi-square	<i>p</i>
<i>RAI Pro-Black/Anti-White Response</i>			
Race	1	2.65	0.104
PRAM II	2	0.40	0.820
Race by PRAM II	2	2.69	0.261
<i>RAI Pro-White/Anti-Black Response</i>			
Race	1	0.711	0.400
PRAM II	2	15.90	0.001
Race by PRAM II	2	5.80	0.033
<i>RAI Both Response</i>			
Race	1	0.27	0.602
PRAM II	2	2.42	0.298
Race by PRAM II	2	2.68	0.262
<i>RAI Neither Response</i>			
Race	1	10.30	0.001
PRAM II	2	3.13	0.21
Race by PRAM II	2	2.85	0.24
<i>RAI I Don't Know Response</i>			
Race	1	0.03	0.85
PRAM II	2	1.98	0.37
Race by PRAM II	2	0.09	0.96

relatively more pro-White/anti-Black than White children living in Oregon. While interpreting our data to make these comparisons was not the study purpose, the results are very interesting.

Discussion

The research reported here provides evidence that the RAI is a technically adequate measure of child racial bias and that item-level response behavior is systematic. Scoring of item responses as “pro-Black,” “neutral,” or “pro-White” fit the Rasch rating scale model. Psychometrically, it is important to demonstrate that the RAI functions invariantly across the population of Black and White children, though Black children do score lower (more Pro-Black) than White children. The DIF modeling does support the hypothesis of equivalent measurement. Alternative procedures for testing measurement dimensionality tended to support the hypothesis that the RAI is a unidimensional measure. These results render interpretable the frequency of RAI response option use and, specifically, which children tended to use them.

Most notably, we found no differences between the PRAM II bias categories in frequency of response option use, suggesting that racial bias classifications may depend more on the

measurement procedure. We tested the relation between the PRAM and RAI, focusing most closely on how those children classified as biased on the PRAM II used the expanded RAI response options. A reasonable hypothesis is that children who are pro-Black/anti-White or pro-White/anti-Black would respond to RAI items using response options other than “both,” “neither,” or “I don’t know,” but this was not supported in these data. However, the basis for including these response options was more related to the following hypothesis: if children are actually provided with the options of “both,” “neither,” or “I don’t know,” then they will in fact use them. The results clearly indicated that the response options “both,” “neither,” and “I don’t know” were used, and children classified as racially biased on the PRAM II used them. Our findings suggest that the simple forced-choice format may misidentify racial bias. This finding has important implications for measurement of childhood racial bias and the use of simple forced-choice response formats.

The findings from study three provide a starting point to think through more sophisticated measurement methodologies that are sensitive to the relevant measurement issues outlined in the literature (Cameron, Alvarez, Ruble, & Fuligni, 2001; Killen & McKown, 2005; Kowalski, 2003): limited response options, race of examiner effects and social desirability (Krumpal, 2013), and single-perceiver and single-target instruments. In addition, it would be optimal to integrate the existing knowledge base on measurement into one instrument (Clark & Tate, 2008). First, while there have been efforts to add additional response options to explicit measures of children’s racial attitudes, the production of different scales that use different response choices and scoring procedures makes it difficult to compare scales, which hinders the theoretical advancement of intergroup relations in childhood (Killen & McKown, 2005). We believe that the RAI response options have the potential to provide more nuanced understanding of children’s racial prejudice and advance theorizations on the locus of children’s racial attitudes. In the RAI five-option response space, choosing one child or the other child indicates that a respondent has a clear preference. The “both” option can be interpreted as a statement of certainty (I have enough information, and the mini-story applies to both). The “neither” option is understood as an affirmative statement that the mini-story applies to neither, suggesting no clear preference. The “I don’t know” option indicates that the respondent does not have enough information. While some researchers might oppose the inclusion of an “opt out” response, such as “I don’t know” –and, thus, forcing a child to choose a response even if she truly does not know – including “I don’t know” improves the quality of the data and allows researchers to interpret the other four response options as statements of “relative certainty” because the respondent has the option of stating she does not know (Clark & Tate, 2008). This type of response set renders children’s responses as more clearly

interpretable and has the potential to expand the theoretical understanding of, for example, minority children's lack of ingroup preference as outlined by the system justification theory, and it would lend precision to existing constructs by providing much clearer and interpretable response options. Of course, these patterns need to be replicated in other samples, but this paper has demonstrated an important proof of concept: that children can respond to more than two forced-choice options and that the pattern of their responses in the larger format reveals subtleties in how researchers can assess racial bias.

A great deal of social science research on young children's racial attitudes has focused on Black-White dynamics, with few exceptions (Gibson et al., 2015; Griffiths & Nesdale, 2006; Killen, Henning, Kelly, Crystal, & Ruck, 2007). As a result, much of what we know about intergroup attitudes among the young, particularly in the US context, is based upon majority-minority relations (e.g., Whites' attitudes toward Blacks, Blacks' attitudes toward Whites) with current gaps in our understanding, in particular, of Latina/o children's attitudes (Stokes-Guinan, 2011). Importantly, demographic shifts necessitate the construction of instruments that reflect the multiracial milieu and can capture intricacies of intergroup dynamics, starting in childhood (Bobo & Hutchings, 2003). Our solution to this issue is not to merely include more research on specific racial and ethnic groups in cross-sectional designs using independent rating techniques that make it difficult to compare measures (Killen & McKown, 2005), rather it is to study racial and ethnic groups as both targets and perceivers (Clark & Tate, 2008), in one instrument that is easy and relatively inexpensive to administer. The RAI could be repurposed to include multiple-perceivers and multiple-targets (e.g., Asian, Latino, Black, and White) in *one* instrument and would allow for comparisons to be calibrated to the racial identity of the perceiver (Clark & Tate, 2008). It could also specify whether outgroup derogation, for example, is specific to a particular racial outgroup or generalizes to all racial outgroups. This type of response architecture could also generate additional theoretical constructs that single target-single perceiver instruments cannot (see Clark & Tate, 2008). Currently, there are few measurement approaches that can render such multi-group attitudinal comparisons (Clark & Tate, 2008).

To address concerns about social desirability bias based on the race-of-examiner effect, the use of A-CASI methodology can minimize social desirability by allowing children to answer questions on their own, without the help of an adult administrator to make their answer selections, which is the standard way that the PRAM II is administered. Moreover, A-CASI standardizes the administration process (Borgers et al., 2000) and provides a low cost strategy for administering explicit scales to large numbers of respondents.

On a practical level, the measurement of children's racial attitudes using multiple-response options, multiple perceivers,

and multiple targets is the type of measurement approach that has implications for the development of educational interventions designed to reduce racial bias. Interventions could be developed based on theoretical insights regarding the precise contours of children's racial attitudes, and the efficacy of anti-bias interventions could be evaluated with this type of instrument.

Limitations and future research While we demonstrate the conceptual argument for using RAI response categories, a main limitation related to the three ordered category scoring system is the interpretation of the "both" response. For example, the scoring system assumes that choosing "both" in relation to positive items means the same thing as choosing "both" in relation to negative items. It is possible that neither group is "favored" in the comparative sense. Assigning a score of 1 to the responses of "both" and "neither," irrespective of item valence, may be problematic since stating that both children are positive is likely not equivalent to stating that neither child is positive. The modeling of "both" and "neither" as neutral did indicate that children who use these response options appear to not be, pro-Black or pro-White, at least as measured by the PRAM II. Clearly, this may be a measurement artifact. More research is necessary to better understand the process underlying the use of "both" and "neither" and the extent to which these responses are similar. Conducting cognitive interviews with young children on the both and neither responses would be a good starting point. Although we provided instructions on the meaning of the RAI response options in the instructional training segment, it is unclear if children understood the difference in saying "neither" child is good or "both" children. In addition, future research should explore alternative scoring techniques to better capture the meaning of "both" and "neither" response options by item valence.

Another limitation of this study is related to race-of-examiner effects and social desirability. Although A-CASI technology implicitly controls for race-of-examiner effects to the extent that the interviewer is removed from the administration of the instruments, we did not directly test this effect. In addition, the racial composition of the research team was mixed. The first author who was also the research team Principal Investigator is a light-skinned Black woman, two other research members are White, and one research member is a dark-skinned Black woman, and the composition of the research team varied over the course of the study. Each research participant wore headphones so that the questions were delivered privately, and we instructed research staff in the computer rooms to be as unobtrusive as possible. Children were instructed to raise their hand if they experienced any issues with the computer. As such, we did everything possible to minimize race-of-examiner effects and

social desirability issues. Finally, the findings with this population of Black and White children from the Pacific Northwest and the southeast may not be generalizable to other White and Black populations from different regions or other socioeconomic groups.

Conclusion

In an increasingly multiracial and globally interconnected society, empirical and interdisciplinary study of racial attitudes is critical. Societies with problematic racial and ethnic relations clearly exhibit negative economic and health sequelae (Dulin-Keita, Hannon, Fernandez, & Cockerham, 2011). Racial and ethnic attitudes develop during childhood; however, it is within this period that racialized ways of thinking are most malleable. Without clear measurement of children's racial biases, school curricula, race-targeted interventions and policies are likely to be ineffective.

Acknowledgments The development of the RAI and data collection efforts was supported by grants from: National Science Foundation (Dissertation Improvement Grant- Methodology, Measurement, and Statistics #061886) and the American Sociological Association Fund for the Advancement of the Discipline to Khaya D. Clark (with Dr. Tyrone Forman). The National Science Foundation supported this research in the form of a postdoctoral fellowship to Khaya D. Clark at Emory University (#0805729). Special thanks to Drs. George Engelhard (Emory University) and Tyrone Forman (University of Illinois-Chicago) for the contributions to the development of this research study. The authors are grateful for editing and publication assistance from Ms. LeNeve Spires, Publications Manager, Department of Family Medicine, Oregon Health & Science University, Portland, OR, USA.

Author contributions K.D. Clark and C.U. Tate developed the study concept. K.D. Clark developed the study design and received the funding as Principal Investigator. K.D. Clark and a small team (Liz Alexander, Maggie Haggerman, and Adria Welcher, all from Emory University) collected data in Atlanta Public schools. K.D. Clark and P. Yovanoff performed data analysis. P. Yovanoff and K.D. Clark led the interpretation of the data analysis; C.U. Tate provided comments on the data analysis. P. Yovanoff and C.U. Tate provided revisions to the manuscript. All authors approved the final version of the manuscript for submission.

References

- Aboud, F. (1988). *Children and prejudice*. New York: Basil Blackwell, Ltd.
- Aboud, F. (2003). The formation of in-group favoritism and out-group prejudice in young children: Are they distinct attitudes? *Developmental Psychology*, 28, 48–60.
- Aboud, F., & Doyle, A. B. (1996). Does talk of race foster prejudice or tolerance in children? *Canadian Journal of Behavioural Science*, 28, 161–170.
- Akritas, M., Arnold, S. F., & Brunner, E. (1997). Nonparametric hypotheses and rank statistics for unbalanced factorial designs. *Journal of the American Statistical Association*, 92, 258–265.
- Andrich, D. A. (1978). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement*, 2, 581–594.
- Babbie, E. (1995). *The practice of social research*. Belmont: Wadsworth Publishing Co.
- Baron, A., & Banaji, M. (2004). *The stability and change of implicit and explicit prejudice across development*. Paper presented at the 34th Annual Meeting of the Jean Piaget Society, Toronto, ON.
- Baron, A., & Banaji, M. (2009). Evidence of system justification in young children. *Social and Personality Psychology Compass*, 3, 1–9.
- Bernat, G., & Balch, P. (1979). The Chicano Racial Attitude Measure (CRAM): Results from an initial investigation. *American Journal of Community Psychology*, 7(2), 137–146.
- Bernstein, J., Zimmerman, T., Werner-Wilson, R. J., & Vosburg, J. (2000). Preschool children's classification skills and multicultural education intervention to promote acceptance of ethnic diversity. *Journal of Research in Childhood Education*, 14(181–192).
- Best, D. D. (2005). [Conversation about computerized administration of the PRAM II and CMT II].
- Bigler, R. S., & Liben, L. S. (1993). A cognitive-developmental approach to racial stereotyping and reconstructive memory in euro-American children. *Child Development*, 64, 1507–1518.
- Blumer, H. (1958). Race prejudice as a sense of group position. *Pacific Sociological Review*, 1, 3–7.
- Bobo, L., & Hutchings, V. (2003). Perceptions of racial group competition: Extending Blumer's theory of group position to a multicultural social context. *American Sociological Review*, 61, 951–972.
- Borgers, N., de Leeuw, E., & Hox, J. (2000). Children as respondents in survey research: Cognitive development and response quality. *Bulletin de Methodologie*, 66, 60–75.
- Cameron, J. A., Alvarez, J. M., Ruble, D. N., & Fuligni, A. J. (2001). Children's lay theories about ingroups and outgroups: Reconceptualizing research on prejudice. *Personality and Social Psychology Review*, 5, 118–128.
- Chou, Y. T., & Wang, W. C. (2010). Checking dimensionality in item response models with principal component analysis on standardized residuals. *Educational and Psychological Measurement*, 70, 717–731.
- Clark, K., & Tate, C. (2008). Measuring Racial Prejudice in a multiracial world: New methods and new constructs. In M. Morrison & T. Morrison (Eds.), *The psychology of modern prejudice*. Hauppauge: Nova Science Publishers.
- Conover, W., & Iman, R. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician*, 35, 124–129.
- Cross, W., & Cross, T. (2008). Theory, research, and models. In S. Quintana & C. McKown (Eds.), *Handbook of race, racism and the developing child*. Hoboken, NJ: John Wiley & Sons, Inc.
- Dayton, C. M. (2003). Model comparisons using information measures. *Journal of Modern Applied Statistical Methods*, 2, 281–292.
- de Leeuw, E., Hox, J., & Kef, S. (2003). Computer-assisted self-interviewing tailored for special populations and topics. *Field Methods*, 15, 223–251.
- de Leeuw, E., & Nicholls, W. (1996). Technological innovations in data collection: Acceptance, data, quality and costs. *Sociological Research Online*, 1(4), 1–19.
- Doyle, A. B., & Aboud, F. (1995). A longitudinal study of white children's racial prejudice as a social-cognitive development. *Merrill-Palmer Quarterly*, 41, 209–228.
- Dulin-Keita, A., Hannon, L., Fernandez, J., & Cockerham, W. (2011). The defining moment: Children's conceptualization of race and experiences with racial discrimination. *Ethnic and Racial Studies*, 34(4), 622–682.
- Dunham, Y., Baron, A. S., Banaji, M. R. (2007). Children and social groups: A developmental analysis of implicit consistency in Hispanic Americans. *Self & Identity*, 6, 238–255.

- Engelhard, G. (2013). *Invariant measurement*. New York: Routledge.
- Gibson, B., Robbins, E., & Rochat, P. (2015). White bias in 3- to 7-year-old children across cultures. *Journal of Cognition and Culture*, *15*, 344–373.
- Glover, R. J., & Smith, C. A. (1997). Racial attitudes of preschoolers: Age, race of examiner, and childcare setting. *Psychological Reports*, *81*, 719–722.
- Griffiths, J. A., & Nesdale, D. (2006). Ingroup and outgroup of ethnic majority and minority children. *International Journal of Intercultural Relations*, *30*, 735–749.
- Jones, E. F., Parker, B. L., Joyner, M. H., & Steiner-Ulku, B. (1999). The influences of behavior valence and actor race on black and white children's moral and liking judgements. *Journal of Psychology*, *33*, 194–204.
- Jost, J., Banaji, M., & Nosek, B. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, *25*(881–919).
- Katz, P. A., & Kofkin, J. A. (1997). Race, gender and young children. In S. S. Luthar, J. A. Burack, D. Cicchetti, & J. R. Weisz (Eds.), *Developmental psychology: Perspectives on adjustment, risk, and disorder* (pp. 51–74). New York: Cambridge University Press.
- Katz, P. A., & Zalk, S. (1976). *The Katz-Zalk projective prejudice test: A measure of racial attitudes in children*. Retrieved from Washington D.C.
- Killen, M., Henning, A., Kelly, M. C., Crystal, D., & Ruck, M. (2007). Evaluations of interracial peer encounters by majority and minority US children and adolescents. *International Journal of Behavioral Development*, *31*, 491–500.
- Killen, M., & McKown, C. (2005). How integrative approaches to intergroup attitudes advance the field. *Applied Developmental Psychology*, *26*, 616–622.
- Kowalski, K. (2003). The emergence of ethnic and racial attitudes in preschool-aged children. *The Journal of Social Psychology*, *143*(6), 677–690.
- Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality and Quantity*, *47*, 2025–2047.
- Krysan, M. (1998). Privacy and expression of white racial attitudes—a comparison across three contexts. *Public Opinion Quarterly*, *62*, 506–544.
- Levy, S., & Hughes, J. (2009). Development of racial and ethnic prejudice among children. In T. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 23–42). New York: Psychology Press.
- Linacre J. G. (2012). Winsteps (3.75) [Computer software]. Beaverton, OR
- McGlothlin, H., & Killen, M. (2006). Intergroup attitudes of European American children in homogenous schools. *Child Development*, *77*(5), 1375–1386.
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York, New York: Routledge.
- Nesdale, D. (2000). Developmental changes in children's ethnic preferences and social cognitions. *Journal of Applied Developmental Psychology*, *20*, 501–519.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill.
- Rice, A., Ruiz, R. E., & Padilla, A. M. (1974). Person perception, self-identity and ethnic group preference in Anglo, Black, and Chicano preschool and third-grade children. *Journal of Cross Cultural Psychology*, *5*(1), 100–108.
- Schmeiser, C. B., & Welch, C. J. (2006). Test development. In R. L. Brennan (Ed.), *Educational measurement* (pp. 307–353). Westport, CT: Praeger Publishers.
- Smith R. (2000). Fit analysis in latent trait measurement model. *Journal of Applied Measurement*, *1*, 199–218.
- Smith, E. R., & Mackie, D. M. (2002). Commentary. In D. M. Mackie & E. R. Smith (Eds.), *From prejudice to intergroup emotions: Differentiated reactions to social groups* (pp. 285–299). New York: Psychology Press.
- Stokes-Guinan, K. (2011). Age and skin tone as predictors of positive and negative racial attitudes in Hispanic children. *Hispanic Journal of Behavioral Sciences*, *33*(1), 3–21.
- Thissen, D., Rosa, K., & Mcleod, L. D. (2001). Item response theory for items scored in more than two categories. In D. Thissen & H. Wainer (Eds.), *Test scoring*. Hillsdale, NJ: Lawrence Erlbaum.
- Williams, J. E., & Morland, J. K. (1975). *Preschool racial attitudes measure II and color meaning test: General information and manuals of directions*. Retrieved from Wake Forest.
- Williams, J. E., & Morland, J. K. (1976). *Race, color and the young child*. Chapel Hill: University of North Carolina Press.
- Williams, J. E., Best, D. L., Boswell, D. A. (1975). Children's racial attitudes in the early school years. *Child Development*, *46*, 494–500.
- Williams, J. E., Best, D. L., Boswell, D. A., Mattson, L. A., Graves, D.J. (1975). Preschool racial attitudes measure II. *Educational and Psychological Measurement*, *35*, 3–18.