

Quantitative measures of respect and social inclusion in children: overview and recommendations

Colin G. Tredoux^{a*}, Noraini M. Noor^b and Lisa de Paulo^a

^a*Department of Psychology, University of Cape Town, Cape Town, South Africa;*

^b*Department of Psychology, International Islamic University of Malaysia, Kuala Lumpur, Malaysia*

(Received 4 September 2009; final version received 30 October 2009)

An important task in empirical research on prejudice, respect and inclusion in children is the measurement of attitudes and behaviours. Although the literature on measuring attitudes and behaviour in adults is of some assistance, it is also clear that the task of measuring children's attitudes and behaviour presents particular problems and challenges. We provide an overview in this paper of a number of currently used measures. Their strengths and limitations are outlined, and recommendations are made for contemporary practice. This review has been undertaken by the Una Quantitative Methods Learning Group. Una is a global learning initiative on children and ethnic diversity (see: www.unaglobal.org).

Keywords: children; ethnic diversity; attitudes; prejudice; respect; social inclusion

Introduction

Social scientists have long been interested in the development of prejudice in young children. This is part of a larger tradition in social science that addresses the global phenomenon of ethnic conflict. However, an important goal is to extend beyond this focus on prejudice and also consider the promotion of respect and social inclusion in children and adults. To this end, social scientists have created measures to monitor levels of respect and inclusion, and to evaluate changes as a result of (i) naturally occurring development, and (ii) programmes that aim to improve group relations.

We use the following general definitions of 'prejudice', which we will consider to be the opposite pole of liking and respect (an attitude), and of 'discrimination', which is the opposite pole of inclusion (a behaviour). In the latter case, the definition assumes that the social categories of race and ethnicity are the group characteristic at issue, but we believe that the definition is useful for other group characteristics.

Prejudice is a unified, stable and consistent tendency to respond in a negative way towards members of a particular ethnic group because of their group affiliation. (Aboud, 1988, p. 6)

Prejudice involves holding derogatory attitudes or beliefs, expressing negative affect, or displaying hostile or discriminatory behavior towards members of a group on account of their membership of that group. (Brown, 1995, p. 8)

*Corresponding author. Email: colin.tredoux@uct.ac.za

A social science definition of racial discrimination... includes two components: (1) differential treatment on the basis of race that disadvantages a racial group and (2) treatment on the basis of inadequately justified factors other than race that disadvantages a racial group (differential effect). (Blank, Dadaby, & Citro, 2004, p. 39)

An important orienting observation is that there are particular practical difficulties that researchers must engage and deal with as a consequence of assessing children. Our goal, ultimately, is to recommend a set of reliable and valid measuring instruments for researchers in the field, and criteria for selecting ones appropriate for the purpose.

The typical formats in the child attitude literature have been to use evaluation or liking judgments, about pictorially represented targets, and administered in a one-on-one interview. In these tasks, special efforts are made to represent race or ethnicity as clearly and simply as possible, for example with photographs or drawings of comparably attractive individual children from different ethnic groups, in the hope that if this is done carefully enough, the child's response can be generalised to the category as a whole. In contrast, behavioural observation focuses on children's responses to a person in context, and so there can be no assumption about its generalisability.

The particular problems that measuring prejudice in children poses has been addressed in a number of ways, over a period of some 80 years. Attempts have ranged from the elicitation of direct ethnic preference, stereotype and trait evaluation, structured and unstructured interviews, observations of behaviour in context, friendship choice and sociogram mapping, among others. Some of the main controversies over the years have involved:

- (1) Whether children should be asked to declare their attitudes or be asked to evaluate race targets. Most researchers now ask children to evaluate or indicate a desired social distance from targets, and do not attempt to infer their mental states.
- (2) Whether children should directly compare two or three targets, or evaluate each separately. Most now allow children to preview all targets and then evaluate one at a time, providing response options that include saying the target does or does not have the attribute.
- (3) Whether only positive approach items should be included, or whether both positive and negative items should be included. Most now include both positive and negative items, because item analysis has shown the negative ones to be particularly discriminating.
- (4) Whether to include stereotypic traits or age-appropriate evaluative terms. This depends on the context, in that some young children do not know the stereotype, yet apply age-appropriate evaluations.
- (5) Whether behavioural intentions constitute a measure of behaviour or of attitude. Most studies have found that behavioural intentions are not translated into behaviour (e.g., negative intentions do not lead to behaviour but positive intentions are uncorrelated with behaviour). Consequently, they are akin to attitude measures of social distance.
- (6) Whether only two racial and ethnic groups should be included, or more than two. This of course depends on the context and the purpose. Less prejudice is found when three groups are evaluated, compared to two, yet more than three can confuse and distract young children.

There are additional controversies surrounding the selection and measurement of determinants or correlates of prejudice and discrimination. One concerns how to define and measure self-categorisation and self-identity in children (cf. Bennett & Sani, 2008). Another concerns the role of common developmental trajectories versus different social influences. A third focuses on the role of the context, namely the family, neighbourhood, school and the culture, such as individualist and collectivist cultures. These issues are important but beyond the scope of the present paper.

We consider measures and methods used to assess prejudice in children, under several broad headings, which reflect the nature or intention of the measures. First, we consider those attitude measures based on respectful–disrespectful evaluations of particular ethnic or race groups. Second, we consider measures that assess attitudes in terms of personal relations that children have with members of particular groups. Liking, friendship, peer relations and social distance are the primary forms of relating assessed in the extant literature, though some have also assessed the quality of the relationship. Third, we consider behavioural measures of inclusion and exclusion derived for the most part from observations of children in context. These include the degree of voluntary association, spatial proximity of children in free interaction or encounter, and specific inclusive behaviours such as helping and negotiating. Only currently used measures will be described. It should be noted that these build on earlier attempts (including the Clark & Clark Doll Choice Technique, 1947; the Katz-Zalk Projective Prejudice Test – Katz, Sohn, & Zalk, 1975) which have since been rejected for a number of reasons (see Aboud, 1988; Katz & Zalk, 1974).

Attitudes: evaluations

The PRAM II

The PRAM II shows the child 24 racial items and 12 filler gender items (Williams, Best, & Boswell, 1975). The 24 racial items require children to select between white-skinned and black-skinned figures in response to evaluative adjectives. One item reads: ‘Here are two girls. One of them is an ugly girl. People do not like to look at her. Which is the ugly girl?’ Another item is: ‘Here are two boys. One of them is a kind boy. Once he saw a kitten fall into a lake and picked up the kitten to save it from drowning. Which is the kind boy?’ (see Figure 1, described in Aboud, 1988).

Children are presented with a picture of a black and a white figure and have to choose which one best fits the description. The strength of the child’s positive or negative attitude is ascertained by adding the number of pro-white and anti-black choices made, or the reverse. The scoring ranges from 0 to 24. Low scores are indicative of pro-black/anti-white bias and high scores are indicative of pro-white/anti-black bias. Mid-range scores between 5 and 8 show lack of bias (Aboud, 1988).

Many studies measuring prejudice in children have used traits from the PRAM II. Empirical evaluation of the measurement properties of the PRAM has suggested that the instrument may have adequate internal consistency but questionable test–retest reliability: although Williams and colleagues reported Spearman-Brown adjusted coefficients of between .78 and .88, Nagata (2001) reports a test-retest reliability of .55, which is low.

Aboud (1988) noted that the PRAM II assesses children on many evaluative attributes, in numerous contexts, and results generalise to contexts beyond that of the test itself. However, there are certain weaknesses. The most problematic is that the

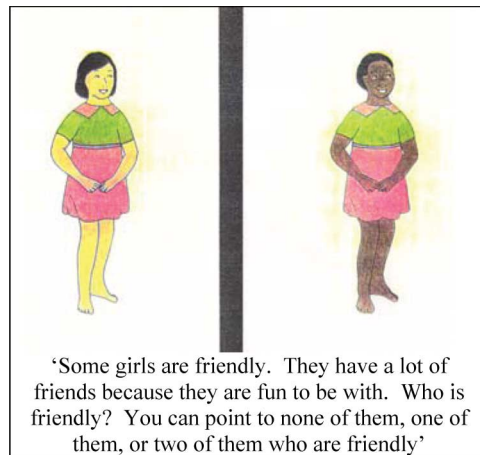


Figure 1. Example item from the multiPRAM (from Aboud, 2002).

PRAM II presents targets concurrently and requires children to make a forced choice between them. As children have to select one of the targets, it implies rejection of the other when in fact the children may not be expressing any attitude to the one not chosen. More importantly, children are not permitted to assign the same trait to more than one target. In addition, the evaluative terms used in this test are simple; they are appropriate for children up to seven or eight years but most are not appropriate for older children. On the positive side, the test is acceptable to young children, easy to administer and score, and findings tend to correlate with a number of interesting predictors. However, at this point the forced-choice format limits its usefulness and it may be considered to measure ingroup favoritism rather than outgroup prejudice (Aboud, 2003) – choosing one's own group does imply ingroup favoritism, but it does not imply outgroup prejudice, and the forced-choice procedure removes the ability to assess this.

A multiresponse version of the Preschool Racial Attitude Measure (multiPRAM) was used in a study conducted by Aboud (2002) to combine the multiple gender and age targets of the PRAM with the multi-response options of the MRA (discussed next). The multiPRAM was used to determine children's own attitudes and assumptions about the attitudes of white and black communicators both before and after hearing an anti-bias message. The multiPRAM asks children to assign six positive and six negative attributes and four filler items to pictures of white- and brown-skinned children.

Children are asked to answer by pointing to none, one or two pictured stimulus individuals whom they think have the attributes in question. The stimulus drawings are presented in pairs, and black and white people, males and females of different ages, represented in the drawings are alike, except for skin color and hair texture. An outgroup score is formed by subtracting the number of negative evaluations ascribed to the outgroup people from the number of positive evaluations ascribed to the outgroup (positive minus negative). Likewise a score for ingroup evaluation is calculated. Test-retest reliabilities over two weeks were .43. The major limitation with this measure is that it assesses attitudes to males and females of different ages, including children, adolescents and adults. It requires the use of an artist's coloured drawings which have to be requested from the authors or redrawn with the correct features varied.

Wright and Tropp (2005)

Wright and Tropp (2005) developed an instrument that has considerable similarities to the PRAM. Their instrument was created by pre-testing several hundred head-and-shoulder Polaroid photographs of five- to seven-year-old Latino and white children. Three Latino and three white adult raters rated photographs on the following three dimensions, using five-point scales: (a) the clarity of the photograph; (b) the positivity of the child's facial expression; and (c) the child's physical attractiveness. After collecting the ratings, Latino/white pairs were selected according to their match on these dimensions. Four (two male and two female) Latino/white matched pairs were combined to create a set of eight photographs. When the instrument is used, children are asked to select one of the photographs in each matched pair, in response to six evaluative items, namely those: (1) 'who are good at lots of things'; (2) 'who are smart'; (3) 'who have lots of friends'; (4) 'who are nice'; (5) 'who are happy'; and (6) 'who like school'. The number of times that Latino and white targets are chosen in these trials is counted, creating separate ratings of ingroup evaluation and outgroup evaluation, which are summed across pairs, and the six evaluative items, to produce overall measures of Ingroup Evaluation and Outgroup Evaluation. Internal reliability coefficients for these measures are good (Cronbach's $\alpha > .8$).

The MRA measure

Whereas the previously described measures assess attitudes to a single outgroup, the Multi-Response Racial Attitude (MRA) measure assesses attitudes to multiple groups. In Doyle and Aboud's 1995 study this was in relation to white, black and Chinese Canadian groups, but in principle other groups could be evaluated with the MRA (see Figure 2).

Following a similar logic to the PRAM, and its variants, the MRA requires children to assign 20 evaluative attributes to multiple groups. Ten positive attributes (e.g., plays fair, is friendly) and 10 negative attributes (e.g., is bossy, excludes others), were taken from the original Preschool Racial Attitude Measure, when used with children

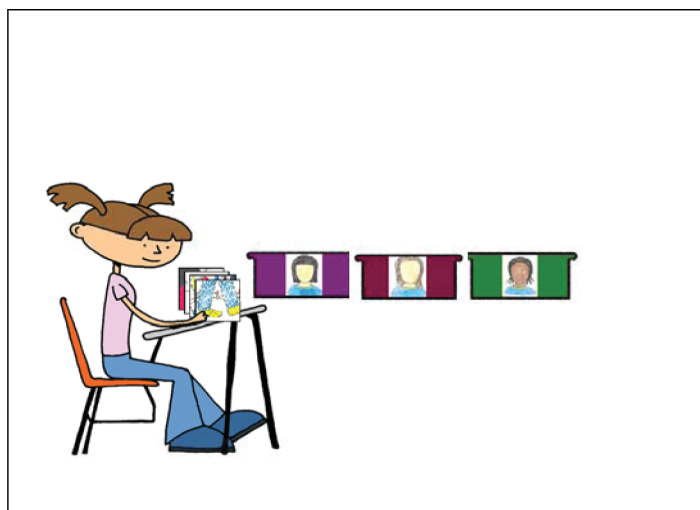


Figure 2. Three-box version of the MRA (from Doyle & Aboud, 1995).

under seven years, and from the Revised Class Play measure when used with older children. Attributes are either presented in picture form for younger children, or written on index cards for older children. Each attribute is presented on as many cards as there are groups, making a total number of cards equal to 20 times the number of groups. Children are asked to sort the cards by placing them in boxes (as many as there are groups). One box is indicated to belong to a child of the same group as the respondent, and the other two to children from the other groups. A drawing of a child of the same sex (and same, or different group) as the respondent is attached to the front of each box. The drawings vary only in skin colour and hair texture, with facial features minimised. For each evaluation, children are given identical cards and instructed to insert the cards in the box or boxes of people who are that way. For example, 'Some children are bossy. Who is bossy? Is it the white child, the black child, the Chinese child, or more than one of them who is bossy?' (Aboud, 2003; Doyle & Aboud, 1995). The number of evaluative attributes assigned to each group is tallied, treating positive and negative attributes as having values +1 and -1, respectively. If required, a composite intergroup bias score may be calculated as ingroup minus outgroup.

The MRA was designed to measure bias through an assessment of both positive and negative attitudes, and was also intended to partially decouple ingroup and outgroup attitudes: children can allot an evaluation to more than a single group and are not forced to evaluate one as positive and the other as negative, as happened with the earlier, forced-choice instruments. It has high internal consistency with alpha coefficients in one study being .90 for positive white, .89 for negative white, .83 for positive black and .79 for negative black. Test-retest reliabilities across a two-week interval for an independent sample of children from four to seven years were: $r(22) = .48$ for positive white, .66 for negative white, .68 for positive black and .71 for negative black (cited in Aboud, 2003, p. 49). Among white children, positive white reliabilities are often lower because the range of scores is narrower. Criterion validity has been demonstrated in relation to how children talk with their friend about racial attitudes (Aboud & Doyle, 1996) and to their actual friend selections (Aboud, Mendelson, & Purdy, 2003).

The Intergroup Attitude Measure

Cameron, Rutland and Brown (2007) created a modified version of the MRA called the Intergroup Attitude Measure to allow for more independent evaluations of groups. Instead of having children evaluate the groups simultaneously, they present children with the two groups as a preview (so they would mentally consider a comparison) and then asked for evaluations one group at a time. The stimulus groups are presented as a collage of people from each ethnic group. First the two collages are shown and labeled, and then one is removed while the other is evaluated. A second difference is that the response format is similar to that used by Bigler, Jones and Lobliner (1997) in that children judge the frequency or prevalence of the attribute in the group: none, some, most or all. Four squares of stick figures depict the quantities (see Figure 3). While looking at the collage, children are asked, 'How many people are that way? Point to the square of stick figures to indicate whether none, some, most or all are that way'. Once again, there are 10 positive (clean, happy, friendly, good, hardworking, helpful, kind, nice, unselfish and polite) and 10 negative attributes (bad, dirty, nasty, unhelpful, sad, selfish, rude, unkind, lazy and unfriendly). These attributes are appropriate for children under eight years but perhaps not for older children who may resist

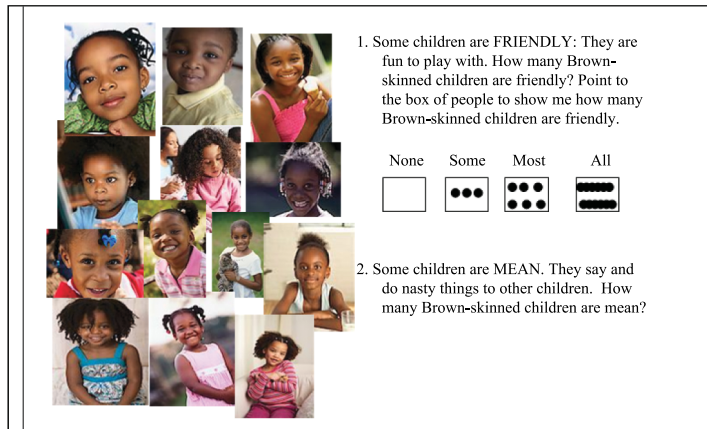


Figure 3. MRA version in current use.

evaluating people as 'dirty' or 'lazy'. Mean scores are calculated on a one to four scale for positive items and negative items for each group; an intergroup bias score may also be calculated. This measure includes important age-appropriate and measurement features: comparative but separate evaluations of groups, a collage of same-age stimulus persons who show the natural variation among group members, and a four-point attitude response. Alpha coefficients in one study were above .80. However, indicators of validity are not yet available.

Implicit attitudes

An important new development in adult and child research on attitudes concerns so-called 'implicit attitudes'. Most attitude measures require self-report and therefore rely on the frankness of the respondent. Implicit attitudes are less subject to conscious control by respondents. They are 'actions or judgments that are under the control of automatically activated evaluation, without the performer's awareness of that causation' (Greenwald, McGhee, & Schwartz, 1998, p. 1464). Because they tap automatic and spontaneous processes, implicit measures are assumed to be less prone to social desirability and self-presentational concerns.

The most widely used method or test of implicit attitudes is the Implicit Association Test (IAT; see Figure 4; Greenwald, McGhee, & Schwartz, 1998). In practice, the IAT is usually delivered by a computer. Stimuli belonging to two categories are shown on the screen, and the respondent is instructed to respond to these differentially, by pairing a positive attribute with one of the categories and a negative attribute with the other. This must be done as quickly as possible, typically by pushing the left or right shift key on the keyboard. For instance, a respondent might be asked to apply negative attributes to white faces and positive attributes to black faces. This is done for a number of trials and then the procedure is reversed, i.e., the negative attributions are now applied to faces of the other group and vice versa. The measure of interest is the average difference in reaction time when applying negative attributes versus positive attributes to outgroup faces. If the target categories are differentially linked with the attribute dimension, the respondent should presumably find it significantly simpler to assign positive evaluations than vice versa. The validity of this procedural format as a

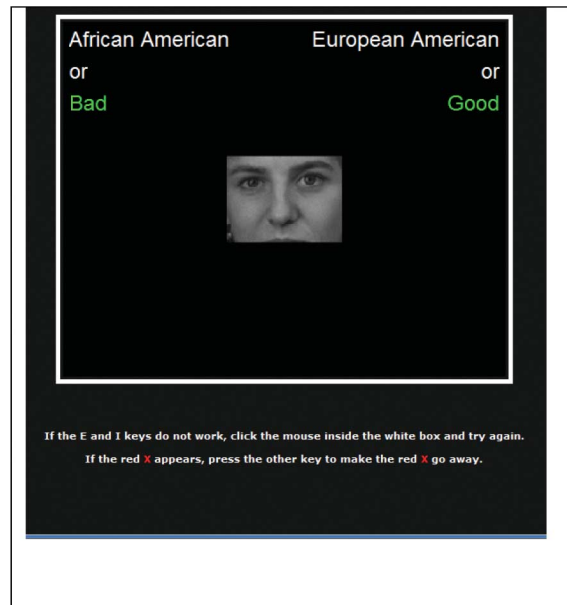


Figure 4. Sample screenshot of IAT race test.

measure of implicit attitudes and its reasonably good internal consistency has been demonstrated by several studies. In particular, in a meta-analysis conducted by Hofmann, Gawronski, Gschwender, Le, and Schmitt (2005), in which 50 studies – using adult populations – were considered, an average internal reliability of .79 was computed (see also a more recent meta-analysis by Greenwald, Poehlman, Uhlmann, & Banaji, 2009, and a review focused on construct validity by Nosek, Greenwald, & Banaji, 2007).

There is now a child version of the IAT, Child IAT, developed by Baron and Banaji (2006). As in the traditional IAT procedure, ethnic groups are presented by pictorial stimuli of children (e.g., faces of African American and European American children are used to represent race). Positive and negative attributes are presented via headphones to control for differences in reading ability. The traditional computer keyboard is replaced with a keyboard that has two large response buttons, in order to facilitate rapid responding in young children who may lack motor control. Lastly, the experimenter sits with each child and carefully explains the instructions during the task. Several studies have used the method and found children as young as five are able to complete the procedure (Baron & Banaji, 2006; Dunham, Baron, & Banaji, 2006; Banaji et al., 2008; Rutland, Cameron, Milne, & McGeorge, 2005; also Thomas, Smith & Bell, 2007, who administered the IAT to children as young as three years). Another version based on priming has been developed by Steele (Williams, Steele, & Durante, 2009).

The IAT measure remains controversial. Ostensibly measuring the link between unconscious prejudice and behaviour, its critiques have called into question the measure's construct validity, criterion validity and external validity (the degree to which laboratory experiments of the relationship between IAT measures hold up in the real world) (Blanton, Jaccard, Klick, Mellers, Mitchell, & Tetlock, 2009; Blanton & Jaccard, 2006). These criticisms are also applicable to use of the IAT with children.

In addition, due to the developmental stage of children tested with the IAT, the instrument may be assessing familiarity to target objects rather than prejudice. Even if the IAT were a valid measure, we do not know whether it is indeed needed with very young children, that is, whether it adds something to what is obtained by explicit measures. For instance, if young children respond without being affected by any desirability concern, implicit and explicit measures should be highly correlated. As far as we know there is no published study with preschool-aged children reporting correlations between implicit and explicit measures. In unpublished work, Castelli and colleagues have found moderate correlations (Castelli, personal communication, September 2009). We also wish to point out that the IAT is a differential measure and ingroup liking is confounded with outgroup prejudice as in many explicit measures of the past.

Attitudes: personal relations

These measures cover social distance, behavioural intentions, liking and peer relations.

The Social Distance Scale

One of the oldest measures in the adult literature for assessing prejudice and ethnocentrism is the Social Distance Scale, devised by Bogardus (1925) and used in hundreds of studies in the intervening years. This scale asks respondents to indicate to what extent they would accept members of an outgroup into a spatial or personally proximate relationship, in a set of increasing steps (e.g., into their country, neighbourhood, school, or in marriage). Several variants of the social distance measure have been used to assess prejudice in children.

Verna (1981) used a social distance scale to measure race preference. Children were asked to respond to a series of target-figures of different races and sexes including the following: black boy, black girl, white boy and white girl. A single figure was printed on each page of a printed booklet, and each figure appeared twice in the booklet. The children were to indicate the desired distance a figure representing themselves should have in relation to the target-figures by positioning themselves on a line in relation to the target figure. The distance was measured in cm or inches. Test reliability of the Verna social distance scale is satisfactory (Spearman-Brown = .87, Verna, 1981; Split-half reliability = .72, Katz, 1973).

A more recent variation is reported by Connolly, Fitzpatrick, Gallagher, and Harris (2006), from Queens University Belfast, and is closer in its implementation to the original Bogardus measure. A picture of a child (boy or girl, 'ethnically marked') is placed in front of the child who is being assessed, and the following questions are asked with response options such as 'YES', 'yes', 'no', 'NO'.

Imagine that this boy/girl has just started at your school and this is his/her first day in your class. I am going to ask you some questions about what you would do.

Would you go up to him/her and say 'hello'?

Would you let him/her play with you and your friends at lunchtime?

Would you share your things with him/her in class, like your coloured pencils?

Would you invite him/her to come to your house to play after school?

Would you share a secret with him/her?

Similar items were used by Cameron et al. (2007). Using scenes from a park, children were asked: how much they would like to play with the target, how much they would like to have them over to their house for a meal, and to stay overnight. Responses were made on a '1' (big frown) to '5' (big smile) scale. Although some researchers call these 'behaviour intentions' they are essentially attitudinal measures of social distance which may or may not be translated into actual behaviour.

Another measure akin to 'social distance', but perhaps less prone to demand characteristics than those mentioned above, requires children to indicate the position that they typically occupy in a predefined physical space, e.g., a playground or classroom. Children are presented with a diagram or map of the space and then asked to mark the position they typically occupy, but also may be asked to mark the position of their classmates on the map (Clack, 2007). Various measures can be derived from these maps, for each participant, and at a group level, such as (i) the relative distance of the closest ingroup and outgroup members, or (ii) the difference in the number of ingroup and outgroup members within a certain distance from the participant. Figure 5 presents an example of a mapping of 122 schoolchildren from a study by Dowdall, Tredoux and Dixon (2009).

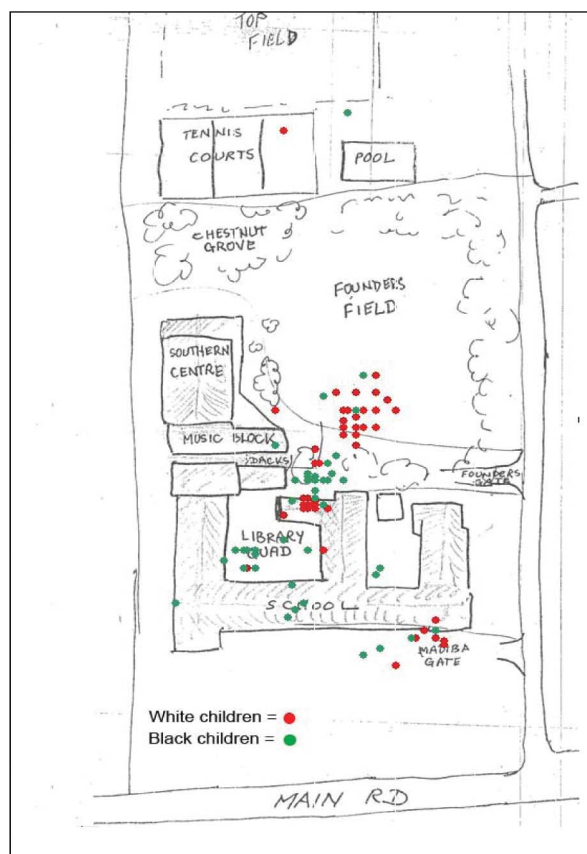


Figure 5. Mapping of self-reported position in a school playground by black and white South African, children (from Dowdall et al., 2009).

The 'Liking Board'

A measure that is similar in general procedure to the aforementioned social distance scales is the 'Liking Board'. Aboud and Mitchell (1977) presented a flat rectangular sheet of cardboard, with dimensions 20 cm × 60 cm, to children. The board had linear, centimeter scale markings. Children were asked to place photographs of members of different ethnic groups that they liked close to themselves and those that they did not like farther away on the board. This distance was recorded. Other researchers (e.g., Nesdale, Maas, Durkin & Griffiths, 2005) have simply asked children to rate how much they like the target persons. Neither is a reliable measure of attitude when only a single item is used.

Peer relationship and friendship measures

Important measures of inclusion and exclusion take peer relations as their context. Especially in situations of contact, children may be asked who are their friends, who they like to play with and who they avoid. In Pettigrew's (1988) reformulation of Allport's (1954) contact theory, intergroup friendship was proposed as a key mediator of the positive potential effects of contact between members of different groups. It is useful, therefore, to consider measures of friendship amongst children of different ethnic groups.

Commonly, social developmental researchers have examined peer relations, such as friendship, by asking children who are their best friend(s). Children are either allowed to nominate up to three or are asked to rate each classmate on a '5' (best friend) to '2' (not a friend) scale, leaving blank those they do not like at all (Aboud, Mendelson, & Purdy, 2003). Mutual best friends are those reciprocated by the nominated friend, thus requiring both friends to name each other. Another procedure is used to identify companions who may or may not be friends. 'Companions' are those rated reliably by classmates as children who hang around with each other; as such they constitute a looser but voluntary association of friends or friends of friends. Researchers then identify which peers are from the same and which from another ethnic group. Some researchers have instead asked children to name their contacts or friends from each ethnic group, so the ethnic designation is transparent to respondents. This is unlikely to be appropriate for children under eight years. Turner, Hewstone and Voci (2007) asked students to rate the following statements on a '1' (strongly disagree) to '5' (strongly agree) scale: 'I spend a lot of time doing things with Asian/white friends' and 'Asian/white friends often come around to my house'. There is still some debate as to whether one-way nominations of friends are a good measure of a relationship, in that they are not necessarily reciprocated; however, in most cases someone designated as a best friend would likely reciprocate with at minimum a 'good friend' designation. On this basis, many researchers consider one-way nominations to be valid.

Mendelson and Aboud (1999) developed the McGill Friendship Questionnaire, which consists of five items for each of six friendship provisions (reliable alliance, exciting companionship, help, intimacy, emotional security and self-validation), along with five items focusing on positive feelings (affection and satisfaction). For each item, respondents choose one of two options to reflect whether the named friend is, or is not, that way. They then decide how frequently this is the case on a four-point scale, where '1' = 'most of the time does not show the quality' and '4' = 'most of the time does show the quality' (Aboud et al., 2003). Average scores for each of the qualities,

or a composite score, is calculated to reflect the quality of the same- or cross-race friendship. Similarly, Turner et al. (2007) asked about self-disclosure (similar to intimacy) and anxiety reduction (similar to emotional security) when in the presence of cross-ethnic peers: 'How likely would it be that you would disclose a personal problem to a member of the outgroup?' Whereas Turner et al.'s items were generated on the basis of intergroup attitude research, Aboud et al.'s items were derived from research on children's peer relationships. Those wanting to use these items might modify them for particular age groups, and for the purpose of the research.

Inclusive behaviours

A vein of research on prejudice, peer relation and friendship in children has used observational methods, as an alternative to self-report. The primary advantage of observational measures is that they are typically unobtrusive and therefore more likely to accurately reflect 'real-life' patterns.

Observation of voluntary associations during free play

Because young children spend a significant amount of their time playing, most researchers consider their choice of playmate an appropriate and useful indicator of who they enjoy associating with. This is particularly useful in playschools with heterogeneous populations. For example, Fishbein and Imai (1993) observed children during free play for 30 minutes and coded behaviours over two days for six five-minute periods. This measure is commonly used to study children's play. Whenever the target child is engaged in interactive or parallel play, the race and sex of the dyadic partner is recorded. The observed proportion of playmate choices for all race/sex categories is then compared to chance expectation, or statistically controlled for numbers of classmates available from each ethnic group.

Finkelstein and Haskins (1983) observed categories of behaviour – e.g., talk, negative acts, commands, sharing toys and working together. Every 10 or 20 seconds, researchers would rate who the child was playing with and what behaviour was observed. This was done for two days, rotating the rating through a sample of 10 or 12 children during their free play. Examples of behaviours examined in Finkelstein's study were (i) attempts to enter into group play, and (ii) negative behaviours, such as engaging in physical or verbal conflict. Patterson and Bigler (2006) also observed free play. If children were playing with others, the playmates were coded as ingroup members, outgroup members or both. The percentage of interaction with ingroup members only, outgroup members only or both was calculated.

An analysis co-varying the number of available playmates of each race and sex is recommended. In addition, it is important to obtain inter-observer reliability scores for this kind of observation; for example, Howes and Wu (1990) obtained above .90 in their kappa coefficients.

Spatial proximity

Race and ethnic prejudice (along with other forms of discrimination) are perhaps too frequently physically decontextualised. When they occur in school and other settings there is a clear spatiality and temporality about them. Often the 'flashpoint' for an intergroup dispute is precisely the use of, and access to, physical space (e.g.,

land), and often the history of an intergroup dispute is ‘written’ into the environment as a pattern of spatial (e.g., residential and educational) segregation. A tradition of research that has its roots in both environmental psychology and urban geography recognises this, and directly assesses the spatiality and temporality of intergroup relations. In a particularly influential paper, Massey and Denton (1988) propose five dimensions for measuring spatial segregation and suggest measures of each.

Methods for assessing the spatial segregation or integration of intergroup protagonists in shared, ‘everyday’ spaces – that is, at a resolution much closer to the interpersonal than the approaches typically taken in urban geography – are suggested in several studies by Tredoux, Dixon, Durrheim, and colleagues (Tredoux & Dixon, 2009; Dixon, Tredoux, & Durrheim, 2008; see www.contactecology.com for an overview). These have been applied to university students and to schoolchildren (Schrieff, Tredoux, Dixon, & Finchilescu, 2005; Clack, 2007; Dowdall, Tredoux, & Dixon, 2009).

One observational method for assessing spatial segregation maps protagonists in a particular space, with a fixed periodicity. This can be done by preparing a physical map of the space beforehand (e.g., Schrieff et al. created maps of an empty student refectory), and having observers record the position of everybody in this space, using a suitable temporal sampling frame. It can also be done by taking digital photographs of the space with a fixed periodicity and then coding the position of everybody in the space afterwards (see Tredoux, Dixon, Underwood, Nunez, & Finchilescu, 2005, who built a digital model of a cascade of steps and used bespoke software to record and display the position of people in the space, see Figure 6).

A variety of measures can be derived from the observational data recorded with these methods. The space can be partitioned, for example along natural categories or axes, such as tables, or steps, or artificially into quadrants, and classic ‘segregation’ indices such as dissimilarity, and/or interaction, computed (see Massey & Denton, 1988). Alternatively, the spatial measure can be preserved and indices of concentration and/or clustering can be computed (again, see Massey & Denton, 1988).

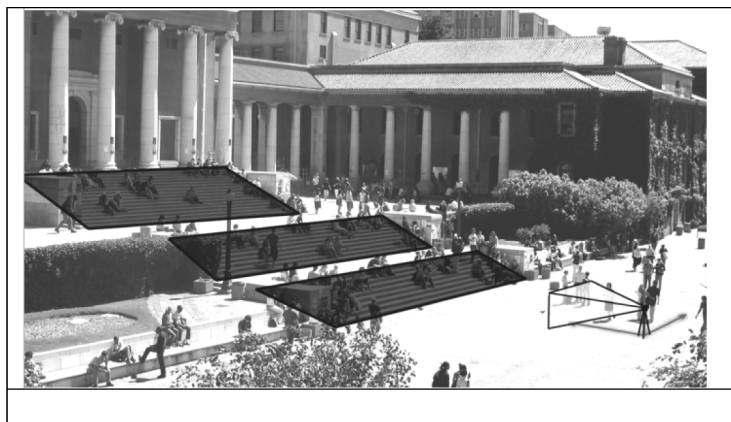


Figure 6. Setup for recording micro-segregation and integration in a public space (taken from Tredoux et al., 2005).

Helping, responding to name-callers, dyadic conversations

Cooperative learning programs often measure positive behaviours such as giving support for others' learning, asking and answering questions, and sharing information. When researchers observe voluntary contact, they often also rate the kind of behaviour being shown ranging from positive (e.g., facilitating play or sharing toys), to negative (e.g., fighting, crying), to neutral. Inter-observer ratings seem to be slightly lower than those for free-play observations.

Because bullying has become such an important issue in children's interactions, researchers have focused on types of bullying/negative behaviour, such as physical fighting, name-calling, teasing and exclusion. Observations of these negative behaviours are normally difficult but have been conducted using videos and audio recording (Hawkins, Pepler, & Craig, 2001).

Self-reports of bullying behaviour are now commonly part of surveys conducted in schools. The findings tend to converge with observational data. For example, using the Safe School Survey, Aboud and Miller (2007) found high levels of self-reported name-calling and teasing of others. Students were also fairly reliable in reporting the witnessing of bullying and being victimised by bullies. Replies to the name-caller by witnesses/bystanders have been studied by retrospective self-report or through simulated audio scenarios (Aboud & Joong, 2007; Aboud & Miller, 2007). In the intervention reported by Aboud and Miller (2007), children's verbal responses to the simulated name-calling were subsequently coded on a '0' to '6' scale of assertiveness. High levels of assertiveness are reflected in stating a behavioural rule to stop name-calling, or stating the value that name-calling is immoral, mean and/or not acceptable or warranted.

Lamb, Bigler, Liben, and Green (2007) presented children with a real-life scenario, with the help of a confederate acting as a bully. Researchers candidly observed children's attempts to stop name-calling as a victim. (This is ethically appropriate if children have been previously taught to respond in a non-confrontational manner, as they had been in this study.) For coding, an assertive scoring scale, much like Aboud and Miller's, was used to evaluate victims' responses in this real-life scenario.

Dyadic interactions can reveal children's positive and negative contact with peers. They can also reveal children's attitudes, if the task is set up appropriately. For example, Leman and Lam (2008) asked pairs of same- or cross-race dyads to select a playmate from peers belonging to various ethnic groups. The children were shown photographs of eight potential playmates and were told to imagine that these children would be new students in their class. Their task was to pick only one of the eight, and they were told that it must be the one that they would most want to play with. They were also instructed to 'talk about it together before picking'. Dyads had three minutes to discuss and make their choice. A video camera was used to film the interaction, but audio recording would have been sufficient. The researchers also looked at how assertive the children were when negotiating their selection. This behavioural category might be relevant if a difference in behaviour was expected between minority children and majority children, but a rationale for this difference would be necessary. Inter-rater agreement of behaviours was over 70%, with kappa coefficient between .72 and .84.

Aboud and Doyle (1996) observed verbal interactions between pairs of same-race children only. In this case, they were asked to talk about evaluations previously made

of ethnically different stimulus persons, represented in photographs. Each child was expressly paired with a partner who had made a different evaluation from themselves. In the dyadic discussion the task was simply to explain evaluations they had made previously, and there was no requirement that the children agree on their selection. Verbal utterances were coded using a framework based on known social-cognitive underpinnings of prejudice, such as between-race similarities, within-race differences, use of 'we' and 'they' pronouns and known exemplars. Number of verbal utterances made by each participant could then be correlated with the amount of change shown by the partner in a new evaluation of the stimulus person in question. This would reveal whether certain verbal arguments were more convincing than others. Inter-rater agreement ranged across the different codes from 75% to 100%, with an average of 91%.

Recommendations

We include here some recommendations for measuring prejudice in children, taking note of current shortcomings in existing measures:

The construct that is to be measured must be clearly defined and assessed via measures that children can relate to; they must be age-appropriate, for example, using evaluative terms that are important for children of that age. Terms that are too simple (such as 'dirty') will be resisted by older children, yet terms that are too difficult (such as 'bossy') might be difficult for younger children.

Previously used measures with known levels of acceptable reliability and validity should be employed; if modified, their reliability and validity should be reassessed. It is not a good idea to create new measures from scratch unless you are an expert. Measures may need to be tailored to the specific context, but many evaluative terms and peer relations can be used in most contexts.

The full range of measures should be considered and the best chosen for the purpose of study. Attitudes are the most commonly measured construct, but observation of relationships and behaviour is also recommended. The value of implicit measures should be weighed against current evidence for their reliability and validity.

If there is a need to assess individual children, the parent-child dyad, the school or neighbourhood climate and the society, these different measures and levels of measurement require a multi-level analysis. This might be relevant when measuring the effectiveness of interventions to change attitudes at the classroom or school level, rather than at the individual level.

Acknowledgements

This paper has been prepared collectively by the Una Quantitative Methods Learning Group. Una is a global learning initiative on children and ethnic diversity (see www.unaglobal.org) funded by the Bernard van Leer Foundation and The Atlantic Philanthropies. The membership of the group is as follows:

- Frances Aboud, McGill University, Canada (Co-Chair)
- Colin Tredoux, University of Cape Town, South Africa (Co-Chair)
- Christia Brown, University of Kentucky, USA
- Luigi Castelli, University of Padua, Italy
- Charlotte Cole, Sesame Workshop, New York, USA
- Ulrike Niens, Queens University, Belfast, Northern Ireland
- Noraini Noor, International Islamic University of Malaysia, Malaysia

- Linda Tropp, University of Massachusetts, Amherst, USA
- Buhle Zuma, University of Cape Town, South Africa.

References

- Aboud, F.E. (1988). *Children and prejudice*. Oxford: Blackwell.
- Aboud, F.E. (2002, June). *Changing racial attitudes of young children with antibias messages*. Paper presented at the annual meeting of the Society for the Psychological Study of Social Issues, Toronto.
- Aboud, F.E. (2003). The formation of ingroup favouritism and outgroup prejudice in young children: Are they distinct attitudes? *Developmental Psychology*, 39, 48–60.
- Aboud, F.E., & Doyle, A.B. (1996). Does talk of race foster prejudice or tolerance in children? *Canadian Journal of Behavioural Science*, 28, 161–170.
- Aboud, F.E., & Joong, A. (2007). Intergroup name-calling and conditions for creating assertive bystanders. In S.R. Levy, & M. Killen (Eds.), *Intergroup attitudes and relations in childhood through adulthood* (pp. 249–260). Oxford: Oxford University Press.
- Aboud, F.E., Mendelson, M.J., & Purdy, K. (2003). Cross-race peer relations and friendship quality. *International Journal of Behavioral Development*, 27, 165–174.
- Aboud, F.E., & Miller, L. (2007). Promoting peer intervention in name-calling. *South African Journal of Psychology*, 37, 803–819.
- Aboud, F.E., & Mitchell, F.G. (1977). Ethnic role taking: The effects of preferences and self-identification. *International Journal of Psychology*, 12, 1–17.
- Allport, G. (1954). *The nature of prejudice*. Cambridge, MA: Addison-Wesley.
- Banaji, M.R., Baron, A.S., Dunham, Y., & Olson, K. (2008). The development of intergroup social cognition: Early emergence, implicit nature, and sensitivity to group status. In S.R. Levy, & M. Killen (Eds.), *Intergroup attitudes and relations in childhood through adulthood* (pp. 87–104). Oxford: Oxford University Press.
- Baron, A.S., & Banaji, M.R. (2006). The development of implicit attitudes: Evidence of race evaluations from ages 6, 10 and adulthood. *Psychological Science*, 17, 53–58.
- Bennett, M., & Sani, F. (2008). Children's subjective identification with social groups. In S.R. Levy, & M. Killen (Eds.), *Intergroup attitudes and relations in childhood through adulthood* (pp. 19–31). Oxford: Oxford University Press.
- Bigler, R.S., Jones, L.C., & Loblin, D.B. (1997). Social categorization and the formation of intergroup attitudes in children. *Child Development*, 68, 530–543.
- Blank, R.M., Dabady, M., & Citro, C.F. (2004). *Measuring racial discrimination*. Washington, DC: National Academies Press.
- Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist*, 61, 27–41.
- Blanton, H., Jaccard, J., Klick, J., Mellers, B., Mitchell, G., & Tetlock, P.E. (2009). Strong claims and weak evidence: Reassessing the predictive validity of the IAT. *Journal of Applied Psychology*, 94, 567–582.
- Bogardus, E.S. (1925). Measuring social distances. *Journal of Applied Sociology*, 9, 299–308.
- Brown, R. (1995). *Prejudice: Its social psychology*. Cambridge, MA: Blackwell.
- Cameron, L., Rutland, B., & Brown, R. (2007). Promoting children's positive intergroup attitudes towards stigmatized groups: Extended contact and multiple classification skills training. *International Journal of Behavioral Development*, 31, 454–466.
- Castelli, L., Zogmaister, C., & Tomelleri, S. (2009). The transmission of racial attitudes within the family. *Developmental Psychology*, 45, 586–591.
- Clack, B.T. (2007). *Racial contact and isolation in a post-apartheid South African school*. (Unpublished Ph.D. dissertation.) University of Lancaster, UK.
- Clark, K.B., & Clark, M.K. (1947). Racial identification and preference in Negro children. In T. Newcomb, & E. Hartley (Eds.), *Readings in social psychology* (pp. 169–178). New York: Holt, Rinehart, & Winston.
- Connolly, P., Fitzpatrick, S., Gallagher, T., & Harris, P. (2006). Addressing diversity and inclusion in the early years in conflict-affected societies: A case study of the Media Initiative for Children, Northern Ireland. *International Journal of Early Years Education*, 14, 263–278.
- Doyle, A.B., & Aboud, F.E. (1995). A longitudinal study of White children's racial prejudice as a social cognitive development. *Merrill-Palmer Quarterly*, 41, 210–229.

- Dowdall, G., Tredoux, C.G., & Dixon, J.A. (2009, August). *Micro-ecology of intergroup contact in a Cape Town High School*. Paper presented at 15th Annual Congress of the Psychological Society of South Africa, Cape Town.
- Dunham, Y., Baron, A.S., & Banaji, M.R. (2006). From American city to Japanese village: A cross-sectional investigation of implicit race attitudes. *Child Development*, 77, 1268–1281.
- Finkelstein, N.W., & Haskins, R. (1983). Kindergarten children prefer same-color peers. *Child Development*, 54, 502–508.
- Fishbein, H.D., & Imai, S. (1993). Preschoolers select playmates on the basis of gender and race. *Journal of Applied Developmental Psychology*, 14, 303–316.
- Greenwald, A.G., McGhee, D.E., & Schwartz, J.K.L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464–1480.
- Greenwald, A.G., Poehlman, T.A., Uhlmann, E.L., & Banaji, M.R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97, 17–41.
- Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*, 31, 1369–1385.
- Hawkins, D.L., Pepler, D.J., & Craig, W.M. (2001). Naturalistic observations of peer interventions in bullying. *Social Development*, 10, 512–527.
- Howes, C., & Wu, F. (1990). Peer interactions and friendships in an ethnically diverse school setting. *Child Development*, 61, 537–541.
- Katz, P.A. (1973). Stimulus predifferentiation and modification of children's racial attitudes. *Child Development*, 44, 232–237.
- Katz, P.A., & Zalk, S.R. (1974). Doll preferences: An index of racial attitudes? *Journal of Educational Psychology*, 66, 663–668.
- Katz, P.A., Sohn, M., & Zalk, S.R. (1975). Perceptual concomitants of racial attitudes in urban grade-school children. *Developmental Psychology*, 11, 135–144.
- Lamb, L., Bigler, R.S., Liben, L.S., & Green, V. (2007, March). *When children bully peers for sex 'inappropriate' behavior: Theoretical and practical implications of anti-bullying interventions*. Poster presented at the biennial meeting of the Society for Research in Child Development, Boston, MA.
- Leman, P.J., & Lam, V. (2008). The influence of race and gender on children's conversations and playmate choices. *Child Development*, 79, 1329–1343.
- Massey, D.S., & Denton, N.A. (1988). The dimensions of residential segregation. *Social Forces*, 67, 281–315.
- Mendelson, M.J., & Aboud, F.E. (1999). Measuring friendship quality in late adolescents and young adults: The McGill Friendship Questionnaires. *Canadian Journal of Behavioral Science*, 31, 130–132.
- Nagata, D.K. (2001). The relationships between ethnic attitude and egocentrism measures. *Journal of Genetic Psychology*, 146, 151–160.
- Nesdale, D., Maass, A., Durkin, K., & Griffiths, J. (2005). Group norms, threat and children's racial prejudice. *Child Development*, 76, 652–663.
- Nosek, B.A., Greenwald, A.G., & Banaji, M.R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In J.A. Bargh (Ed.), *Automatic processes in social thinking and behavior* (pp. 265–292). New York: Psychology Press.
- Patterson, M.M., & Bigler, R.S. (2006). Preschool children's attention to environmental messages about groups: Social categorization and the origins of intergroup bias. *Child Development*, 77, 847–860.
- Pettigrew, T.F. (1998). Intergroup contact theory. *Annual Review of Psychology*, 49, 65–85.
- Rutland, A., Cameron, L., Milne, A., & McGeorge, P. (2005). Social norms and self-presentation: Children's implicit and explicit intergroup attitudes. *Child Development*, 76, 451–466.
- Schrieff, L., Tredoux, C., Dixon, J.A., & Finchilescu, G. (2005). Racial contact and seating patterns in university dining halls. *South African Journal of Psychology*, 35, 433–443.
- Thomas, S., Smith, R.B., & Bell, P. (2007). Implicit attitudes in very young children: An adaptation of the IAT. *Current Research in Social Psychology*, 13, 75–85.
- Tredoux, C.G., & Dixon, J.A. (2009). Mapping the multiple contexts of racial isolation: The case of Long Street, Cape Town. *Urban Studies*, 46, 761–777.

- Tredoux, C.G., Dixon, J.A., Underwood, S., Nunez, D. & Finchilescu, G. (2005). Preserving spatial and temporal dimensions in observational data of segregation. *South African Journal of Psychology*, 35, 412–432.
- Turner, R.N., Hewstone, M., & Voci, A. (2007). Reducing explicit and implicit outgroup prejudice via direct and extended contact: The mediating role of self-disclosure and intergroup anxiety. *Journal of Personality and Social Psychology*, 93, 369–388.
- Verna, G. (1981). Use of a free-response task to measure children's race preferences. *The Journal of Genetic Psychology*, 138, 87–93.
- Williams, A., Steele, J.R., & Durante, S. (2009). The development of implicit racial bias among 6- and 9-year-olds: Is forced categorization by race a precondition for bias? Manuscript, York University, Toronto.
- Williams, J.E., Best, D.L., & Boswell, D.A. (1975). The measurement of children's racial attitudes in the early school years. *Child Development*, 46, 494–500.
- Wright, S.C., & Tropp, L.R. (2005). Language and intergroup contact: Investigating the impact of bilingual instruction on children's intergroup attitudes. *Group Processes and Intergroup Relations*, 8, 309–328.