

The Calculation of Discrepancy Scores in the Context of Biased Self-Perceptions of Acceptance

Haley F. Stephens · Janet A. Kistner · Rebecca J. Lynch

Published online: 21 October 2014
© Springer Science+Business Media New York 2014

Abstract Children's biased self-perceptions of social acceptance have been linked to increased aggressive and depressive symptoms. The examination of biased self-perceptions as a risk factor for negative socio-emotional outcomes is an important line of research but inconsistency exists within the field regarding which method of difference score calculation (i.e., difference or residual scores) is used to quantify the degree to which a person holds biased self-perceptions. This study compared the methods on three issues: 1) the extent to which the methods agree about which participants hold biased self-perceptions; 2) the strength of the estimated relationship between biased-self perceptions of social acceptance and socio-emotional outcomes (i.e., aggression and depressive symptoms); 3) the strength of those same estimated relationships while controlling for a potentially confounding variable (actual social acceptance). Data from a previously published study (David and Kistner *Journal of Abnormal Child Psychology: An Official Publication of the International Society for Research in Child and Adolescent Psychopathology*, 28(4), 327–337, 2000) of 871 third through fifth grade students who completed measures of their depressive symptoms as well as their own, and their peers', social acceptance and aggression, was used in the present study. Results showed the difference and residual methods demonstrated moderate agreement regarding bias classification. The difference method correlated more strongly with aggression than did the residual method. Conversely, the residual method correlated more strongly with depressive symptoms than did the difference method. Once actual social acceptance was added to the difference score prediction model the methods became equivalent in their associations with both outcomes. Implications regarding

the importance of statistically controlling for the actual social acceptance component of the bias score are discussed.

Keywords Self-perceptions · Discrepancy measurement · Peer relationships · Positive illusory bias

A large body of research exists on children's self-perceived competence in a variety of domains (e.g., behavioral conduct, academic achievement, social acceptance). In the past two decades, a subset of this research has constrained its focus to the study of biased self-perceptions (i.e., self-perceptions that are either more positive or more negative than objective indicators warrant) and how these self-perceptions relate to adverse outcomes. Positively biased self-perceptions of social acceptance have been shown to be associated with Attention Deficit Hyperactivity Disorder (AD/HD) symptoms (Hoza et al. 2000; Owens and Hoza 2003; Hoza et al. 2004) and aggression (David and Kistner 2000; Orobio de Castro et al. 2007; White and Kistner 2011), while negatively biased self-perceptions of social acceptance have been linked to depressive symptoms (Campbell and Fehr 1990; Cole et al. 2001). Given these associations, consideration of children's biased self-perceptions of social acceptance is warranted for the identification of children who may be at risk for negative socio-emotional outcomes. Despite growing interest in children's biased self-perceptions of social acceptance, considerable debate exists in the literature over the appropriate method for calculating a bias score (i.e., a measure of the extent to which children systematically differ in their self-perceived social acceptance from others informants' perceptions of their social acceptance). Investigators typically obtain a rating of perceived social acceptance (PSA) from the target child and a rating of the child's actual social acceptance (ASA) from an informant (e.g., parent, teacher, peer). The inconsistency within the biased social self-perception literature concerns the

H. F. Stephens (✉) · J. A. Kistner · R. J. Lynch
Psychology Department, Florida State University, 1107 W. Call
Street, Box 207900, 32306-4301 Tallahassee, FL, USA
e-mail: hstephens@psy.fsu.edu

choice of method used to compare the ratings of the target child's PSA and ASA. The mathematical properties demonstrated by each method have the potential to produce different results as an artifact of the data (e.g., differences between the variances of the participants' ratings or the correlation between the participants' ratings). Furthermore, the research question posed and the characteristics of the population in question may lend themselves better to one type of computation method than another. The inconsistency with respect to choice of bias score calculation method is a barrier to comparing results across studies. A consensus about when it is appropriate, or even preferable, to use each type of calculation method is necessary to create coherence in the growing field of biased self-perception research. The two most commonly used methods to measure bias are the difference between the participants' standardized PSA and ASA ratings and the residual difference between the participants' PSA and ASA ratings. The standardized difference score (DIF) is the subtraction of the standardized ASA rating from the standardized PSA rating (e.g., z score of child's PSA rating minus z score of teacher's ASA rating). The DIF approach is typically used to place participants' PSA and ASA ratings on the same metric and/or to increase the interpretability of the difference score by placing it on a scale of variability with a mean of 0 and a standard deviation of 1 (e.g., the z distribution). Unlike DIF, the residual difference score (RES) is calculated by using the ASA rating to predict the PSA rating and measuring the difference between the predicted PSA rating and the observed PSA rating. The ASA rating is used as the independent variable, and the PSA rating is used as the dependent variable, in a regression equation model. The difference between the predicted and observed PSA rating is RES. This score is typically standardized by dividing it by the standard deviation of the unstandardized residual difference scores. The mathematical properties of the different computations dictate that the relationship between the bias score and variables of interest (e.g., AD/HD, aggression, depression) will be different depending on the differences between the variances of the participants' ASA and PSA ratings and the correlation between the ASA and PSA ratings. For example, when the variance of the ASA ratings is equal to the variance of the PSA ratings (e.g., when standardized), both sets of ratings contribute equally to the difference score, as is the case with the DIF method. The RES approach is used to account for the theoretical assumption that PSA is based, at least in part, on ASA (via feedback from peers, homework assignments, etc.). The RES approach defines perceptual bias as the degree to which PSA is higher or lower than would be predicted by ASA and it controls for the potentially confounding effects of ASA in the predictive relationship between perceptual bias and the outcome variable of interest. Research in the area of perceptual bias often examines the extent to which biased self-perceptions are related to some variable of interest (e.g. AD/

HD, aggression, depression). When using the DIF approach, the resulting DIF bias scores are systematically correlated with the ASA and PSA ratings from which they were created (Cronbach 1958; Zuckerman and Knee 1996; Griffin et al. 1999). Suspicion is warranted that it is not the DIF bias score that is related to the variable of interest, but rather the ASA rating that accounts for the association between the DIF bias score and the variable of interest. To mitigate this concern, researchers may choose to control for ASA when statistically testing for a relationship between bias and the variable of interest. To date, all studies which attempt to control for ASA use the RES approach because ASA is partialled out during the creation of the RES bias score. This leaves RES to act as a semi-partial correlate when it is used to predict a variable of interest. Studies which utilize a DIF approach *and* control for one of the components of the bias score are markedly absent in the perceptual bias literature. Never would a researcher conclude that race is associated with intelligence without considering the confounding variable of socioeconomic status. Similarly, a researcher should not conclude that children's biased self-perceptions of their social acceptance are associated with a variable of interest without considering the confounding variable of the children's actual social acceptance (i.e., as rated by others). Why then, is there a dearth of studies that have controlled for well-known associations between ASA and the variable of interest? Further, would those studies that examined the relationship between bias and a variable of interest using DIF have found the same results if ASA had been controlled in the analyses? And finally, how would those results compare to results generated with the use of RES to examine the same associations? In a seminal study on discrepancy measurement, De Los Reyes and Kazdin (2004) examined the differences among these two methods using a single sample of children with externalizing behavior problems. DIF and RES scores were computed from mother and child ratings of the children's externalizing symptoms. The results showed a low ($r=.14$), but significant, correlation between mother-rated and child self-rated externalizing symptoms. As a result, RES correlated almost perfectly with the mothers' ratings, which were used as the dependent variable in the regression analysis used to compute RES. This was discussed as problematic because any correlation between the RES and a variable of interest was essentially identical to the correlation between the mothers' ratings and the same variable of interest. DIF, however, correlated equally with each of the component variables from which it was created. The authors placed emphasis on reconciling informant discrepancies and the importance of each informant's rating to contribute equally to the discrepancy score, and thus DIF scores were recommended as the best method for measuring informant discrepancies. Recent research has called into question the use of DIF scores for the study of informant discrepancy due to the fact that interpretations based on

analyses that use a DIF score can be inconsistent with results from equivalent models that utilize both component variables separately (Laird and Weems 2011). An association between the DIF score and an outcome variable may be driven by a strong association between only one of the component variables and that outcome variable, rather than the difference between component variables. The study by De Los Reyes and Kazdin (2004) is widely cited as justification for the use of DIF for the calculation of a bias (Diamantopoulou et al. 2008; Whitton et al. 2008). However, informant discrepancies, as examined in the De Los Reyes and Kazdin (2004) study, and biased self-perceptions are theoretically different constructs. The former assumes there is no true measure of the variable of interest (e.g., child externalizing problems). Instead, the goal is to obtain a measure of the discrepancy between the two participants' ratings of that variable with the assumption that both participants are equally accurate in their ratings. In contrast, the examination of biased self-perceptions is concerned with the degree to which the target child's self-views are either overly positive or over negative when compared to another person's view of the target child. A bias score is a measure of the degree to which the target child's rating is biased, or systematically different from others' ratings of the target child. Despite this theoretical difference, many researchers in the area of perceptual bias choose to utilize the DIF method without consideration of the RES approach. While DIF may be the appropriate method for calculating bias scores for *some* studies it may not be the optimal measure for *all* research questions and populations of interest. The biased social self-perception literature typically designates the person or persons rating the target child (i.e., peers, teachers, or parents) as the criterion against which to compare the target child's self-perceptions. While it is the case that truly *objective* observers of children's behaviors do not exist, it is often presumed that the people rating the child are more objective than the child himself (Hoza et al. 2002; Owens and Hoza 2003; David and Kistner 2000). The term "actual acceptance" is merely a shorthand notation used to distinguish the criterion ratings from the child's self-perceptions. There is evidence, however, to suggest that children over- or underestimate their social acceptance when self-perceptions are compared with social acceptance ratings across different types of informants (e.g., peers; teachers). For example, a study conducted by Hoza and colleagues (2004) demonstrated that children with AD/HD were more likely to overestimate their social acceptance regardless of whether the teacher, mother, or father was used as the criterion rater. Further, in a study conducted by Owens and Hoza (2003), children with AD/HD overestimated their scholastic competence when math achievement (i.e., standardized test scores) was used as the criterion and also when teacher report was used as the criterion. Finally, Swanson et al. (2012) showed that girls with ADHD tended to overestimate their competence relative to control children

across domains and regardless of the criterion rater. Specifically, for the area of scholastic competence, girls' self-perceptions were systematically different from both standardized achievement tests and teacher ratings. The same was true for the area of social acceptance when the criterion was measured by peers, teachers, and mothers. Taken together, these studies provide support for the notion that there tends to be agreement across informants about children's social acceptance and that children are likely to show the same perceptual bias regardless of whether the criterion measure (i.e., "actual" social acceptance) is provided by peers, teachers or parents.

The Present Study

The aim of the present study is to address inconsistencies in the field of self-perception research concerning which method is used to calculate the bias score. Comparisons of the results obtained using each method of bias score calculation are conducted in an effort to address three controversial issues. The first issue of interest concerns the notion that the methods of bias score calculation may identify different children within the same sample as holding positively or negatively biased self-perceptions. The extent to which the two methods agree about which children are classified as positively/negatively biased was examined. It was hypothesized that the agreement level would be lower than researchers would like to see if the methods were to be used interchangeably in the literature.

The second issue of interest in the present study concerns the idea that researchers may find different answers to a research question depending on which method of bias score calculation is used in the analysis. The present study utilized measures of children's ASA and PSA to calculate a measure of their biased self-perceptions of social acceptance (bias). The relationship between bias and aggression (David and Kistner 2000; Orobio de Castro et al. 2007) and the relationship between bias and depressive symptoms (Campbell and Fehr 1990; Rudolph and Clark 2001; Cole et al. 2001) are well-documented. In the present study, these two relationships were examined using each of the two bias score calculation methods. Comparisons of the results were conducted to determine whether differences exist in the strengths of the estimated relationships between bias and outcome variables (i.e., aggression and depressive symptoms).

The final issue of interest in the present study considers the potential importance of controlling for ASA when examining the relationship between bias and some outcome variable. If ASA is associated with the outcome variable then it should be controlled in the analysis of the relationship between bias and the outcome variable. The association between children's ASA and aggression (Pope et al. 1991; Kupersmidt and Patterson 1991; Coie et al. 1992) and the

relationship between ASA and depressive symptoms (Hymel et al. 1990; Zimmer-Gembeck et al. 2007) are also well-documented. As previously noted, there should be no need to control for ASA when examining the relationship between the RES method and outcome variables. The present study examined the relationships between DIF and the outcome variables while controlling for variance in those relationships that was accounted for by ASA and then compared those results with the zero-order relationships between RES and the outcome variables. In contrast to the previous set of analyses, it was hypothesized that the two methods would become nearly equivalent in their respective predictions of the outcome variables of interest.

The empirical examination of these three issues has important implications for the future of perceptual bias research. If the methods of bias score calculation demonstrate notable disagreement about which children are classified as holding positively biased self-perceptions, investigators should be encouraged to consider how they want to conceptually define bias in their studies and choose the method that best aligns with their definition. Further, method disagreement would be an indicator that the methods operate differently, and thus, should not be used interchangeably. This notion would be bolstered if the methods produce significantly different estimates of the relationships between bias and the outcome variables. Finally, an examination of the estimates of the relationships between bias and the outcome variables when controlling for ASA may illustrate the potential importance of controlling for the actual competence component variable in all perceptual bias studies. Overall, the proposed study may contribute to the goal of method consistency in the field of biased self-perception research.

Method

Participants

The present study is based on archival data from David and Kistner (2000). To recruit the sample, parent permission to participate was solicited from 1717 children in the third, fourth, and fifth grades at nine public elementary schools located in rural and suburban areas of northern Florida. Consent was obtained from 962 children (56 % consent rate). At the time of testing, four children declined to participate and three did not participate due to absences. The present study relies on peer nominations as measures of social acceptance and aggression. Therefore, only participants from classes in which a clear majority (i.e., more than 60 % of the class) of the classroom students participated were included in the analyses in order to increase confidence

in the validity of the measures (60 participants from nine classrooms were dropped from the sample). Two participants were missing all self-report data and two participants lacked variability in their peer ratings. As such, those participants were also dropped from the sample. Of the remaining sample (N=890), complete data was obtained from 871 children (one participant was missing the PSA measure and 18 were missing the depression measure). The 19 children with incomplete response profiles were excluded from the study analyses. As shown in Table 1, the 19 participants who were dropped did not differ from the 871 participants used in the sample on important demographic and study variables.

Analyses in the present study are based on 871 participants from 56 classrooms. Of this sample, 46 % were male and 54 % were female. The distribution of ethnicity of this sample was 67 % Caucasian, 30 % African American, 2 % Asian, and 1 % Hispanic. The average age of the participants was 9.6 years (SD=1.1). There were 286 third, 256 fourth, and 329 fifth graders in the sample.

Measures

Actual Social Acceptance (ASA)

Sociometric ratings were used to measure each participant’s actual social acceptance. Participants were given a roster of student names from their own classroom and were asked to rate on a 5-point scale (ranging from 1=“do not like at all” to 5=“like very much”) how much they liked each classmate. Ratings were summed, averaged, and standardized within class to form a measure of social acceptance. Peer ratings have been demonstrated to be reliable and valid indices of social acceptance (Hartup 1983). Test-retest reliabilities of 4- and 6-week intervals have been reported to be .81 and .84 respectively (Asher et al. 1979; Oden and Asher 1977). In the present sample, 6 month test-retest reliability was .76.

Table 1 Comparison of participants dropped and included in the study

	Dropped (n=19)		Included (n=871)		t	p
	M	SD	M	SD		
Age	9.53	.96	9.63	1.05	-.42	.67
PSA	2.92	.73	2.92	.71	-.03	.98
ASA	2.86	.54	3.03	.69	-1.11	.27
Aggression	.24	.29	.24	.27	.01	.99
	Boys	Girls	Boys	Girls	χ^2	p
Sex	n=8	n=11	n=400	n=471	.01	.92

Perceived Social Acceptance (PSA)

The Self-Perception Profile for Children (SPPC; Harter 1985) was used to measure each participant's perceived social acceptance. The SPPC is made up of six subscales measuring children's self-perceptions of social acceptance, scholastic competence, athletic competence, physical appearance, behavioral conduct, and global self-worth. Each subscale is comprised of six questions for a total of 36 questions, each rated on a 4-point Likert-type scale. Only the social acceptance subscale was used in the present study. Answers to the six pertinent questions were averaged, with higher numbers indicating more positive self-perceptions. Cronbach's alphas for the subscales as reported by Harter (1985) based on four samples ranged from .71 to .86, indicating acceptable internal consistency. Cronbach's alpha for the social acceptance subscale of the SPPC in the present sample was .71.

Perceptual Bias (bias)

Two measures of perceptual bias in the social acceptance domain were created using the two methods of discrepancy score calculation.

Standardized The DIF bias score was created using a two-step process. First, participants' ASA and PSA ratings were standardized within the sample by placing them on z distributions with a mean of 0 and a standard deviation of 1. Second, participants' respective standardized ASA ratings were subtracted from their respective standardized PSA ratings. The resulting values acted as the standardized difference bias scores.

Residual The RES bias score was created using a three-step process. First, participants' PSA ratings were regressed on their ASA ratings. Second, the remaining variances were saved as residual scores. Third, the residual scores were standardized within the sample by dividing them by the standard deviation of the unstandardized residual scores. The remaining values acted as the residual difference bias scores.

Aggression

Participants' levels of aggression were assessed using a peer nomination measure (Crick et al. 1996; Crick and Grotpeter 1995) that consists of three subscales (i.e., overt aggression, relational aggression, and prosocial behavior). Only the overt and relational aggression subscales were considered in the current study. The overt aggression scale consists of five items related to verbal (e.g., "call others mean names") and physical (e.g., "hit, kick, or punch other kids") aggression. The relational aggression scale consists of five items related to behaviors intended to hurt another peer's relationships (e.g., "when

they are mad at a person, they get even by keeping the person from being in their group of friends"). Participants were given a roster of student names from their own class and asked to nominate three classmates who fit each of the behavioral descriptors. For each item, the number of nominations that each child received was summed and divided by the total number of possible nominations. The average number of nominations across all ten overt and relational aggression items was summed, averaged, and standardized within class to form a measure of aggression. Both of the aggression subscales have been shown to be highly reliable with internal consistency reliabilities of .94 for overt aggression and .83 for relational aggression (Crick and Grotpeter 1995). In the present sample, internal consistency reliabilities are .97 for overt aggression and .89 for relational aggression (David and Kistner 2000).

Depression

Participants' level of depressive symptoms was measured using the Children's Depression Inventory (CDI; Kovacs 1992). It is a commonly used 27-item child-report measure of the presence and severity of symptoms of depression during the previous 2 weeks. The CDI has good internal consistency, test-retest reliability, and discriminative validity (Carlson and Cantwell 1979).

Procedure

Participants completed the measures described above as part of a larger test battery during two 60-minute data collection sessions. The two sessions were conducted approximately 1 week apart and at least 3 months into the school year in order to ensure that all students were acquainted with each other. Trained graduate and undergraduate research assistants presented instructions to small groups of children at the start of each testing session and then the participants were allowed to complete the measures at their own pace. Participants were encouraged to respond to all items and they were individually queried about any omitted items. If participants reported that they lacked sufficient information to provide a rating or a nomination, the question was left blank.

Results

Descriptive Statistics

Means, standard deviations, and concurrent correlations between bias score components, calculation methods, and variables of interest are presented in Table 2. Children's sex was significantly related to ASA. Thus, sex was retained as a

Table 2 Means, standard deviations, and correlations

		<i>M</i>	<i>SD</i>	1.	2.	3.	4.	5.	6.	7.	8.
1.	Age	9.63	1.05	-							
2.	Sex	–	–	–.03	–						
3.	Aggression	.24	.27	.01	–.30**	–					
4.	Depression	10.07	8.01	.01	.00	.16**	–				
5.	PSA	2.92	.71	.05	–.01	0.03	–.48**	–			
6.	ASA	3.03	.69	–.03	.16**	–.28**	–.18**	.18**	–		
7.	DIF	.00	1.28	.06	–.14**	.24**	–.23**	.64**	–.64**	–	
8.	RES	.00	1.00	.06	–.04	.08*	–.46**	.98**	.00	.77**	–

N=871; **p*<.05; ***p*<.01

covariate in the regression analyses because ASA is used to create both calculation methods. Children’s age was not significantly related to any other variable. As such, age was dropped as a covariate for all further analyses. As expected, DIF correlated strongly and evenly with its component variables and RES demonstrated a near perfect correlation with PSA and zero correlation with ASA. A weak, but statistically significant correlation was found between the two bias score component variables. DIF demonstrated a very strong correlation with RES. Both methods were significantly negatively correlated with Depression and significantly positively correlated with Aggression.

Bias Classification Analyses

A reliability analysis using the Kappa statistic was performed to determine the consistency with which participants were classified as positively or negatively biased by the DIF and RES methods. For both methods, participants were coded as positively biased if their bias score was greater than or equal to zero and negatively biased if their bias score was less than zero. The reliability estimate of the DIF and RES methods demonstrated moderate agreement (78.9 %; $\kappa=.58, p<.001$). A depiction of where the DIF and RES methods disagreed is shown in Fig. 1. Participants were sorted and plotted according to the magnitude of their DIF bias scores and then the participants’ RES bias scores were plotted against their DIF scores. Quadrant 1 denotes bias score classification disagreement characterized by a positive RES score and a negative DIF score, whereas quadrant 4 denotes disagreement in the opposite direction. Quadrant 2 holds participants who were classified as positively biased according to both methods and quadrant 3 captures participants who were classified as negatively biased according to both methods.

In a separate set of analyses, a cut score of one standard deviation above or below the mean of bias was used to identify children who held either strongly positively or strongly negatively biased views of their social acceptance according to each calculation method. Means and standard

deviations of the standardized component variables can be found in Table 3 for participants who were identified as either strongly positively or strongly negatively biased. As shown in Fig. 2, those who had strong DIF bias scores tended to have extreme PSA and ASA scores whereas participants who had strong RES bias scores tended to have even more extreme PSA scores but average ASA scores. Taken together, these results indicate that the two methods are capturing different subsamples of participants.

Concordance Across Calculation Methods

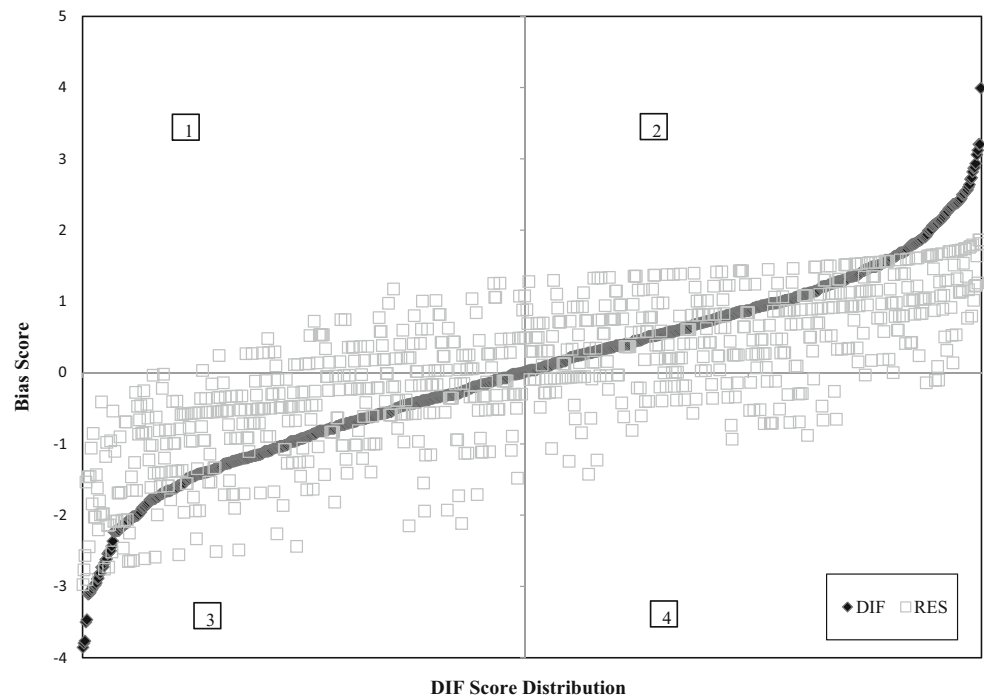
Associations with Aggression and Depression

The Hotelling-Williams test for difference of correlations within a dependent sample was used to detect differences in the strengths of the correlations between the calculation methods and the outcome measures (Steiger 1980). It was confirmed that the relationship between DIF and Aggression, $r(871)=.24$, was significantly stronger than the relationship between RES and Aggression, $r(871)=.08$; Williams $t(871)=7.26, p<.01$. The results of a separate Hotelling-Williams test demonstrated that the relationship between RES and Depression, $r(871)=-.46$, was significantly stronger than the relationship between DIF and Depression, $r(871)=-.23$; Williams $t(871)=11.53, p<.01$. According to zero-order correlations, the DIF method produced a stronger association with Aggression than did the RES method. Conversely, the RES method produced a stronger association with Depression than did the DIF method.

Controlling for Actual Social Acceptance

A stepwise regression analysis was conducted to examine the relationship between DIF and Aggression while controlling for variance accounted for by ASA. DIF was entered as a predictor variable in step one and ASA was added as a predictor in step two. A separate regression analysis was conducted to examine the relationship between RES and

Fig. 1 Bias classification method agreement along the DIF score distribution



Aggression in a way that would produce numbers against which to compare the DIF stepwise regression analysis. Results of both analyses are presented in Table 4. When ASA was controlled in the model, the strength of the relationship between DIF and Aggression decreased. For the DIF method, the partial and semi-partial correlations were notably weaker than the zero-order correlation with Aggression but were more aligned with the zero-order correlation between the RES method and Aggression.

A stepwise regression analysis was also conducted to examine the relationship between DIF and Depression while controlling for variance accounted for by ASA. Again, a separate regression analysis was run for the relationship between RES and Depression to aid method comparisons. Results of both analyses are presented in Table 5. Once again, the partial and semi-partial correlations for the DIF method became aligned with the zero-order correlation for the RES method once ASA was controlled. DIF remained significantly

associated with Depression once ASA was added to the models but this time, the relationship between DIF and Depression became stronger as a result of controlling for ASA in the model. This suppression effect occurred because ASA was not as strongly correlated with depressive symptoms as PSA. Thus, when ASA was controlled, the predictive relationship between DIF and Depression became more dependent on PSA. These results highlight the notion that predictive relationships between bias and an outcome variable are highly dependent upon correlations between the bias score's component variables and that outcome variable.

Discussion

Researchers in the field of perceptual bias utilize bias scores to quantify the degree to which children over- or under-estimate

Table 3 Means and standard deviations of standardized component variables by valence of bias

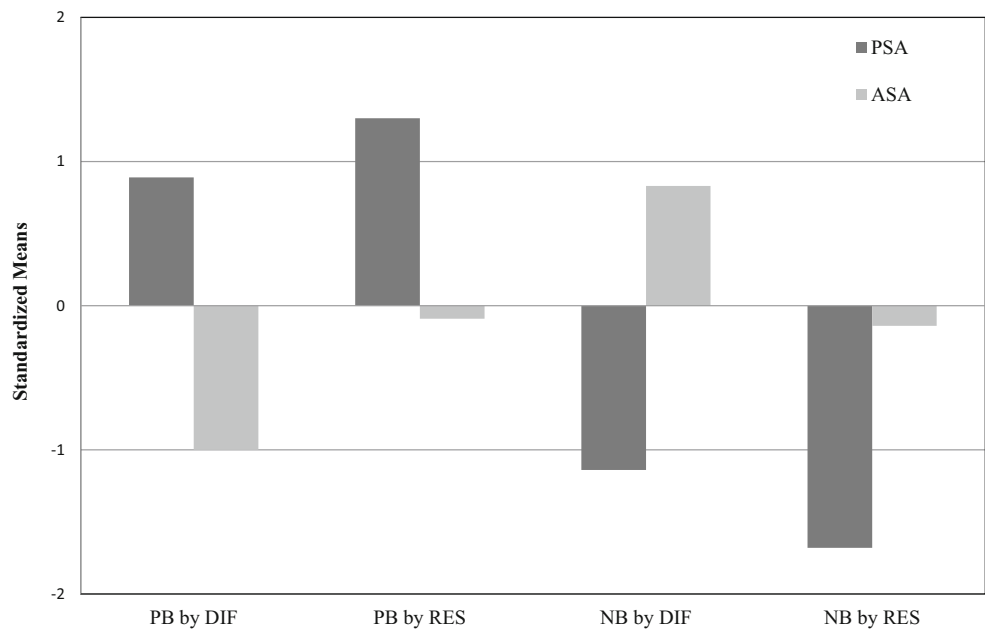
	Strongly PB ¹ by DIF		Strongly PB ¹ by RES		Strongly NB ² by DIF		Strongly NB ² by RES	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
PSA	0.89	0.58	1.30	0.25	-1.14	0.86	-1.68	.54
ASA	-1.00	0.74	-0.09	0.96	0.83	0.84	-0.14	0.99

PB, Positively biased; *NB*, negatively biased; *DIF*, standardized difference score; *RES*, residual difference score; *PSA*, perceived social acceptance; *ASA*, actual social acceptance

¹ Strongly PB group defined as participants whose bias score was more than one standard deviation above the mean of that method

² Strongly NB group defined as participants whose bias score was more than one standard deviation below the mean of that method

Fig. 2 Mean actual and perceived social acceptance for strongly biased groups



Groups by Valence and Method of Bias Score

their social acceptance. The methods of bias score computation examined in this study have unique mathematical properties which have the potential to impact the magnitude of the resulting difference scores. Due to inconsistency with respect to which bias score calculation method is utilized, researchers should not expect to easily compare results across studies. Findings from the present study illuminate the frequency of method disagreements and demonstrate important differences in the predictive potential of each method.

The DIF and RES methods demonstrated only moderate agreement about which children held positively versus negatively biased self-perceptions of social acceptance. Further, participants who held strongly biased self-perceptions according to the DIF method tended to have opposing extreme scores for ASA and PSA, whereas participants who held strongly biased self-perceptions according to the RES method only tended to have extreme scores for PSA. The notion that the methods capture different sub-samples of children as holding either positively or negatively biased self-perceptions should be troubling for researchers in the field of perceptual bias. When those children who clearly hold positively biased self-perceptions according to the DIF method are identified as

holding negatively biased self-perceptions by the RES method (or vice versa), investigators should not expect to find consistent results across the methods. The correlates of holding biased self-perceptions according to one method will likely differ from the correlates of holding the biased self-perceptions of the same magnitude according to the other method. These findings support the idea that the DIF method should not be used interchangeably with the RES method. Instead, investigators may wish to think about how bias is conceptualized in their studies and choose the method which corresponds best to their definitions.

The idea that the DIF and RES methods capture different sub-samples of children as holding either positively or negatively biased self-perceptions should be worrisome for practitioners as well. One way in which practitioners may wish to utilize bias scores is to aid in identification of children who are at-risk for the negative outcomes that are associated with bias (i.e., aggression and depression). In the present study, the DIF and RES methods identified different children as holding bias. Therefore, clinicians may not be able to reliably identify at risk-children based on a bias score if those clinicians use different calculation methods.

Table 4 Hierarchical regression analyses predicting aggression

Step	Method	B	t	p	r	pr	sr
Step 1	DIF	.20	6.31	<.001	.24	.21	.20
	RES	.07	2.11	.04	.08	.07	.07
Step 2	DIF	.09	2.23	.03	.24	.08	.07
	ASA	-.18	-4.29	<.001	-.28	-.14	-.13

Table 5 Hierarchical regression analyses predicting depression

Step	Method	B	t	p	r	pr	sr
Step 1	DIF	-.24	-7.11	<.001	-.23	-.23	-.23
	RES	-.46	-15.08	<.001	-.46	-.46	-.46
Step 2	DIF	-.59	-15.36	<.001	-.23	-.46	-.45
	ASA	-.56	-14.58	<.001	-.18	-.44	-.43

The zero-order associations between the methods of bias score calculation and two outcomes were examined in the present study. The DIF method demonstrated the stronger relationship with aggression and the RES method demonstrated the stronger relationship with depression, corroborating the notion that the correlates of perceptual bias differ depending on which method is used to calculate the bias score. These results were to be expected based on the fact that the DIF bias score is always correlated equally with the component variables (i.e., ASA and PSA) while the RES bias score will always be strongly correlated with PSA and not at all related to ASA. In this study, aggression was more strongly correlated with children's ASA than their PSA, thus increasing the likelihood of finding a relationship between bias and aggression when using the DIF method. The opposite pattern was true when examining depressive symptoms. PSA was more strongly correlated with depressive symptoms than was ASA, making investigators more likely to find a relationship between bias and depression when using the RES method. The relationships between the bias score component variables and the outcome of interest prove to be an important consideration when interpreting perceptual bias study results.

One limitation to the present study concerns the influence that intra-rater bias may have had on the associations between the bias score component variables and the outcome variables. Specifically, children provided ratings of PSA and depression, whereas peers provided ratings of ASA and aggression. This shared method variance serves to increase the likelihood that PSA will be more strongly related to depression and ASA will be more strongly related to aggression. To the author's knowledge, no existing study that examines the relationship between biased self-perceptions of social acceptance and aggression or depression is free from this concern. As previously mentioned, there are studies that demonstrated that children's self-perceptions systematically differed from informants' perceptions of those children regardless of the informants' relationship to the children (i.e., peer, teacher, parent, test scores, etc.) and therefore, the method used to collect those ratings (Owens and Hoza 2003; Swanson et al. 2012). These studies supported the notion that the associations between the bias score component variables and the outcome variables in the present study were not entirely due to intra-rater bias.

The present study utilized a self-report measure of depressive symptoms because these symptoms are internalizing by their very nature. Thus, an outside observer would be a less valid rater of a child's own depressive symptoms. Any study examining the association between biased self-perceptions and depressive symptoms will have a difficult time circumventing the problem of intra-rater bias. Studies wishing to examine the relationship between biased self-perceptions and aggression, however, have a somewhat easier route. Future research examining biased self-perceptions of social acceptance and aggression should aim to replicate this well-

documented relationship using methods that do not allow for intra-rater bias. One example would be to obtain a self-report of PSA, a peer report of ASA, and a teacher report of aggression. Further, this strong research design should also be employed in future studies that investigate biased self-perceptions of behavioral competence or make use of outcomes other than aggression (e.g., social adjustment).

In the childhood peer-relationship literature it is routinely demonstrated that ASA is uniquely associated with a multitude of socio-emotional outcomes, including aggression and depressive symptoms. As such, ASA may be viewed as a potential confounding variable when considering the relationship between bias and socio-emotional outcomes and should, therefore, be controlled in the prediction model. When the DIF method was used in this study the respective strengths of the relationships between bias and the two outcomes were notably different when ASA was controlled. Once again, the associations between the bias score component variables and the outcomes of interest determined whether the relationships between bias and the outcomes were strengthened or weakened as a result of controlling for ASA. The strength of the relationship between bias and aggression was reduced once ASA was controlled because it had been the ASA component of the DIF bias score that was driving the relationship between bias and aggression. The opposite occurred when using the DIF score to examine the relationship between bias and depression. This relationship was actually strengthened by controlling for ASA because it was PSA that was driving the association between bias and depression. The mathematical properties of the DIF bias score ensure that the relationship between bias and an outcome of interest will change when a child's ASA is controlled in the analysis (assuming ASA is related to the outcome variable). The RES bias score is not subject to the same criticism because this calculation method controls for the ASA component variable during creation of the RES bias score. The RES bias score is already acting as a semi-partial correlate of the outcome variable it predicts. Thus, the DIF method will have nearly the same predictive potential as the RES method when ASA is considered as a covariate in during data analysis. Taken together, the results of the present study indicate that researchers should not expect concordant findings across methods. This method inconsistency can be remedied, however, by controlling for the ASA component variable when using the DIF bias score. If it were to become routine for investigators to control for ASA, results could be more confidently compared across studies that employ different methods of bias score calculation.

The mathematical certainty that the association between a DIF bias score and an outcome variable will be strongly influenced by its component variables has important implications for psychological phenomena that have been founded on use of the DIF method. The AD/HD literature, in particular, has exclusively used the DIF method to suggest that children

with AD/HD tend to hold exceptionally unrealistic positively biased self-perceptions compared to control children (Scholtens et al. 2012; Swanson et al. 2012; Owens et al. 2007). This finding has been so well-supported using the DIF method that it is regarded as a psychological truth in the AD/HD literature. This is concerning given the mounting evidence that DIF scores are often driven by associations between the component variables and the outcome of interest. With respect to the AD/HD literature, there is a well-documented negative correlation between social acceptance and AD/HD symptoms (Melnick and Hinshaw 1996; Greene et al. 2001). Thus, researchers and clinicians should be concerned that it is the relationship between social acceptance and AD/HD that is causing children with AD/HD to appear more positively biased in their self-perceptions than control children. Further, the research community may want to question whether the reason children with AD/HD appear to be the most positively biased about their area of greatest competence deficit (e.g., behavioral, academic, social, athletic, etc.; Hoza et al. 2002, 2004) is largely driven by their actual acceptance scores. Future research should examine the positive illusory biases of children with AD/HD using the RES method to confirm findings that have resulted from the DIF method. Alternatively, future research may choose to control for the actual acceptance component of the DIF bias score; a strategy that would have the same effect as using the RES method.

Recently, the use and interpretation of results gathered from the use of the DIF scores have been called into question. Laird and Weems (2011) demonstrated that models which use DIF to predict an outcome are equivalent to models which use the two components of that difference score to predict the same outcome. It was shown that despite model equivalence, the output information from the DIF model can often lead to an interpretation of the results that is inconsistent with the output information from the model that utilized the separate component variables. The authors called for careful interpretation of the data considering the associations between the component variables and the outcome variable. This recommendation is well-aligned with the empirical demonstration conducted in the present study in which it was important to consider whether it was ASA or PSA that carried most of the weight in the relationship between DIF and the outcome. More recently, Laird and De Los Reyes (2013) discouraged the use of DIF scores entirely due to two serious problems. First, any correlation computed between DIF and an outcome variable is fully determined by the variances of the two component variables and the correlations of each component variable with the outcome variable. If the component variables have equal variances and equal correlations with the outcome variable, DIF cannot mathematically be significantly associated with that outcome variable. Further, testing a directional DIF score is assuming that one of the component variables will be positively associated with the outcome and the other

component variable will be negatively associated with the outcome. This is rarely the case with biased self-perceptions because even when ASA and PSA disagree, they still correlate positively. The authors proposed testing interaction terms in polynomial regression analyses as a more valid method for examining discrepant reports. The RES score is not subject to the concerns raised by Laird and Weems (2011) and Laird and De Los Reyes (2013) because the ASA component is partialled out of the RES score and, therefore, needs no further consideration when interpreting results.

The RES method is not without criticism. Opponents of the RES method point to the fact that the RES score is often correlated nearly perfectly with the PSA component variable, and thus, is no more informative than using PSA as the sole predictor of an outcome. But, the correlation between PSA and RES depends strongly on the relationship between PSA and ASA. If PSA and ASA are not highly correlated then it is quite true that PSA will yield similar results as RES. But, if PSA and ASA are strongly associated then PSA may yield very different estimates than RES. Utilization of the RES method across all perceptual bias studies would be one way to consistently account for the relationship between the component variables in every sample.

Some may question why a researcher would go to the trouble of calculating a bias score when using the two component variables of that bias score in a regression model may provide similar results. As illustrated by this study, there are two reasons why the RES bias score may hold an advantage to a separate score regression model. First, the RES score has a strong theoretical alignment with the concept of *biased* self-perceptions whereas the PSA component variable merely represents self-perceptions. The RES score represents self-perceptions that are either more or less positive than would be expected based on informants' ratings of the child's actual acceptance. Secondly, the RES bias score can be used flexibly for the identification of participants as holding particularly biased self-perceptions. The use of ASA and PSA separately to identify children as having problematic self-perceptions would be more cumbersome and it would require multiple arbitrary cut points.

Given that the distributions of variables, ranges of variability, and intercorrelations among variables will be different in every sample, the exact findings from this study are not expected to generalize to other samples. What is expected to generalize, however, is that the DIF and RES methods will always capture a slightly different sub-sample of participants as holding positively (versus negatively) biased self-perceptions of competence in any given domain. Further, the DIF bias score (versus the RES bias score) will always be a stronger predictor of outcomes that are more heavily influenced by the actual competence/ASA component (versus the perceived competence/PSA component) of the bias score. When

the outcome variable is more strongly influenced by the perceived competence/PSA component the opposite will be true and the RES bias score will result in a stronger relationship with that outcome. One way to bring consistency to the field of perceptual bias research would be to control for the actual competence/ASA component variable by either adding it as a covariate in analyses with the DIF score, or simply, by using the RES score.

Conflict of Interest Haley F. Stephens, Janet A. Kistner, and Rebecca J. Lynch declare that they have no conflict of interest.

Experiment Participants All procedures performed in this study involving human participants were in accordance with ethical standards of the institutional review board at Florida State University and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from the parents of all individual participants (children) included in the study.

References

- Asher, S. R., Singleton, L. C., Tinsley, B. R., & Hymel, S. (1979). A reliable sociometric measure for preschool children. *Developmental Psychology, 15*(4), 443–444.
- Campbell, J. D., & Fehr, B. (1990). Self-esteem and perceptions of conveyed impressions: is negative affectivity associated with greater realism? *Journal of Personality and Social Psychology, 58*(1), 122–133.
- Carlson, G. A., & Cantwell, D. P. (1979). A survey of depressive symptoms in a child and adolescent psychiatric population. *Journal of the American Academy of Child Psychiatry, 18*, 587–599.
- Coie, J. D., Lochman, J. E., Terry, R., & Hyman, C. (1992). Predicting early adolescent disorder from childhood aggression and peer rejection. *Journal of Consulting and Clinical Psychology, 60*(5), 783–792.
- Cole, D. A., Jacquez, F. M., & Maschman, T. L. (2001). Social origins of depressive cognitions: a longitudinal study of self-perceived competence in children. *Cognitive Therapy and Research, 25*(4), 377–395.
- Crick, N. R., & Grotpeter, J. K. (1995). Relational aggression, gender, and social-psychological adjustment. *Child Development, 66*(3), 710–722.
- Crick, N. R., Bigbee, M. A., & Howes, C. (1996). Gender differences in children's normative beliefs about aggression: how do I hurt thee? Let me count the ways. *Child Development, 67*(3), 1003–1014.
- Cronbach, L. J. (1958). Proposals leading to analytic treatment of social perception scores. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior* (pp. 353–379). Stanford: Stanford University Press.
- David, C. F., & Kistner, J. A. (2000). Do positive self-perceptions have a “dark side”? Examination of the link between perceptual bias and aggression. *Journal of Abnormal Child Psychology: An Official Publication of the International Society for Research in Child and Adolescent Psychopathology, 28*(4), 327–337.
- De Los Reyes, A., & Kazdin, A. E. (2004). Measuring informant discrepancies in clinical child research. *Psychological Assessment, 16*, 330–334.
- Diamantopoulou, S., Rydell, A., & Henricsson, L. (2008). Can both low and high self-esteem be related to aggression in children? *Social Development, 17*(3), 682–698.
- Greene, R. W., Biederman, J., Faraone, S. V., Monuteaux, M. C., Mick, E., DuPre, E. P., Fine, C. S., & Goring, J. G. (2001). Social impairment in girls with ADHD: patterns, gender comparisons, and correlates. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*(6), 704–710.
- Griffin, D., Murray, S., & Gonzalez, R. (1999). Difference score correlations in relationship research: a conceptual primer. *Personal Relationships, 6*(4), 505–518.
- Harter, S. (1985). Manual for the Self-Perception Profile for Children. Unpublished manuscript, University of Denver, CO.
- Hartup, W. (1983). Peer relations. In E. M. Hetherington (Ed.) & P. H. Mussen (Series Ed.), *Handbook of child psychology (Vol. 4): Socialization, personality, and social development* (pp. 103–196). New York: Wiley.
- Hoza, B., Waschbusch, D. A., Pelham, W. E., Molina, B. S. G., & Milich, R. (2000). Attention-deficit/hyperactivity disorder and control boys' responses to social success and failure. *Child Development, 71*(2), 432–446.
- Hoza, B., Pelham, W. E., Dobbs, J., Owens, J. S., & Pillow, D. R. (2002). Do boys with attention-deficit/hyperactivity disorder have positive illusory self-concepts? *Journal of Abnormal Psychology, 111*(2), 268–278.
- Hoza, B., Gerdes, A. C., Hinshaw, S. P., Arnold, L. E., Pelham, W. E., Molina, B. S. G., & Wigal, T. (2004). Self-perceptions of competence in children with ADHD and comparison children. *Journal of Consulting and Clinical Psychology, 72*(3), 382–391.
- Hymel, S., Rubin, K. H., Rowden, L., & Lemare, L. (1990). Children's peer relationships: longitudinal prediction of internalizing and externalizing problems from middle to late childhood. *Child Development, 61*(6), 2004–2021.
- Kovacs, M. (1992). *The children's depression inventory manual*. Toronto: MultiHealth Systems.
- Kupersmidt, J. B., & Patterson, C. J. (1991). Childhood peer rejection, aggression, withdrawal, and perceived competence as predictors of self-reported behavior problems in preadolescence. *Journal of Abnormal Child Psychology: An Official Publication of the International Society for Research in Child and Adolescent Psychopathology, 19*(4), 427–449.
- Laird, R. D., & Weems, C. F. (2011). The equivalence of regression models using difference scores for each informant: implications for the study of informant discrepancies. *Psychological Assessment, 23*(2), 388–397.
- Laird, R. D., & De Los Reyes, A. (2013). Testing informant discrepancies as predictors of early adolescent psychopathology: why difference scores cannot tell you what you want to know and how polynomial regression may. *Journal of Abnormal Child Psychology, 41*(1), 1–14.
- Melnick, S. M., & Hinshaw, S. P. (1996). What they want and what they get: the social goals of boys with ADHD and comparison boys. *Journal of Abnormal Child Psychology, 24*(2), 169–185.
- Oden, S., & Asher, S. R. (1977). Coaching children in social skills for friendship making. *Child Development, 48*(2), 495–506.
- Orobio de Castro, B., Brendgen, M., Van Boxtel, H., Vitaro, F., & Schaeppers, L. (2007). “Accept me, or else”: disputed overestimation of social competence predicts increases in proactive aggression. *Journal of Abnormal Child Psychology: An Official Publication of the International Society for Research in Child and Adolescent Psychopathology, 35*(2), 165–178.
- Owens, J. S., & Hoza, B. (2003). The role of inattention and hyperactivity/impulsivity in the positive illusory bias. *Journal of Consulting and Clinical Psychology, 71*(4), 680–691.
- Owens, J. S., Goldfine, M. E., Evangelista, N. M., Hoza, B., & Kaiser, N. M. (2007). A critical review of self-perceptions and the positive illusory bias in children with ADHD. *Clinical Child and Family Psychology Review, 10*(4), 335–351.

- Pope, A. W., Bierman, K. L., & Mumma, G. H. (1991). Aggression, hyperactivity, and inattention-immaturity: behavior dimensions associated with peer rejection in elementary school boys. *Developmental Psychology, 27*(4), 663–671.
- Rudolph, K. D., & Clark, A. G. (2001). Conceptions of relationships in children with depressive and aggressive symptoms: social-cognitive distortion or reality? *Journal of Abnormal Child Psychology, 29*(1), 41–56.
- Scholtens, S., Diamantopoulou, S., Tillman, C. M., & Rydell, A. (2012). Effects of symptoms of ADHD, ODD, and cognitive functioning on social acceptance and the positive illusory bias in children. *Journal of Attention Disorders, 16*(8), 685–696.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 112*, 485–499.
- Swanson, E. N., Owens, E. B., & Hinshaw, S. P. (2012). Is the positive illusory bias illusory? Examining discrepant self-perceptions of competence in girls with ADHD. *Journal of Abnormal Child Psychology, 40*(6), 987–998.
- White, B. A., & Kistner, J. A. (2011). Biased self-perceptions, peer rejection, and aggression in children. *Journal of Abnormal Child Psychology, 39*(5), 645–656.
- Whitton, S. W., Larson, J. J., & Hauser, S. T. (2008). Depressive symptoms and bias in perceived social competence among young adults. *Journal of Clinical Psychology, 64*(7), 791–805.
- Zimmer-Gembeck, M. J., Hunter, T. A., & Pronk, R. (2007). A model of behaviors, peer relations and depression: perceived social acceptance as a mediator and the divergence of perceptions. *Journal of Social and Clinical Psychology, 26*(3), 273–302.
- Zuckerman, M., & Knee, C. R. (1996). The relation between overly positive self-evaluation and adjustment: a comment on colvin, block, and funder (1995). *Journal of Personality and Social Psychology, 70*(6), 1250–1251.