Thin-Slice Judgments of Children’s Social Status and Behavior

Tessa A. M. Lansu & Yvonne H. M. van den Berg

To cite this article: Tessa A. M. Lansu & Yvonne H. M. van den Berg (2020): Thin-Slice Judgments of Children’s Social Status and Behavior, The Journal of Experimental Education, DOI: 10.1080/00220973.2020.1808943

To link to this article: https://doi.org/10.1080/00220973.2020.1808943

© 2020 The Author(s). Published with license by Taylor and Francis Group, LLC

Published online: 25 Aug 2020.
**Thin-Slice Judgments of Children’s Social Status and Behavior**

Tessa A. M. Lansu and Yvonne H. M. van den Berg

Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands

**ABSTRACT**

The moment a child walks into a new classroom, teachers and classmates form an impression based on minimal information. Yet, little is known about the accuracy of such impressions when it concerns children’s social functioning at school. The current study examined the accuracy of children’s, teachers’ and adults’ impressions of 18 unacquainted children based on thin slices of behavior. The likeability, popularity, prosocial behavior, aggression, and exclusion of these children were judged by 101 children, 79 elementary school teachers, and 68 young adults based on 20-second video clips. Judges were better than chance in predicting popularity and prosocial behavior, but worse than chance in predicting aggression and exclusion. Female judges were more accurate judging social exclusion of same-sex than other-sex targets. Teachers were more accurate than children in their judgments of prosocial behavior. The current study shows that confidence in one’s impression of aggression and exclusion in unacquainted children based on minimal information is not warranted.

**KEYWORDS**

Accuracy; children; social functioning; teachers; thin slices

The moment a child walks into a new classroom, teachers and classmates form an impression. However, it is unclear how accurate such impressions are. As these types of impressions are the basis of subsequent expectations and behavior toward that person (Ambady et al., 2000), inaccurate impressions may be the prelude of negative social interactions. The current study therefore examines the accuracy of children’s, teachers’, and adults’ impressions regarding a child’s social behavior and status in the classroom context.

**Impressions of Children**

Numerous studies have examined the reliability of the impressions people form about an unacquainted adult (Hall et al., 2016). Even when these impressions are based on observations of less than 5 minutes, so called ‘thin slices’, studies have shown that people’s initial impressions of a person’s personality, political preference, and sexual orientation are quite accurate. (see Ambady et al., 2000 for an overview). Yet, it remains unknown whether people – adults and age mates - can also accurately estimate children’s social behavior and status.

Previous studies regarding impressions of children predominantly focused on children’s academic behaviors. These studies showed that teachers quickly form impressions about hypothetical as well as their own students’ future academic behavior and performance (Ritts et al., 1992; Rubie-Davies, 2018). These impressions are based on a variety of characteristics, including...
students’ physical attractiveness, race, gender, working habits, and even characteristics of older siblings or parents (Ritts et al., 1992; Rubie-Davies, 2018; Timmermans et al., 2016). These initial perceptions of students’ academic performance have shown to be reasonably accurate in predicting students’ future performance (Jussim & Eccles, 1992; Jussim & Harber, 2005).

However, teachers’ inaccurate impressions of students’ academic performance are also known to be predictive of students’ future behavior. An example of the effects of teachers’ inaccurate impressions on students’ functioning, is the classic Pygmalion study by Rosenthal and Jacobson (1968), which showed that students randomly described as ‘bloomers’ to their teachers increased more in IQ than students not labeled as bloomers. Moreover, when ‘non-bloomers’ did show intellectual growth, teachers seemed to think more negatively about them. More recent studies also showed that teachers’ inaccurate expectations and first impressions can lead to self-fulfilling prophecies and result in poor(er) academic performance (Rubie-Davies, 2018; Sorhagen, 2013). Studies even show that teachers’ implicit prejudiced attitudes, reflecting automatic negative associations with an ethnic minority group, contribute to an achievement gap between ethnic minority and ethnic majority students on standardized tests (Peterson et al., 2016; van den Bergh et al., 2010). Together, these studies demonstrate that perceptions held by teachers can shape children’s actual academic behavior and functioning at school.

Yet, people’s responses to children being shaped by their impressions of them is likely not only confined to the academic domain. How people perceive children’s social characteristics (whether a child is expected to be aggressive, helpful, well embedded in the group, a leader) may just as well shape their interactions with them, and subsequently children’s opportunities to develop social skills and a good social position. Kindergarten teachers have been found to not only hold different beliefs and attitudes toward prosocial, withdrawn and aggressive children, they also intend to behave differently toward these children (Arbeau & Coplan, 2007). Similarly, children’s behavior toward another child seems to be affected by information about the social functioning of that child. Harris and colleagues (1992) showed that children’s manipulated expectancy as well as the actual diagnostic status of their partner’s problem behavior negatively affected children’s social interaction with the other child.

In addition to beliefs and expectations regarding children’s social behavior that are formed explicitly, early brief impressions and expectations may subtly influence how these children are approached too. Inaccurate impressions of a child, even when based on little information, may negatively bias people’s attention for behavioral cues, and cause them to respond more negatively to the child. A few studies have looked into the accuracy of thin slice judgments of children’s social characteristics. Thin slices are defined as brief excerpts of expressive behavior, sampled from the behavioral stream, that contain dynamic information and are less than 5 minutes long (Ambady et al., 2000; Ambady & Rule, 2007). One study showed that, based on thin slice video fragments, youth’s externalizing problems and temperament can be judged above chance levels by unacquainted observers (Tackett et al., 2017). Another study using photographs showed that children’s trustworthiness can be judged above chance levels by unacquainted adults and children as well (Li et al., 2019). Judgments of the children’s trustworthiness were mediated by the extent to which the children were well-liked among their classroom peers. This suggests that unacquainted observers’ judgments of trustworthiness in fact are related to children’s social status in the peer group. In the current study, we focus on the accuracy of judgments of children’s social functioning in their peer group.

**Social Characteristics of the Child being Judged**

In the current study, we will focus specifically on children’s social behavior and social status, as these are relevant characteristics that are being judged in everyday life on a daily basis (Rose-Krasnor, 1997). With regard to social behavior, we specifically focus on prosocial behavior (e.g.,
voluntary behavior intended to benefit others; Eisenberg, 1986) and aggression (e.g., behaviors intended to harm others; Card et al., 2008). Social status is conceptualized in terms of social preference and popularity (Cillessen & Marks, 2011; van den Berg and Lansu share first authorship of this paper & Cillessen, in press). Social preference indicates likeability and being highly preferred by the peer group, whereas popularity indicates children’s impact, visibility, and reputation in a group (Cillessen & Marks, 2011).

The extent to which positive or negative behavior is expected of a person affects children’s behavior toward that person. For example, when someone has features that make others judge them as trustworthy, children tend to show more prosocial behavior toward that person (Ewing et al., 2015). And when children expect another child to have behavioral problems, this negatively affects children’s social interaction with the other child (Harris et al., 1992). Similarly, those who are seen as having high popularity or likeability may be perceived and treated differently than those low(er) in status. Previous studies have indeed demonstrated this for both popularity and likeability. LaFontana and Cillessen (2001) showed that when it comes to popular hypothetical peers, children attributed more hostility and more stable negative attributions, than when judging hypothetical peers with neutral social status. And Hymel (1986) similarly showed that when it comes to disliked peers, children attributed more hostility, more stable negative attributions and more blame for children’s negative behavior, than when judging peers that were well-liked.

Moreover, characteristics that are more available and easier to detect are likely to be estimated with higher accuracy (Funder, 2012). We therefore expect higher accuracy for social behaviors than for status, as social behavior is directly observable whereas social status is more indirectly inferred. Based on this premise we also expect higher accuracy in judging positive social behavior than negative social behavior, as prosocial behavior occurs more frequently than aggressive behavior (Ladd & Profilet, 1996).

In addition, the characteristics of a target can only be estimated accurately when observable and relevant cues of these characteristics are present in the ‘thin slices’ (i.e., realistic accuracy model; Funder, 2012). In the current study, thin slices were therefore extracted from a videotaped cooperative interaction between classmates. This cooperative task gave the videotaped child the chance to display both positive and negative social behavior and exert influence and dominance. Moreover, it is an interaction with an actual classroom peer, making it likely that a child’s classroom behavioral profile and status position is reflected in the behavior shown (Lansu & Cillessen, 2015).

Finally, we took the sex of the child being observed into account. Research has shown that in men, sadness, angry facial expressions are more easily recognized (Bijlstra et al., 2010), whereas in women positive emotions and body postures are more easily recognized (Bijlstra et al., 2010; Bijlstra et al., 2019). It is therefore possible that judges are more likely to see negative behavior in boys, and positive social behavior in girls, affecting the accuracy of their judgment. Boys might be more difficult to read, as men have been shown to suppress their emotions more than women (Gross & John, 2003). Moreover, the dyadic setting in which children were being filmed may have been more comfortable for girls as they have more experience with such dyadic situations (Hops et al., 1997; Leaper, 1991). This could lead to a more naturalistic behavioral pattern in girls than in boys, making it easier to accurately judge girls’ than boys’ behavior and status among peers. We therefore expect girl targets to be judged more accurately than boy targets.

**Characteristics of the Judge**

Accuracy of first impressions also depends on characteristics of the judge. The realistic accuracy model specifies judgmental ability of the judge as a moderator of judgment accuracy (Funder, 2012). One could think of judgment ability as the degree to which a judge is familiar with or knowledgeable about the group being judged. Research for example has demonstrated that
cultural familiarity facilitates emotion recognition accuracy (Elfenbein & Ambady, 2003), and also for detecting homosexual sexual orientation (Rule et al., 2007), and Mormon religious beliefs (Rule et al., 2010), in-group advantages in judgment accuracy have been found. In the current study, familiarity with the social dynamics in children’s peer groups might facilitate more accurate judgment. For child judges, close resemblance of the social world of the judge with the person judged probably increases sensitivity to important social cues, making them able to form reliable impressions. Teachers may not be directly part of the social dynamic of the peer group, but they are exposed to these dynamics on a daily basis, and are trained to attend to and understand children’s social functioning. Young adults who do not work with groups of children are probably the least accurate, as they are not likely to be exposed to the social dynamic in children’s peer groups. We therefore expected children to be most accurate, followed by teachers, followed by adults.

Sex is another characteristic related to judge’s ability to make accurate estimations, as it is known to play a role in picking up social cues and accuracy of social judgment. In judgments of emotion, women are more accurate than men (Hall, 1985), even when they see the targets only for a very brief moment (Hall & Matsumoto, 2004). This higher accuracy in females has also been found in child and adolescent judges (Lawrence et al., 2015), and has especially been found when cues were less informative (i.e. had lower emotional intensity; Hoffmann et al., 2010). Moreover, women tend to judge neuroticism more accurately than men when basing their judgment on thin slices (Lippa & Dietz, 2000). Given these findings, we expected female judges to be more accurate than male judges.

Finally, we examined the interaction between the judge’s sex and the sex of the child being observed. We expected judges to rate targets of their own sex more accurately, because of more personal experience with interactions with people of their own sex than of the other sex (Mehta & Strough, 2009). Through this difference in experience, people may have developed better sensitivity for sex specific cues concerning social functioning for same-sex than for other-sex targets.

**Method**

**Phase 1: Stimulus Selection and Criterion Variables**

**Stimulus Selection**

To select the video stimuli for the current study, we used the video recordings of 18 Dutch children (9 boys, 9 girls, $M_{age} = 10.98$ years, SD = 0.74 years), with equal distribution of popularity levels within each gender. The videos were previously collected in a larger study on children’s social relations and dyadic interactions (Lansu & Cillessen, 2015). Two same-sex classmates were asked to plan a classroom party, for which they had to decide on the day and time of the party, the activities that would take place, and the drinks and snacks that would be available. Each couple had 10 minutes to complete this task while sitting in front of a laptop with a webcam recording the interaction.

For the current study, the videos were edited in such a way that only one child from a given dyad was visible on screen. For each target, we selected the first 10 seconds in response to the question “When is the party held?” and the first 10 seconds in response to the question “What are we going to do at the party?”. By combining these two 10-second clips, one 20-second video clip for each target was created. Videos with a duration of 20 seconds have been successfully used in thin slices research before (Ambady & Rosenthal, 1992). In addition, short videos were also chosen to keep particularly the child judges motivated to attend to all 18 clips.
Criterion Variables
Targets’ actual social status and behavior in the classroom were measured using peer nominations (Cillessen & Marks, 2011). Each child within the classroom of the target was asked to nominate classmates they liked most, liked least, who were most popular, and who were least popular. Three items were used to assess social behaviors, including prosocial behavior (“who helps others often”), aggression (“who argues a lot”), and social exclusion (“who is being excluded by others”). They could nominate as many or as few classmates as they wanted for each question, allowing both same-sex and other-sex choices. The number of nominations received was counted for each question and standardized to z-scores within classrooms. Targets’ scores for prosocial behavior, aggression, and exclusion were used as criterion variables. A score for likeability among classmates was computed as the difference between the standardized liked most and liked least scores, which was standardized again within classrooms. A score for popularity was computed as the difference between standardized most popular and least popular scores, which was standardized again within classrooms (Cillessen & Marks, 2011). All of these peer nomination items are regularly used in research on peer relations, and have proven to have good validity and reliability (van den Berg & Cillessen, 2013).

Phase 2: Measuring First Impressions
Participants
The videos were presented to 101 children (M_age = 10.7 years, 47.5% female, 96.0% Dutch ethnicity), 79 elementary school teachers (M_age = 39.5 years, 84.8% female, 82.3% Dutch ethnicity), and 68 young adults (M_age = 22.1 years, 82.4% female, 69.1% Dutch ethnicity) who judged the video clip of each target. All judges were unacquainted with the targets presented in the video clips.

Procedure
Children and teachers were recruited with a letter to schools explaining the project. Parental informed consent was obtained for all children. The young adults were recruited from the human subject pool of the university and they received course credits. The children and teachers were visited at school. Children were taken from the classroom in pairs. They each watched the videos on their own computer while wearing headphones. After each video, they completed the thin-slice judgments. Teachers completed the judgment task after school time. The young adults came to the laboratory at the university and completed the thin slice judgment task in a cubicle. All participants self-reported their ethnic background and age before completing the thin slice judgment task. All procedures have been reviewed and approved by the IRB of the authors’ host institute (ECG2012-2402-0020).

Thin Slice Judgments
After watching each 20-second video on a laptop, the judges rated on that same laptop each target on the five criterion variables (e.g. social status and behavior). For each target they were asked: Do you think this child 1) is liked among her classmates? (likeability), 2) is popular among her classmates? (popularity), 3) helps her classmates often? (prosocial), 4) often argues with classmates? (agression), 5) is being excluded by her classmates? (social exclusion). A 7-point rating scale was used ranging from 1 (not probable) to 7 (very probable).
Table 1. Mean accuracy for each construct.

<table>
<thead>
<tr>
<th></th>
<th>Sex Target</th>
<th>Sex Judge</th>
<th>Familiarity Judge</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Boy</td>
<td>Girl</td>
</tr>
<tr>
<td>Likeability</td>
<td>−.00</td>
<td>.02</td>
<td>.00</td>
</tr>
<tr>
<td>Popularity</td>
<td>.16*</td>
<td>.18*</td>
<td>.29*</td>
</tr>
<tr>
<td>Prosocial behavior</td>
<td>.21*</td>
<td>.11*</td>
<td>.07*</td>
</tr>
<tr>
<td>Aggression</td>
<td>−.14*</td>
<td>−.05*</td>
<td>−.05†</td>
</tr>
<tr>
<td>Social exclusion</td>
<td>−.09*</td>
<td>−.16*</td>
<td>−.09*</td>
</tr>
</tbody>
</table>

Note: Mean accuracy values were calculated using Fisher-z transformed accuracy correlations and were transformed back into $r$ for presentation. One-sample $t$-tests testing correlations against zero were performed on the Fisher-z transformed accuracy scores, *$p < .05$. Means with different subscripts were significantly different between groups.

**Accuracy**

A trait-based approach was used to calculate the accuracy scores (Back & Nestler, 2016; Vogt & Colvin, 2003). For each criterion, the judge’s first impression was compared with targets’ actual score. For instance, each judge’s rating of targets’ popularity was correlated with the targets’ actual popularity based on the peer nomination score obtained in the target’s classroom, subsequently averaging these correlations across targets. This correlation coefficient was used as the judge’s individual accuracy score for this criterion; higher scores indicated more agreement between the judge’s first impression and the targets’ actual status or behavior measured through peer nominations. Given a non-normal distribution of the profile correlations, the correlation coefficients were transformed into Fisher’s-$z$ scores before running the analyses and converted back into $r$ for presentation.

**Results**

**Accuracy Across Judges**

A repeated measures ANOVA with criterion (likeability, popularity, prosocial behavior, aggression, exclusion) as within subjects factor on the accuracy scores was conducted. As Mauchly’s Test of Sphericity indicated that the assumption of sphericity had been violated, $\chi^2(2) = 48.58, p < .001$, we applied a Greenhouse-Geisser correction to within subjects tests and demonstrated a significant difference between criteria, $F(3.63,883.15) = 124.39, p < .001, \eta_p^2 = .34$. Post hoc analyses employing a Bonferroni correction revealed that all criteria differed significantly from each other. Table 1 shows the accuracy of each criterion. One sample $t$-tests showed that judges were significantly better than chance in predicting targets’ popularity $t(245) = 12.60, p < .001$ and prosocial behavior $t(246) = 13.58, p < .001$. However, they were significantly worse than chance in predicting aggression $t(245) = −9.39, p < .001$ and social exclusion $t(246) = −6.74, p < .001$. For likeability, judges were at chance level $t(244) = −0.25, p = .803$.

To test whether accuracy varied by target’s sex, a within subjects ANOVA was conducted with criterion (likeability, popularity, prosocial behavior, aggression, exclusion) and target’s sex (boy, girl) as within subjects factors. There was no overall effect of target’s sex $F(1,944) = 2.26, p = .134, \eta_p^2 = .01$. However, a significant interaction was found between the criterion being judged and targets’ sex, $F(3.86,911.63) = 5.51, p < .001$ (Table 1). Post hoc paired-sample $t$-test revealed that judges were generally more accurate when judging popularity of girl targets compared to boy targets, $t(245) = −3.83, p < .001$, and less inaccurate when judging social exclusion of girls compared to boys, $t(243) = −2.50, p = .01$. Judge’s accuracy of likeability, prosocial behavior and aggression did not vary by target’s sex.

**Differences between Judges**

As shown in Figure 1, there was considerable variance in judges’ accuracy, indicating that some judges were more accurate than others in rating unacquainted children’s social behavior and
status. First, we examined whether judges being familiar with the social dynamics of children’s peer groups is related to their accuracy of judgment. A mixed ANOVA on the accuracy scores with familiarity of observer (children vs. teachers vs. young adults) as between subjects factor and criterion (likeability, popularity, prosocial behavior, aggression, exclusion) as within subjects factor showed no main effect of familiarity of observer, $F(2, 241) = 1.09, p = .337, \eta_p^2 = .01$. Yet, there was a significant interaction between criterion and familiarity of observer, $F(7.29, 877.92) = 2.15, p = .034, \eta_p^2 = .02$. Post hoc analyses showed that teachers were more accurate than children when judging prosocial behavior, $F(2,244) = 3.02, p = .051, \eta_p^2 = .02$ (see Table 1). Young adults’ accuracy did not differ from teachers’ or children’s accuracy. Children, teachers, and young adults were equally accurate when judging the other four constructs ($p$’s $>.05$).

Next, we examined whether accuracy varied by the judges’ sex. A mixed ANOVA on the accuracy scores with judge’s sex (male, female) as between subjects factor and criterion (likeability, popularity, prosocial behavior, aggression, exclusion) as within subjects factor showed that female judges were equally accurate as male judges in general, $F(1, 242) = 0.22, p = .643, \eta_p^2 = .00$, and that this did not differ across specific criteria, $F(3.64, 879.60) = 0.36, p = .816, \eta_p^2 = .00$.

Finally, for each of the criteria judged, a mixed ANOVA on the accuracy scores with observer’s sex (male, female) as between subjects factor and target sex (boy, girl) as a within subjects factor was run. A significant interaction between observer sex and target sex was found for social exclusion, $F(1, 242) = 4.70, p = .031, \eta_p^2 = .02$. Post hoc t-tests comparing the accuracy of the judgment of exclusion of boy targets and girl targets for male and female observers separately show that male observers are equally accurate in judging targets from both sexes ($M$’s are $0.11$ and $0.13$ respectively) $t(76) = 0.34, p = .74$. However, female observers are more accurate in judging same-sex targets’ exclusion ($M = -0.08$) than other-sex targets’ exclusion ($M = -0.20$), $t(166) = -3.53, p = .001$. No other effects of judge’s and target’s sex on accuracy were found for any of the other criteria.

**Discussion**

Numerous studies have addressed the degree to which first impressions of an adult target are accurate, even after a very short exposure to this target (Ambady et al., 2000). However, whether children’s behavior and social status can also be accurately estimated based on thin slices of information was unknown until now. The results demonstrate that people’s first impression of children’s social position and behavior ranged from worse to better than chance, and depended on the characteristic being judged, the target being judged, and the characteristics of the judge.
In line with our expectations, judges were more accurate than chance in estimating children’s prosocial behavior. It could be that observers picked up on the vagal activity of the targets, as Kogan and colleagues (2014) showed that people with low or very high levels of vagal activity are less likely to behave prosocially. The thin slices part of study by Kogan and colleagues suggests that people might pick up on how ‘tense’ the targets in thin slices are, and use this information to increase the accuracy of their prediction whether a target is likely to engage in prosocial behavior. This because vagal activity of the target was (quadratically) related to prosociality judgments by unknown judges looking at 20-sec clips of the targets.

Unexpectedly, in the current study, judges were also better than chance in estimating children’s level of popularity. Whereas we expected that popularity in itself would be difficult to observe, as it is imposed on children by their peers, there might be features associated with popularity that are more easily perceived. For example, very concrete features such as linguistic patterns (Labov, 2006), and clothes (Gillath et al., 2012), may signal a certain social status. In addition, research on judgments of social class suggests that when people engage in social interactions, they communicate their social status to observers through their behaviors and cultural practices (Kraus & Keltner, 2009; Kraus et al., 2017). A similar process could also apply to popularity and children’s behavior. Having a position of influence and power among classmates might create a certain behavioral pattern or way of responding to others. This pattern in turn then is recognized by others as being reflective of high popularity.

In the current study, judges were inaccurate in estimating children’s level of aggression and social exclusion. Whereas previous research has shown that people’s judgment of 2-sec displays of pictures of sex offenders’ faces corresponds with actual violent history of the offenders (Stillman et al., 2010), the levels of minor aggression assessed in the children in the current study might be less easy to perceive. And whereas men scoring high on psychopathy were able to accurately identify targets that had previously been victimized based on thin-slice clips (Wheeler et al., 2009), non-psychopathic judges were not able to correctly make such judgments when viewing thin slices clips of children interacting in the current study. Extreme aggressiveness on the side of the targets, or psychopathic traits on the side of the judges might be necessary prerequisites in order for aggression and victimization to be correctly identified. Moreover, as indicators of these behaviors may be less likely to be displayed during the interaction task and are less frequent to begin with (Ladd & Proffit, 1996) compared to other characteristics being estimated by our judges, it may be more difficult to correctly assess these characteristics. It is surprising though that the estimations in the current study are not just at chance level, but that judges are even significantly incorrect in estimating these characteristics in children (e.g. estimating those with relatively high aggression scores to be low in aggression and vice versa). People thus should be careful in drawing conclusions about children’s functioning with regard to aggression and being socially excluded after first impressions based on minimal information, as these are likely to be inaccurate.

**Characteristics of the Judge and Child being Judged**

Judges’ familiarity with the targets’ social world did not make a difference for the accuracy of judgment, except for impressions of prosocial behaviors. Whereas all three groups were accurate, teachers were a bit more accurate in their estimations of prosocial behavior than child judges. As attention for teachers promoting prosocial behavior in their students is increasing (Kidron & Fleischman, 2006), teachers may engage regularly in observing and analyzing prosocial behavior in their classrooms. Teachers’ increased attention for prosocial behavior in their own classrooms thus may have led to higher accuracy when judging prosocial behavior.

For all other characteristics, the three types of judges were quite similar in their levels of accuracy, suggesting that familiarity with the targets’ social world does not affect thin slices
judgment to a large extent. These first impressions may be based on very basic and universal
cues in the videos, for which you do not need a thorough understanding of the specific variations
in social dynamics applicable to children’s classrooms. Teachers overall not being more accurate
in judging children’s characteristics is in line with the results found by Praetorius and colleagues
(2015), who showed that unacquainted persons’ judgments of students’ academic self-concepts
based on thin sliced information were as accurate as these students’ own teacher’s judgment.

Furthermore, we examined whether accuracy varied by sex. In contrast to previous studies
among adults (Hall, 1985; Hall & Matsumoto, 2004), in the current study female and male judges
were equally accurate. It must be noted though that male judges were underrepresented in the
sample of young adult judges and teacher judges, making it less likely that sex differences in
judgement accuracy could be detected in the current analyses. However, sex was shown to play a
role at the target level, as girl targets were ‘easier to read’ when it concerned popularity and social
exclusion. A potential explanation is that showing assertiveness and dominance may be seen as
less desirable for girls than boys (Sebanc et al., 2003), and that display of such dominant behav-
iors by popular girls is thus more easily noticed in girls. Another explanation could lay in the dis-
play of relational aggression during the task that the targets were engaging in during the video:
planning a party. As relational aggression is more strongly associated with popularity in girls
than boys (Rose et al., 2004), popular girls may have used a strategy of subtle verbal meanness
more often during the task, which was picked up by the judges as an indicator of high popular-
ity status.

The finding that judges were more accurate (c.q. less inaccurate) when judging social exclusion
among girls than boys may be explained by victimization being differentially perpetuated across
contexts for both genders, as a continuation of victimization in a dyadic setting has been found
for girls but not boys (Lansu et al., 2014). This suggests that group reported victimization has a
larger role in the dyadic context for girls than for boys, and thus may be easier to identify in
such a context. Another explanation could lie in the behavior profile of boys and girls who are
socially excluded (Gazelle, 2008). Socially rejected or solitary girls show a more uniform behav-
ioral pattern, whereas rejected boys’ behavioral profile is ‘clouded’ by externalizing behavior to
larger extent (Gazelle, 2008). As a result, the behavioral pattern of girls may be more in line with
judges’ expectations about the behavior of socially excluded children, leading to higher accuracy.
The interaction between target’s sex and judge’s sex for social exclusion showed that this
increased accuracy in judging girl targets applies to female, but not to male judges. It thus might
be the case that for judgments of social exclusion, familiarity with the observed situation (female
judges being more familiar with what social exclusion in female dyads looks like) is a requirement
for higher accuracy.

Limitations and Future Directions

Contextual Differences

The current study demonstrates that some indicators of children’s social functioning in the class-
room (popularity, prosocial behavior) can be accurately estimated based on thin slices of behav-
ior, whereas judgments of other types of behavior (aggression and social exclusion) are below
chance levels. This raises the question why first impressions are (in)accurate. Accuracy could
depend upon the amount of information, as well as the type of situation in which the targets
were recorded and the frequency with which certain behaviors are subsequently displayed (Back
& Nestler, 2016). Maybe judgments of children’s social characteristics would become more accur-
ate if judges are presented with longer video fragments. It would be interesting to examine in
future research if and when (at 40 sec, 1 minute, 2 minutes or 5 minutes) judges become more
accurate in estimating children’s social status and behavioral features.
Moreover, if children would be recorded during a more competitive or threatening situation where there is scarcity or where they face negative feedback, the accuracy of judgments of aggression and social exclusion might be higher. In these situations, children who tend to be aggressive in their classroom might more easily turn to using coercive strategies during the recording (DeRosier et al., 1994), and children who in their classroom tend to be excluded may be ignored by a classmate during the recording, thereby making it more likely for observers to pick up relevant cues.

In addition, which interaction partner the child has may be another factor impacting the results. The video fragments showed only one child, however, the child was interacting with another child. Features of this partner, such as the partner’s social status, are likely to affect the behavior the child displays, and thus the signals judges can use for their thin slice impressions. Examination of the complete set of full-length video recordings from which the current thin slices were drawn has in fact shown that a partner’s popularity had an effect on the child’s behavior (Lansu & Cillessen, 2015). For both boys and girls, the partner’s popularity negatively predicted the child’s use of coercive resource control strategies and negative behavior in the dyad. For girls, in addition, the partner’s popularity also positively predicted their submissiveness and negatively predicted their task influence. Levels of negative behavior and coercive resource control being affected by the partner’s popularity may have interfered with the ‘natural’ levels of aggression signals being sent by the target child. For girls, the levels of submissive behavior being affected by the partner’s popularity may have interfered with social exclusion signals being sent by the target child. Hence, the effect of interaction partner’s popularity on the targets’ behavior may explain the lower accuracy of thin slice judgments of aggression and social exclusion. To minimize the effects of partner’s popularity, future research examining thin slices of children’s social functioning should keep the variability in social status of the target’s interaction partner as low as possible when selecting thin slices.

Differential Use of Cues
Accuracy of first impression may also depend on the cues people use to form their first impressions (Back & Nestler, 2016). Do people pay most attention to specific areas of the face (Langner et al., 2009), emotional expression (Oosterhof & Todorov, 2009) or ease of interaction? One could measure where people direct their attention to while looking at the thin slices using eye tracking (Klin et al., 2002). One could also ask people about the cues they focused on and how they interpreted them in coming to their judgment. Information regarding both judges’ attention and interpretation may help us understand why certain people are more accurate than others. It could be that accurate judges focus on different types of cues or make a different interpretation of the same cues compared to less accurate judges.

Methodological Choices
Finally, accuracy may be affected by study characteristics. In the current study, all targets were White and born in the Netherlands, just like the majority of the judges. Ethnic variation in targets and judges may affect judges’ accuracy, and the current results thus may not generalize to ethnically more heterogeneous samples. Results may also be affected by how target characteristics were measured. In the current study, social status and behaviors were assessed among children’s classmates using peer nominations. This gives us an indication of the relative score for the student compared to the scores of their classmates, but does not directly assess the absolute level of status or behavior. In addition, as these standardized scores were not on the same metric as the observer ratings (Z-scores vs. Likert scale), the accuracy scores reflect rank-order stability rather than absolute agreement between peer and observer ratings.
Conclusion

By using a well-established thin slices approach (Ambady et al., 2000), this study shows that confidence in one’s estimates of aggression and exclusion in unacquainted children based on minimal information is not warranted. It is important to inform people about this, as people do not always seem to have insight into the accuracy of their judgments of others (Dunning et al., 1990; Realo et al., 2003). Moreover, once an impression is formed, people become less open to taking into account objective information that is not in line with their beliefs, which may lead to a confirmation bias (Rabin & Schrag, 1999). Particularly teachers should be made aware that their impressions of aggression and exclusion, made with only minimal information at the start of a school year cannot be trusted.

An important first step is therefore to make teachers aware of their differential impressions about students’ performance and abilities. There are various studies indicating that systematic positive expectancy practices can indeed heighten teachers’ awareness of their (inaccurate) impressions, can successfully alter teachers’ confirmation bias, can prevent self-fulfilling prophecy effects, and ultimately improve students’ academic performance (Gottfredson et al., 1995; Rubie-Davies et al., 2015). Unfortunately, when implementing such programs, raising awareness does not always receive that much attention (de Boer et al., 2018). In teacher training and specific intervention programs, substantial attention should therefore be given to creating awareness and teachers should be encouraged to reconsider their first impressions. Although you never get a second chance to make a first impression, children should get a second chance before teachers make up their mind about whether specific children are likely to be aggressive or excluded.

Acknowledgments

The authors are grateful to the respondents, teachers, and school administrators who made this research possible. We also thank Susanne Doomernik, Ummu Alkan, and Ummit Sahin for their role in the recruitment and data collection.

References


