

A Two-Dimensional Ratings-Based Procedure for Sociometric Status Determination as an Alternative to the Asher and Dodge System

Gerard H. Maassen, *Utrecht University*

Karine Verschueren, *University of Leuven*

The present study compares the results of three two-dimensional procedures for sociometric status determination: the Asher and Dodge (1986) system, the rating scale method SSrat of Maassen, Akkermans, and Van der Linden (1996), and the nomination procedure of Coie and Dodge (1983). The study followed 170 children from third grade through sixth grade. Children were asked (a) to nominate the classmates with whom they most liked and least liked to play, and (b) to rate how much they liked to play with each of their classmates. The SSrat method determines sociometric status with higher validity and stability than do the other two methods. SSrat's methodological strengths and practical advantages as compared to the Asher and Dodge system demonstrate that it is a valuable, if not preferable, alternative for determining children's sociometric status in situations where the use of a rating scale is appropriate.

Sociometry, introduced by Moreno (1934) as a multifaceted theoretical model of sociometric judgment, soon developed into a set of practical tools to determine the popularity of individuals within their social group (Cillessen & Bukowski, 2000). Various instruments establishing the social preferences of the group members have been proposed. For example, one can ask respondents to nominate the group peers whom they feel most attracted to (positive nominations) and then take the number of times a person is nominated as an indication of his or her popularity. Or, alternatively, respondents can be asked to rate all their

Gerard H. Maassen, Faculty of Social Sciences, Department of Methodology and Statistics; Karine Verschueren, Department of Psychology.

Correspondence should be addressed to Gerard H. Maassen at the Department of Methodology and Statistics, Faculty of Social Sciences, Utrecht University, PO Box 80140, 3508 TC Utrecht, The Netherlands. E-mail: g.maassen@fss.uu.nl

Merrill-Palmer Quarterly, April 2005, Vol. 51, No. 2, pp. 192–212. Copyright © 2005 by Wayne State University Press, Detroit, MI 48201.

group peers on a scale; an individual's popularity may be determined by the average rating he or she receives.

From the 1950s on, scholars such as Dunnington (1957) and Thompson and Powell (1951) came to regard such a one-dimensional assessment of (un)popularity as unsatisfactory. If individuals are assigned to the rejected category solely on the basis of a low number of positive nominations, then this category may turn out to be heterogeneously populated. Individuals who receive very few positive nominations are persons to whom few people feel particularly attracted, but on the other hand, only a few of their group peers may actually dislike them. To distinguish between neglected individuals (persons who arouse neither positive nor negative affection in most of their peers) and actively rejected individuals, negative nominations should also be elicited from participants. Peery (1979) was the first to present a two-dimensional system for sociometric status determination. Building on his work, researchers in the early 1980s developed the methods currently used for classifying people on the basis of positive and negative nominations: the standard score model of Coie, Dodge, and Coppotelli (1982) (hereafter referred to as CDnom) and the probability model of Newcomb and Bukowski (1983). Both systems lead to a two-dimensional classification into five status groups: popular, rejected, neglected, controversial, and average. The popular, average, and rejected categories are positioned along the first and main dimension, usually referred to as *social preference*, while the neglected, average, and controversial categories are positioned along the second dimension, usually labeled *social impact*.

The nomination procedures have been applied in numerous studies, and most researchers apparently prefer the procedure of Coie et al. (in 1983 slightly modified by Coie and Dodge). However, many researchers, pedagogues, and parents feel that eliciting from children explicit dislike of other children conflicts with sound educational principles. To meet this objection, Asher and Dodge (1986) proposed a method (hereafter referred to as the AD system) that may be seen as a compromise. They suggested asking for positive nominations from the children, and, in addition, letting them rate on a scale how much they like playing with any of their group peers. The positive and negative information required for a two-dimensional classification is derived from the number of positive nominations received and from the number of lowest like-to-play ratings received, respectively.

During the 1980s, the use of rating scales was overshadowed by the nominations-based methods, although rating scales continued to be used (e.g., French, 1988). The rating scale method developed by Asher, Singleton, Tinsley, and Hymel (1979) is used most often with young chil-

dren (e.g., Ladd, 1990). The principal reason for the prevalence of the nomination methods was probably that the rating methods could not fulfill the requirement for a two-dimensional classification into sociometric status. In the 1990s, Maassen et al. (1996) presented a method based on a probability model (labeled SSrat) that enables researchers to classify subjects into the five aforementioned status groups according to data collected using a bipolar rating scale. In this method, positive and negative information is acquired by splitting the sociometric rating scale into positive and negative halves. An algorithm processing the sums of the received ratings on the positive and negative halves produces a classification into the five common sociometric status groups. (See Hymel, Vaillancourt, McDougall, & Renshaw, 2002, for a more detailed recent account of the history of sociometry, including SSrat.)

Like the AD system, SSrat does not require pupils to nominate classmates they dislike. But SSrat provides several additional advantages. First, it requires only the use of a single rating scale, in contrast to the AD system, which requires an additional nomination measure. Although this extra task may not be too laborious for the participating children, it is far more complex for researchers from the standpoint of design, administration, and scoring (Kane & Lawler, 1978). Second, and more important, sociometric status determination in SSrat is based (a) on all assessors' evaluations of all subjects, and (b) on polytomous score attributions. The latter holds true only when a study employs rating scales with at least five scale points (such as ranging from "dislike very much" to "like very much") instead of dichotomous scores (such as nomination or not, lowest like-to-play rating or not). A classification based on such more refined and differentiated information is expected to be more psychometrically sound (Rubin, Bukowski, & Parker, 1998). Whether this is observed in practice is the main question explored in the present study.

Research Aims

The present study has been designed to explore the sociometric status classifications generated by the nomination method of Coie and Dodge, the system of Asher and Dodge, and the two-dimensional rating scale technique SSrat. Specifically, the aims are to compare the three systems' *validity*, using peer behavioral nominations and teacher ratings of behavior characteristics, as well as their *stability*, using a 3-year interval.

First, as an index of validity, variables were employed that also indicate peer perception—that is, items of a behavioral peer nomina-

tion questionnaire indicating aggressive behavior and shyness or withdrawn conduct. We examined whether and to what extent the classifications yielded by the three methods showed the following hypothesized outcomes: (a) that the popular category, as compared to the average category, was nominated less as being aggressive; (b) that the rejected category, as compared to the average category, was nominated more as being aggressive and as being shy/withdrawn (see Rubin et al., 1998, for evidence that the group of rejected children is on average not only more aggressive but also more withdrawn). Furthermore, we followed other researchers (e.g., Frederickson & Furnham, 1998) in examining whether the neglected category was nominated by peers as more shy/withdrawn than the average category. This should be regarded as an empirical question, since the issue of whether sociometrically neglected children are more socially withdrawn than other children is a matter of debate. Some authors have reported that neglected children are generally more withdrawn (see, e.g., Newcomb, Bukowski, & Pattee, 1993), whereas other researchers have not found such differences (see Rubin et al., 1998, for a review).

Second, teachers' ratings of children's conduct or reputation were employed as validity criteria—in particular, ratings referring to aggression and popularity. We hypothesized that the teacher rates the popular category as more popular and less aggressive than the average category, and that the rejected category is seen as less popular, more aggressive, and more withdrawn. The question of whether neglected children are rated by teachers as more withdrawn was also explored. Teacher ratings have the advantage that they are not linked to the method for determining sociometric status, because they stem from different questionnaires and different informants. On the other hand, they have the disadvantage that they stem from the teachers' perspective, which may differ from the students' views of each other (Newcomb et al., 1993).

The second major aim of the study was to compare the longer-term stability of the three sociometric classification systems. As explained above, SSRAT gathers and processes more refined information than the other two systems, making it plausible to anticipate that the use of SSRAT generates more reliable classifications. Since the most reliable method can be expected to show the greatest stability—under conditions where a high stability is desired and under conditions where a relatively low stability is anticipated—SSRAT was hypothesized to yield more stable sociometric classifications than the other two systems. In this study, the stability was tested using a 3-year interval. Studies in which the stability of sociometric status types has been tested are unexpectedly rare. As noted by Cillessen, Bukowski, and Haselager, "it is

surprising that the stability of status classifications has been addressed in only a dozen studies" (2000, p. 77). As yet, no comparative study of stability includes the SSrat technique. Our study helps to fill this gap in the research literature.

Method

Participants

The study consisted of two phases. In the first phase, 216 children (104 boys and 112 girls) participated. The children were taken from eight third-grade and two fourth-grade regular educational classes across four schools in the neighborhood of Mechelen, a medium-sized city in Flanders (Belgium). The number of children in the classrooms varied from 16 to 28 ($M = 25$). Parental consent was obtained for all the children to participate. The children were between 8 and 10.4 years of age ($M = 9$ years). They were predominantly Caucasian and came from lower-middle-class to middle-class families. Most (more than 93%) had lived with both biological parents since birth. Nomination data are available for all 216 children, and ratings data are available for all but 2.

About 3 years after the first measurement (between 33 and 34 months), the same schools were contacted again. The children who had participated at the first measurement now attended sixth grade. No follow-up data were obtained from the two fourth-grade classes, because they had already left primary school. Over the 3-year period, the composition of four third-grade classes remained more or less the same; percentages of pupils still in the same class ranged from 88 to 92. Two classes have been rearranged into two new classes, according to the then policy of the school in question, and two classes had been regrouped into three smaller ones. The percentage of pupils of the four original classes who remained in the same class was no more than 55. At the second measurement, the number of students of the nine classes varied between 16 and 29 ($M = 22$). Their age was 11.9 years on average, varying between 11 and 14.1 years. The sociometric nomination and ratings data of 198 pupils (95 boys and 103 girls) on this second occasion were available. Nominations from both measurement occasions were available for 170 children and the ratings for 169 of them. Participation rates in the 17 third- and sixth-grade classrooms from the longitudinal sample varied from 85% to 100%. For no clear reason, the two fourth-grade classes had lower participation rates (61 and 71%, respectively). As explained above, these classes did not participate in the follow-up study.

Procedure

At both assessments, testing took place in a classroom setting during normal school hours and was conducted by one female experimenter. In the third- and fourth-grade classes, one male aide also was available. After an introductory explanation, the children were presented with several questionnaires, including the sociometric assessments. The participants were asked to write down from memory positive and negative nominations, nominations for aggression, shyness/withdrawal, and popularity, and to fill out sociometric rating scales (fixed order of presentation based on, e.g., Bukowski, Sippola, Hoza, & Newcomb, 2000; Coie & Kupersmidt, 1983; Dodge, 1983; Dodge & Frame, 1982). The children were encouraged to ask for help if they encountered difficulties in completing the assessments. The testing session lasted no more than 1 hour in each classroom. At the end of the session, each child received a small gift.

The teacher rating scales were explained and given to the teachers. At the first measurement, 11 teachers (8 females) completed the rating scales for the participating children in their classroom. (One of the 10 classes was supervised by two teachers, each of whom filled in roughly half of the questionnaires.) At the second measurement, 9 teachers (all females) filled in these rating scales. After one to two weeks, the completed scales were collected. Full teacher rating data were available for 214 children at the first measurement and for 196 children at the second measurement. For 168 children, full teacher rating data at two time points were obtained.

Sociometric Measures

Children were asked to nominate three classmates with whom they most liked to play (positive nominations) and three with whom they least liked to play (negative nominations). In addition, they were asked to rate on a 5-point scale how much they liked to play with each of their classmates (sociometric ratings). To determine the sociometric status classification, three systems were used: the standard score method of Coie and Dodge based on the nomination measure only; the two-dimensional ratings system SSRAT based on the rating scale only; and the Asher and Dodge system based both on the nomination measure and the rating scale.

The standard score method of Coie and Dodge (CDnom). In the system proposed by Coie and Dodge (1983), positive and negative nominations are summed to “liked most” (*LM*) and “liked least” (*LL*) scores. *LM* and *LL* are standardized within the group. Next, these

standard scores are transformed into two new variables: *SP* (*social preference*) and *SI* (*social impact*) calculated as their difference and sum, respectively, hence $SP = z_{LM} - z_{LL}$ and $SI = z_{LM} + z_{LL}$. The resulting scores are also standardized. Finally, the subjects are attributed to five sociometric status categories as follows: (a) popular, persons with standardized $SP > 1$, $z_{LM} > 0$ and $z_{LL} < 0$; (b) rejected, persons with standardized $SP < -1$, $z_{LM} < 0$ and $z_{LL} > 0$; (c) neglected, persons with standardized $SI < -1$, $z_{LM} < 0$ and $z_{LL} < 0$; (d) controversial, persons with standardized $SI > 1$, $z_{LM} > 0$ and $z_{LL} > 0$; and (e) average, all remaining group members.¹

The Asher and Dodge system (AD system). In the system introduced by Asher and Dodge (1986), both the positive nominations and the lowest like-to-play ratings are processed according to the algorithm of Coie and Dodge, on the understanding that the number of negative nominations received is replaced by the number of lowest like-to-play ratings received.²

The two-dimensional ratings system (SSrat). SSrat, introduced by Maassen et al. (1996), uses data that have been collected with a bipolar rating scale composed of an odd number of scale points. In the present study, a 5-point rating scale was used, with the scores 1 and 2 corresponding to disliking, the neutral scale point 3, and the scores 4 and 5 corresponding to liking. SSrat has been designed on the basis of the same concept as the nomination method CDnom. Because the rating scale comprises a positive half (scores 4 and 5) and a negative half (scores 1 and 2), negative and positive information is collected using a single scale. Instead of processing positive and negative nomination totals, the sums of the ratings on the positive and negative halves are calculated. Again, these total scores are transformed into two new variables: social preference and social impact. These two variables allow classification into the five common sociometric status groups. SSrat is a probability method, and the division into status categories is guided by the test of whether social preference, social impact, and positive and negative rating totals turn out to be higher or lower than expected on the basis of chance. The probability distributions are derived from multinomial distributions estimated from the score patterns of the

1. This procedure is Coie and Dodge's (1983) modification of the method proposed by Coie et al. (1982), who denoted the average category as persons with standardized *SP* and *SI* between -0.5 and < 0.5 , while they distinguished a sixth category of all unclassified group members.

2. The method reported by Asher and Dodge (1986) followed Coie et al. (1982) in defining the average category and distinguishing a sixth category of unclassified children (see previous note). In the present study, classification criteria were used that are fully compatible with the Coie and Dodge (1983) algorithm.

assessors. Thus, unlike the other two techniques, the SSRat algorithm has the advantage of reintroducing an old concept of Moreno (1934), that the opinions of the assessed persons about their group should also be taken into account. For further technical details, see Maassen et al. (1996) or Maassen and Landsheer (1998). A text on an introductory level is provided by Maassen, Van der Linden, Goossens, and Bokhorst (2000).³

Behavioral Measures

Peer behavioral nominations. Following the sociometric nominations, and based on Dodge (1983), children were asked to nominate three classmates with regard to various qualifications. Two of the questions were selected for validity purposes: (1) Who starts to fight and beats other children? and (2) Who is shy and does not play with other children? These two items denote aggression and shyness/withdrawal, respectively. For each child, the total of nominations received was calculated per item and standardized within the class group.

Teachers' behavioral ratings. The teachers of the subjects were presented with a questionnaire consisting of six 6-point rating scales. Three of these items were selected for validation purposes:⁴ (1) How much is this child liked by his or her classmates? (response anchors ranging from "not liked at all" to "liked very much"); (2) How aggressive is this child's behavior towards his or her classmates? (ranging from "not aggressive at all" to "very aggressive"); and (3) To what extent does the child show withdrawn conduct? (ranging from "not withdrawn at all" to "very withdrawn").

Although adult report may offer a qualitatively different perspective on children's peer relationships than peer report, meta-analysis has revealed that adult and peer report are equally successful in detecting differences in aggression, withdrawal, and sociability between the rejected and popular groups on the one hand and the average group on the other hand (Newcomb et al., 1993).

Choice of Cutoff Limits

In the many studies where sociometric status has been determined by the CDnom method or the AD system, the researchers generally have

3. The program SSRat and its manual (Maassen & Landsheer, 1996) can be requested from the first author.

4. The other teacher rating scales were (a) involvement with classroom activities, (b) self-confidence, and (c) school results.

applied these procedures as presented by their authors; that is, with $\pm 1 SD$ as cutoff limits. These procedures have been accepted by the researchers in the field as standards, and therefore in this article they will be referred to as the standard cutoffs. Since SSrat is a probability method, the delimitation of the status categories is guided by a chosen α value. No standard has been developed with regard to the choice of α . On the contrary, the idea of a standard variant may be inadequate, since SSrat is applicable to rating scales with a varied number of scale points and different expressions of social preference as response anchors. The present article compares SSrat with the AD system, implying that SSrat should now use input of 5-point ratings that indicate how much a child likes to play with other children, as is prescribed for the AD procedure. If popular and rejected categories are desired of sizes similar to those yielded by the standard nomination cutoffs, then α can be chosen accordingly. On the other hand, Maassen, Van Boxtel, and Goossens (2005) argued that additional information on the social context of the assessments may indicate that the category sizes yielded by the standard variant of CDnom are inadequate. For example, when almost every child is liked to play with, it is conceivable that a relatively high α value should be taken with respect to the popular category, and a lower α value with respect to the rejected category. In a preliminary analysis of the present study, we examined whether category sizes should be rendered deviant from those yielded by the standard variant of CDnom.

Information on the social context can be gathered by examining the mean average rating received (*ARR*) in the sample. The *ARR* is a more "absolute" measure of the level of acceptance of a pupil within the classroom. The mean *ARR* for the total sample was equal to 2.91 ($N = 214$) on the first occasion and to 2.98 ($N = 198$) on the second occasion. These values are close to the neutral scale midpoint (i.e., 3.0). At Time 1 the *ARR* of 52% of the participants was below the scale midpoint. At Time 2 this percentage was somewhat lower (46%). Since the proportion of students who received a negative average rating and the proportion of students who received a positive average rating each amounted to about half of the sample, the "reservoirs" from which the rejected or popular groups would emerge did not prove to be particularly limited. Hence there seemed to be no reason to apply cutoff limits deviant from the standard cutoff limits ($\pm 1 SD$) when forming the popular or rejected categories using CDnom and the AD system. To find a corresponding cutoff limit for SSrat, yielding sizes similar to those of the standard variants of CDnom and the AD system, SSrat was executed with α values varying between .010

and .050. Results of these analyses showed that the variant of SSRAT with $\alpha = .025$ yielded popular and rejected categories of conventional size on both occasions. Therefore, only this variant of SSRAT will be tabulated and discussed.

Data Analysis

We prefer to express the discriminatory power of the methods in terms of effect sizes (ESs), that is, mean differences divided by the within-groups standard deviation (Cohen, 1992). In our view, t values or p values are less appropriate for comparisons between sociometric classifications. First, the participants were assessed within class groups and hence cannot be regarded as independent units of measurement. T values or p values are based on probability distributions assuming independence of observations, whereas the calculation of ES is not. Second, these statistics depend on the sizes of the sociometric categories involved, which may vary markedly between methods. Nevertheless, for purposes of comparability with other studies and for the convenience of the reader, the significance of the corresponding t values is also indicated in the tables below. In our view, it is inappropriate to interpret small differences of discriminatory power between the three methods. Therefore, the mean value of all the differences between the largest and the smallest ES in the 16 cases considered (see Tables 2 and 3) was calculated, which proved to be .24. Only differences of discriminatory power are seen as substantial and are interpreted further, if the difference between the largest and smallest ES exceeds this value.

As usual, the stability of a classification method will be measured by Cohen's κ (Cohen, 1960), expressing the extent to which the agreement between two classifications exceeds what would be expected on the basis of chance. In addition, the stability of each sociometric status category is also separately expressed in terms of κ (see, e.g., Cillessen et al., 2000; Terry & Coie, 1991). Because a statistical test of the difference between two correlated κ coefficients for polytomous nongraduated category scores is not (yet) available, differences between stabilities are tested using McNemar's test for correlated proportions.

Results

Global Classificatory Results

Table 1 presents the results of the sociometric status classifications according to the different procedures. On both occasions, the popular and rejected categories resulting from the standard variants of CDnom

and the AD system were of conventional size (about 15%). The neglected categories were of similar size. The variant of SSrat with $\alpha = .025$ also yielded popular and rejected categories of conventional size on both occasions, but the neglected and the controversial categories were considerably smaller. This confirms earlier findings (see Maassen et al., 2000, p. 64, for an explanation).

The agreement between the standard cutoffs of CDnom and the AD system expressed in Cohen's κ equaled .48 on the first occasion and .59 on the second, roughly the same values reported by Asher and Dodge (1986). Cohen's κ between the SSrat variant with $\alpha = .025$ and the AD procedure equaled .37 at Time 1 and .41 at Time 2. Thus, the latter system was shown to have more in common with CDnom than with SSrat.

Validity

The social preference dimension. In terms of the validity criteria, Tables 2 and 3 show several characteristics of the popular, average, and rejected categories resulting from the three classification procedures. Because the means of these criteria are not readily interpretable, the tables only report ESs. All the methods corroborated the hypothesis that the popular category is on average nominated as less aggressive than the average category. No clear differences in ESs were seen. Support for the hypothesis that the rejected group is on average considered more aggressive than the average group was much weaker. The t value was significant for none of the CDnom classifications. At Time 2, this also held for the two other procedures. At Time 1, however, ESs clearly differed between methods (difference between largest and smallest ES $> .24$). As can be seen in Table 2, SSrat showed the largest ES and CDnom the lowest. The hypothesis that the rejected category receives

Table 1. Sociometric Status Distributions (in Percentages)
Using Different Classification Procedures

Time	Method	α	N	Pop	Rej	Neg	Con	Ave
1	SSrat	.025	214	12.1	14.5	2.3	1.4	69.6
	CDnom		216	13.9	14.4	16.7	7.4	47.7
	AD		214	15.9	12.6	13.6	5.6	52.3
2	SSrat	.025	198	17.7	16.2	2.5	1.0	62.6
	CDnom		198	13.6	15.7	11.6	2.5	56.6
	AD		198	14.6	14.6	14.6	4.5	51.5

Note. Pop = popular, Rej = rejected, Neg = neglected, Con = controversial, Ave = average.

Table 2. Comparisons of the Popular (P) Versus the Average (A), and the Rejected (R) Versus the Average Categories for Peer Nominations of Aggression and Shyness/Withdrawal in Terms of Effect Sizes (ES) Using Various Classification Procedures

Time	Method	α	N			Aggression		Shyness
			P	A	R	ES		
						P vs. A	R vs. A	R vs. A
1	SSrat	.025	26	149	31	-0.57	0.84	0.96
	CDnom		30	103	31	-0.59	0.51*	0.61
	AD		34	112	27	-0.55	0.74	0.30*
2	SSrat	.025	35	124	32	-0.50	0.02*	1.17
	CDnom		27	112	31	-0.51	0.09*	0.57*
	AD		29	102	29	-0.42	0.16*	0.83

Note. Differences between the largest and smallest ESs of the three classifications that exceed .24 appear in bold and italic.

* t value not significant (one-sided $p > .01$).

more shyness/withdrawal nominations than the average categories was again most strongly supported by the results of SSRat at both occasions. In sum, with regard to the peer behavioral nominations, in all three cases where ESs clearly differed between methods, the results favored SSRat.

Table 3 presents the outcomes of the analyses of the teacher ratings regarding the popularity reputation and aggressive or withdrawn conduct. Unlike the evidence obtained from the peer nominations, on average the teachers saw no difference in aggression between the popular and the average categories, whatever the classification method applied. Regarding the contrast between the rejected and the average categories, a clear difference in ESs between classification methods appeared only at Time 1. Again, SSRat discriminated best and CDnom did worst. In line with the results from peer nominations, no significant differences in aggression were found at Time 2. The hypothesis that the popular category is on average better liked by classmates than the average category was clearly corroborated by the teachers' ratings at Time 2, and less clearly at Time 1. The three methods showed similar performance in this respect. The hypothesis that the rejected category is regarded as less liked by classmates than the average category was strongly supported by the teachers' ratings at both assessments. At Time 1, clear differences between classifications were obtained. Again SSRat performed best and CDnom worst. Regarding the teacher ratings of withdrawn conduct, only SSRat was observed to yield a significant contrast between the rejected and average categories at Time 2.

Table 3. Comparisons of the Popular (P) Versus the Average (A), and the Rejected (R) Versus the Average Categories for Teacher Ratings of Aggression, Popularity, and Withdrawal in Terms of Effect Sizes (ES) Using Various Classification Procedures

Time	Method	α	N			Aggression		Popularity		With-
			P	A	R	P vs. A	R vs. A	P vs. A	R vs. A	drawal
			ES		ES		ES			
1	SSrat	.025	26	148	31	-0.22*	0.88	0.39*	-1.15	-0.04*
	CDnom		29	102	31	-0.16*	0.59*	0.40*	-0.76	-0.02*
	AD		34	112	27	-0.06*	0.68	0.46	-1.00	0.09*
2	SSrat	.025	35	125	31	-0.28*	0.45*	0.82	-1.35	0.77
	CDnom		27	112	30	-0.27*	0.41*	0.78	-1.14	0.41*
	AD		29	102	28	-0.15*	0.61*	0.80	-1.15	0.54*

Note. Differences between the largest and smallest ESs of the three classifications that exceed .24 appear in bold and italic.

* Corresponding t value not significant (one-sided $p > .01$).

Here too, the three methods clearly differed with regard to ES, and once again SSrat performed best, whereas CDnom showed the worst performance.

In sum, with regard to the teacher ratings, in all three cases where ESs clearly differed between methods, the results favored SSrat. The performance of CDnom was evidently the weakest. This corroborates the findings from peer nominations.

The social impact dimension. The question of whether children in the neglected category are more withdrawn than children in the average category was explored using both peer nominations and teachers' ratings of shyness/withdrawal. As stated earlier, evidence for the idea that neglected children are more shy/withdrawn than average children is not equivocal. In this study, no equivocal evidence was found either. The ESs corresponding to SSrat and CDnom were inconsistent, showing positive as well as negative ESs. The AD system yielded the most consistent patterns of ESs. However, none of the t values corresponding with the ESs was significant (one-sided testing with $\alpha = .05$), which makes any interpretation dubious.

Stability of the Classifications

We noted earlier that during the interval between the two assessments the composition of four of the original class groups had drastically changed. This opened an additional opportunity for validation, since

one might anticipate that the stability of the classifications in the four stable classes was comparatively high, which should be shown by all three methods. Therefore, it would be useful to examine the stability of the classifications based on the complete sample as well as within the four class groups that remained overall unchanged.

The stability of the different classification methods can be examined with in several aspects. First, stabilities in terms of Cohen's κ will be considered. Table 4 shows that the κ values were generally low, not surprising given the long interval between the two assessments. The stability of the classifications according to the SSrat ratings method was generally appreciably higher than that of CDnom and the AD system. Moreover, the hypothesis that the stabilities within the four selected class groups were comparatively high was confirmed only when using SSrat, whereas the stabilities of CDnom and the AD system were even lower than those of the complete sample. The table also demonstrates that, particularly within the four stable classes, SSrat produced an appreciably higher stability regarding the popular and the rejected categories. Finally, it is evident that any classificatory method yielded extremely unstable neglected and controversial categories.

When stabilities are expressed as percentages of persons remaining in the same category across assessments, the results were similar (see Table 4). McNemar's test for correlated proportions shows that the stability of the assessment using SSrat was significantly higher than when using the other methods. (For all the participants: SSrat vs. CDnom yielded $\chi^2(1) = 10.2, p < .005$ and SSrat vs. AD $\chi^2(1) = 11.4, p < .001$; for the four stable classes: SSrat vs. CDnom yielded $\chi^2(1) = 14.9, p < .001$ and SSrat vs. AD $\chi^2(1) = 13.4, p < .001$.) Particularly within the four selected classes, the ratings-based variant showed higher percentages for the popular and rejected categories (73 and 38, respectively) than did CDnom (7 and 25) or the AD system (24 and 20).

We also examined the stability (i.e., the within-variable across-time correlation) of the standardized liked most (*LM*), standardized liked least (*LL*), social preference (*SP*) and average received rating (*ARR*) scores. In line with the results obtained from the sociometric classification procedure, stabilities were higher for the *ARR* ($r = .56$) than for the *LM* scores ($r = .34$; using the test for the difference between correlated correlations: $\chi^2(1) = 9.0, p < .005$) and for the *LL* scores ($r = .35$; $\chi^2(1) = 9.0, p < .005$). In the present study the stability of *ARR* was also higher than that of the *SP* scores ($r = .44$), but the difference was not significant ($\chi^2(1) = 3.7, p > .05$).

Table 4. Correlation in Terms of Cohen's κ Between Two Assessments of Sociometric Status in the Total Sample and in the Separate Status Categories

Method	α	N	Pop	Rej	Neg	Con	Ave	Total
All participants								
SSrat	.025	169	.26 (45)	.31 (38)	.20 (25)	-.02 (0)	.12 (69)	.20 (59)
CDnom		170	.25 (33)	.27 (35)	.02 (12)	-.06 (0)	.00 (62)	.10 (42)
AD		169	.23 (34)	.16 (26)	.10 (26)	-.05 (0)	.04 (55)	.10 (41)
Four stable class groups								
SSrat	.025	88	.46 (73)	.32 (38)	-.02 (0)	-.02 (0)	.22 (70)	.29 (62)
CDnom		89	-.05 (7)	.13 (25)	-.11 (0)	-.07 (0)	-.15 (58)	-.05 (34)
AD		88	.10 (24)	.08 (20)	.10 (23)	-.07 (0)	-.02 (51)	.04 (34)

Note. Pop = popular, Rej = rejected, Neg = neglected, Con = controversial, Ave = average. Values in parentheses represent percentages of individuals remaining in the same category with respect to the previous assessment's category size.

Discussion

The central question of this study is whether SSrat, a method for two-dimensional sociometric status classification with ratings (Maassen et al., 1996), is a useful alternative to the AD system of Asher and Dodge (1986). Both SSrat and the AD system are useful tools to meet the educator's objection to soliciting negative nominations from children. The two methods differ in the employment of the available data. It was expected that the processing of more complete and refined information by SSrat would yield more valid and stable sociometric classifications. The main purpose of this study was to test whether this was the case. As a frame of reference, the most widely used nomination procedure, developed by Coie and Dodge (1983), was also included in the comparison.

In the few relevant published studies of the *stability* of sociometric status types, stability of sociometric categories was found to be low, especially over longer time intervals (Cillessen et al., 2000). In two studies in which a sample of third-graders was followed across 2 and 4 years, κ was equal to .19 (Terry & Coie, 1991) and .01 (Coie & Dodge, 1983), respectively. In our study, a sample of third-graders was followed across a 3-year interval, and κ ranged between .10 and .20 for the different classification methods. Results showed that sociometric classifications based on SSrat were more stable than classifications resulting from the other two methods. Furthermore, the hypothesis that the stability of the classification within the four classes whose composition remained roughly unchanged exceeded the stability based on the

complete sample was corroborated only when using SSrat. Within this subsample, the κ value of the popular category reached the criterion for moderate stability ($\kappa = .46$). In contrast, when using the standard nomination or the AD method, very low stabilities were established for this popular group ($\kappa = -.05$ and $\kappa = .10$, respectively). The superiority of the SSrat system in terms of stability was also found for the rejected and the average group. The higher stability of classifications yielded by SSrat, as compared to the other two classification methods, may result from the fact that SSrat is based on obtained *ARR* (average rating received) scores. The *ARR* scores proved more stable than the scores upon which the nominations method, and to a lesser extent the method of Asher and Dodge, are based (including social preference, liked-most, and liked-least scores).

In the present study, the stability of the sociometric status classification yielded by CDnom was extremely low. This has also been reported in several earlier studies (Coie & Dodge, 1983; Gresham & Stuart, 1992; Newcomb & Bukowski, 1984). The stability of the AD system, which is central to the present study, was also very low, suggesting that the nomination component of this procedure and the analogy with CDnom are important. This observation is not new; Asher and Dodge (1986), when introducing their method, emphasized the similarities with CDnom. They reported a high correlation between the liked-least score (the total of liked-least nominations received) and the lowest like-to-play ratings ($r = .80$) and concluded that "the two approaches were found to yield similar classifications overall, as indicated by Cohen's κ ($\kappa = .51$)" (p.446). Terry and Coie (1991) have reported similar agreement between both systems. For third-, fourth-, and fifth-graders they found $\kappa = .47$, $.56$, and $.55$, respectively. Moreover, in the present study, results produced by the AD system showed a higher level of agreement with CDnom classifications than did results produced by SSrat. Although Asher and Dodge were satisfied with the agreement between the results of their system and those of the Coie and Dodge method, the present study shows that this agreement is not necessarily a recommendation, and that the first system may be plagued by the less favorable characteristics of the latter.

Regarding the low stability coefficients of the CDnom and the AD method in the present study, it should be acknowledged that the participants were asked to nominate a limited number of peers (no more than three), as is usual to date. Terry (2000) advocated the use of unlimited nominations, which should increase the stability. Allowing an unlimited number of nominations is not yet common practice. Moreover, in the case of school classes of 30 or fewer children (in this

study the number of pupils varied between 16 and 29) the advantage seems doubtful, since in our experience, for many children in such smaller classes it is difficult to nominate even three peers they dislike very much and three peers they like very much. Thus, in that case, the limit of three nominations seems no serious restriction, whereas several important arguments in favor of the ratings method remain pertinent, such as the attribution of polytomous liking scores and accounting for the different score attribution of the assessors. But of course, this issue remains to be empirically tested in future studies comparing SSrat to CDnom based on unlimited nominations.

In the present study, the *validity* of the sociometric status grouping along the social preference dimension was examined by testing hypothesized differences between the status groups on several behavioral characteristics as measured by peer nominations and teacher ratings. SSrat classifications appeared to be most sensitive in detecting the predicted differences among the popular, average, and rejected status categories, and more specifically between the average and rejected categories. The validity of CDnom was clearly weakest. In particular, the predicted differences between the rejected category and the average category with regard to peer-nominated as well as teacher-rated aggressive and withdrawn conduct and with regard to teacher-rated popularity was found most clearly for classifications based on SSrat and least clearly for classifications based on CDnom. The predicted differences between the popular category and the average category were found for teacher-rated popularity and peer-nominated aggression (not for teacher-rated aggression), but all three classification methods showed equal performance. It should be noted that the validity criteria we used in this study consisted only of peer nominations and one-item teacher rating scales, which may explain why the expected differences were not always found at both measurement times. In each case, it is important to replicate the findings above using other measures of aggression, withdrawal, and popularity before drawing conclusions regarding the (lack of) differences in discriminatory power.

At present, however, all available findings unequivocally suggest that SSrat may be considered a valuable alternative to the standard nomination procedure of Coie and Dodge and the classification procedure of Asher and Dodge. Apart from its favorable performance in terms of validity and stability, SSrat has the practical advantage of using only a rating scale, whereas the AD system requires the use of an additional nomination measure.

The issues of temporal stability and validity of sociometric status addressed in this study are not only of methodological importance;

they also have conceptual or theoretical implications. Recent research, for example, is not only aimed at describing the stability of peer status but also focuses on the antecedents (in child and social environment) and consequents of having a stable peer sociometric status, specifically of being stable rejected (Cillessen et al., 2000). It is evident that the conclusions from these studies depend upon the reliability and validity of the sociometric status classifications that are used. Moreover, several studies have been conducted on the effects of interventions (coaching, social skills training, participation in peer involvement groups, etc.) on social behavior and social status of children (see Asher, Parker, & Walker, 1996; Cillessen et al., 2000, for reviews). As concluded by Cillessen et al. (2000), many of these intervention studies have shown that it is more difficult to improve children's sociometric status than it is to improve their social behaviors and skills. The present study shows that failure to find consistent improvements in sociometric status may result from methodological drawbacks of the classification methods that were used. Hence, for future studies it may be interesting to examine the sensitivity of different sociometric status classification procedures, including the SSRAT procedure, in detecting the hypothesized improvements. In general, we recommend that future empirical research on the stability and the correlates of children's peer status include more alternatives than the standard nomination method or the AD system. As suggested by the findings of this study, including the two-dimensional ratings-based procedure SSRAT may yield a more complete and valid picture of the correlates of peer acceptance and rejection.

References

- Asher, S. R., & Dodge, K. A. (1986). Identifying children who are rejected by their peers. *Developmental Psychology, 22*, 444–449.
- Asher, S. R., Parker, J. G., & Walker, D. L. (1996). Distinguishing friendship from acceptance: Implications for intervention and assessment. In W. M. Bukowski, A. F. Newcomb, & W. W. Hartup (Eds.), *The company they keep: Friendship in childhood and adolescence* (pp. 366–405). Cambridge: Cambridge University Press.
- Asher, S. R., Singleton, L. C., Tinsley, B. R., & Hymel, S. (1979). A reliable sociometric measure for preschool children. *Developmental Psychology, 15*, 443–444.
- Bukowski, W. M., Sippola, L., Hoza, B., & Newcomb, A. F. (2000). Pages from a sociometric notebook: An analysis of nomination and rating scale measures of acceptance, rejection, and social preference. In A. H. N. Cillessen & W. M. Bukowski (Eds.), *Recent advances in the measurement of acceptance and rejection in the peer system* (New Directions for Child

- and Adolescent Development Series No. 88, pp. 11–26). San Francisco: Jossey-Bass.
- Cillessen, A. H. N., & Bukowski, W. M. (2000). Conceptualizing and measuring peer acceptance and rejection. In A. H. N. Cillessen & W. M. Bukowski (Eds.), *Recent advances in the measurement of acceptance and rejection in the peer system* (New Directions for Child and Adolescent Development Series No. 88, pp. 3–10). San Francisco: Jossey-Bass.
- Cillessen, A. H. N., Bukowski, W. M., & Haselager, G. J. T. (2000). Stability of sociometric categories. In A. H. N. Cillessen & W. M. Bukowski (Eds.), *Recent advances in the measurement of acceptance and rejection in the peer system* (New Directions for Child and Adolescent Development Series No. 88, pp. 75–93). San Francisco: Jossey-Bass.
- Cohen, J. (1960). A coefficient agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37–46.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159.
- Coie, J. D., & Dodge, K. A. (1983). Continuities and changes in children's social status: A five-year longitudinal study. *Merrill-Palmer Quarterly, 29*, 261–282.
- Coie, J. D., Dodge, K. A., & Coppotelli, H. (1982). Dimensions and types of social status: A cross-age perspective. *Developmental Psychology, 18*, 557–570.
- Coie, J. D., & Kupersmidt, J. B. (1983). A behavioral analysis of emerging social status in boys' groups. *Child Development, 54*, 1400–1416.
- Dodge, K. A. (1983). Behavioral antecedents of peer social status. *Child Development, 54*, 1386–1399.
- Dodge, K. A., & Frame, C. L. (1982). Social cognitive biases and deficits in aggressive boys. *Child Development, 53*, 620–635.
- Dunnington, M. J. (1957). Investigation of areas of disagreement in sociometric measurement of preschool children. *Child Development, 28*, 93–102.
- Frederickson, N. L., & Furnham, A. F. (1998). Sociometric classification methods in school peer groups: A comparative investigation. *Journal of Child Psychology and Psychiatry, 39*, 921–933.
- French, D. C. (1988). Heterogeneity of peer-rejected boys: Aggressive and nonaggressive subtypes. *Child Development, 59*, 976–985.
- Gresham, F. M., & Stuart, D. (1992). Stability of sociometric assessment: Implication for uses as selection and outcome measures in social skills training. *Journal of School Psychology, 30*, 223–231.
- Hymel, S., Vaillancourt, T., McDougall, P., & Renshaw, P. D. (2002). Acceptance and rejection in the peer group. In P. Smith & C. Hart (Eds.), *Handbook of child social development* (pp. 265–284). London: Blackwell.
- Kane, J. S., & Lawler, E. E. (1978). Methods of peer assessment. *Psychological Bulletin, 85*, 555–586.

- Ladd, G. W. (1990). Having friends, keeping friends, making friends, and being liked by peers in the classroom: Predictors of children's early school adjustment? *Child Development, 61*, 1081–1100.
- Maassen, G. H., Akkermans, W., & Van der Linden, J. L. (1996). Two-dimensional sociometric status determination with rating scales. *Small Group Research, 27*, 56–78.
- Maassen, G. H., & Landsheer, J. A. (1996). *Manual for SSrat 2.0: A program for two-dimensional sociometric status determination with rating scales* (Method Series MS-96-2). Utrecht, the Netherlands: FSW/ISOR Publications, Utrecht University.
- Maassen, G. H., & Landsheer, J. A. (1998). SSrat: The processing of rating scales for the determination of two-dimensional sociometric status. *Behavioral Research Methods, Instruments, and Computers, 30*, 674–679.
- Maassen, G. H., Van Boxtel, H. W., & Goossens, F. A. (2005). *Cutoff limits in nomination and rating scale methods for sociometric status determination*. Manuscript submitted for publication.
- Maassen, G. H., Van der Linden, J. L., Goossens, F. A., & Bokhorst, J. (2000). A ratings-based approach to two-dimensional sociometric status determination. In A. H. N. Cillessen & W. M. Bukowski (Eds.), *Recent advances in the study of acceptance and rejection in the peer system* (New Directions for Child and Adolescent Development Series No. 88, pp. 55–73). San Francisco: Jossey-Bass.
- Moreno, J. L. (1934). *Who shall survive? A new approach to the problem of human interrelations*. Washington, DC: Nervous and Mental Disease Publishing.
- Newcomb, A. F., & Bukowski, W. M. (1983). Social impact and social preference as determinants of children's peer group status. *Developmental Psychology, 19*, 856–867.
- Newcomb, A. F., & Bukowski, W. M. (1984). A longitudinal study of the utility of social preference and social impact sociometric classification schemes. *Child Development, 55*, 1434–1447.
- Newcomb, A. F., Bukowski, W. M., & Pattee, L. (1993). Children's peer relations: A meta-analytic review of popular, rejected, neglected, controversial, and average sociometric status. *Psychological Bulletin, 113*, 99–128.
- Peery, J. C. (1979). Popular, amiable, isolated, rejected: A reconceptualization of sociometric status in preschool children. *Child Development, 50*, 1231–1234.
- Rubin, K. H., Bukowski, M. W., & Parker, J. G. (1998). Peer interactions, relationships, and groups. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (5th ed., pp. 619–700). New York: Wiley.
- Terry, R. (2000). Recent advances in measurement theory and the use of sociometric techniques. In A. H. N. Cillessen, & W. M. Bukowski (Eds.), *Recent advances in the measurement of acceptance and rejection in the peer*

-
- system*. (New Directions for Child and Adolescent Development No. 88, pp. 27–53). San Francisco: Jossey-Bass.
- Terry, R., & Coie, J. D. (1991). A comparison of methods for defining socio-metric status among children. *Developmental Psychology, 27*, 867–880.
- Thompson, G. G., & Powell, M. (1951). An investigation of the rating-scale approach to the measurement of social status. *Educational and Psychological Measurement, 11*, 440–455.