Cillessen, A. H. N. & Marks, P. E. L. (2017). Methodological choices in peer nomination research. In Peter E. L. Marks & Antonius H. N. Cillessen (Eds.), *New Directions in Peer Nomination Methodology. New Directions for Child and Adolescent Development*, *157*, 21–44.

1

Methodological Choices in Peer Nomination Research

Antonius H. N. Cillessen, Peter E. L. Marks

Abstract

Although peer nomination measures have been used by researchers for nearly a century, common methodological practices and rules of thumb (e.g., which variables to measure; use of limited vs. unlimited nomination methods) have continued to develop in recent decades. At the same time, other key aspects of the basic nomination procedure (e.g., whether nonparticipants should be included as nominees, the consequences of pairing code numbers with names on rosters) are underdiscussed and understudied. Beyond providing a general introduction to peer nomination methods and their utility, the current article discusses the advantages and disadvantages of various methodological choices facing researchers who wish to use peer nomination methods, in addition to other considerations that researchers must make in collecting peer nomination data (e.g., establishing reliability and validity, maximizing participation rates, computerized assessments). This article provides recommendations for researchers based on empirical findings (where possible) and the typical practices used in the recent published literature.

Introduction and Background

The term sociometric assessment refers to the methods that are used to assess youths' social relationships in the context of their peer groups (Cillessen & Bukowski, in press). These methods have a long and varied history, from their original development by Moreno (1934) to frequent applications by teachers in schools to map the classroom in a sociogram (Grondlund, 1959) to their most recent implementations in the form of computerized methods for the assessment of classroom peer relationships in various age groups (see, e.g., Endedijk & Cillessen, 2015; van den Berg & Cillessen, 2013). A basic and important part of sociometric assessment is the use of peer nominations within a reference group (classroom or grade). Although there are other methods than peer nominations (e.g., ratings or paired comparisons), peer nominations continue to be the preferred methodology of researchers. Nevertheless, these methods have changed, and the contexts for their use have changed. The scientific context for peer assessment has changed through, for example, the availability of computers and new statistical methods. The practical context for peer assessment has changed as indicated by, for example, changing school policies and also by an increased interest in understanding peer relationships motivated by, for example, the need for bullying prevention, identification of students at risk, or the creation of a positive school climate. Given these changes, it is important to examine peer nomination methodology in its most recent forms and help researchers and practitioners alike to make informed decisions about their use. These are the goals of this article.

Developmental Significance of Peer Relationships. A large body of research has documented the developmental significance of peer relationships (Bukowski, Laursen, & Rubin, in press). Across development, youth spend an increasing amount of time with their peers. Peers are a context for learning prosocial skills and behaviors, such as collaboration, perspective taking, managing conflict, showing respect and support, and for cognitive growth. Peers are also a context for the socialization of aggression and antisocial behavior, such as bullying and substance use. The majority of peer relations research addresses the age range of 4 to 18 years, with fewer studies before the age of 4 and after the age of 18 (for exceptions, see, e.g., Endedijk, Cillessen, Cox, Bekkering, & Hunnius, 2015; Lansu & Cillessen, 2012). During this age range, the major peer context is at school, which provides for clear reference groups (classrooms or grades), and it is for the school context that sociometric assessment methods primarily have been developed. In addition to research, in the school context, sociometric methods can be used for several important applied purposes, such as the identification of children at risk, the improvement of classroom or school climate, and the creation of classroom seating arrangements that foster instruction and learning. Therefore, it continues to be important to develop

optimal methods of sociometric assessment for widespread use in various age groups.

Levels in the Study of the Peer Group. Typically, three levels are distinguished in the study of peer relations, and information about each of these three levels can be derived from a sociometric test using peer nominations. At the individual level, the status of each individual child or adolescent is considered, either continuously or categorically, by focusing on nominations received. For example, based on the numbers of liked most and liked least nominations received from their classmates, children can be assigned continuous scores for peer acceptance, peer rejection, and social preference and they can be assigned to one of five sociometric status categories (see Coie, Dodge, & Coppotelli, 1982). At the dyadic level, dyadic relationships are considered, such as friendships, romantic relationships, enmities, bully–victim dyads, and mutually aggressive pairs. The most common application here is the identification of best friend dyads from reciprocal "best friend" nominations. At the group level, subgroups or cliques can be derived and identified from peer nomination data. This is the domain of social network analysis, and, as for the individual level, continuous information can be determined for each individual such as their social network centrality, or individuals can be assigned to distinct peer groups (subgroups or cliques, with the additional complication of multiple group membership). This article focuses on the collection of sociometric data that can then be used or processed by the peer relations researcher for any of these purposes, but we focus primarily on the individual level of analysis, which is the closest to the original intentions of sociometric assessment and is often its primary purpose in research. But psychometrically sound peer nomination data, once collected, can easily be used for dyadic and network analyses as well.

Sociometric Assessment. In a standard sociometric procedure in a classroom (e.g., Coie et al., 1982; Newcomb & Bukowski, 1983), children are asked to nominate classroom peers they like most (*acceptance*) and they like least (*rejection*). The number of nominations received for each question is then counted for each child and corrected for differences in classroom size in some way (*sociometric standardization*). A continuous score for *social preference* is created by taking the difference between the resulting acceptance and rejection scores and again standardizing the resulting scores within classrooms. A continuous score for *social impact* is created by summing the standardized acceptance and rejection scores and restandardizing the results. Finally, using a set of decision rules, children are assigned to one of five sociometric status types: accepted (liked by many, disliked by few), rejected (disliked by many, liked by few), neglected (neither liked nor disliked), controversial (liked by some and disliked by others), or average (around the means of acceptance and rejection, or simply everyone who is not classified in the first four groups). Group assignments are based on

whether children's scores on the continuous variables are common or rare, using certain cut-off scores.

In a sociometric test so described, there are certain basic elements: reference group, voter population, "votee" population, sociometric criteria, data collection method, quantification method, method of standardization, sociometric dimensions, and classification method. The *reference group* is the group within which status is determined, typically the classroom or the entire grade level of a school. The *voter population* are the children or adolescents who participate as evaluators in a sociometric test. The *votee population* are the children or adolescents who are being evaluated. Ideally, all members of the reference group (e.g., classroom or grade) participate as both voters and votees. In practice, this is often not the case because of absenteeism on the day of testing or lack of permission to participate. The questions on a sociometric test are called *sociometric criteria* (today, usually "Who do you like the most?" and "Who do you like the least?").

Sociometric criteria are subjective evaluations that are personal to the voter—they can also be referred to as "affective" criteria because they describe interpersonal feelings or relationships. Items measuring acceptance and rejection (liked most, liked least, best friends) are in this category. They should be distinguished from reputational criteria that measure perceived behaviors or reputations rather than personal evaluations. Nominations of peers who start fights, cooperate and share, or stay by themselves are reputational items of social behaviors. Nominations of (most) popular, least popular, good looking, good at sports, or good at school are also reputational. In this article, we use the term sociometric assessment for the measurement of sociometric criteria only and the term peer assessment for the measurement of reputational criteria.

Peer Nominations in Comparison to Other Methods. Peer nominations are a form of peer informant measures that have multiple advantages over other methods of evaluation for children and adolescents. Peers tend to see each other most days for large amounts of time and in a variety of situations and on that basis know how individuals behave in the group (Coie & Dodge, 1988; Rubin & Cohen, 1986). Although they are never completely objective, they involve a greater amount of objectivity (when considering social behavior) than self-reports or parental reports (see Clarke & Ladd, 2000). Compared to observational methods, peer reports are much less costly and allow assessment of older individuals, as it is very difficult to make naturalistic observations of adolescents in middle and high school (Rubin & Cohen, 1986). Peer measures also allow for assessments of liking and status that cannot be adequately duplicated by observations (England & Petro, 1998), and peer assessments can assess low-frequency and/or "covert" behaviors (such as relational aggression) that are difficult to observe directly. Finally, there seems to be a consensus that peers have a better idea of each other's social abilities and peer relationships than do teachers (Rubin & Cohen, 1986). Indeed, teachers tend to display fair levels of

inaccuracy when asked to guess which students will be nominated by peers as most liked or most popular (e.g., Babad, 2001; van den Berg, Lansu, & Cillessen, 2015). Although teacher ratings have some advantages (teachers have a reference group against which they can judge student behavior and are better able to distinguish between different behaviors and traits; Clarke & Ladd, 2000; Coie & Dodge, 1988), such measures tend to reflect a single viewpoint (i.e., one teacher), whereas peer-informant data are based on the composite judgments or impressions of several peers.

The focus of this article is on peer nominations, which are not the only method of collecting data from peer informants. Classic sociometric procedures also include paired comparisons, in which participants are asked to compare each peer against each other peer; peer rankings, in which participants rank each peer; and peer ratings, in which participants rate each peer on a numerical scale (see Terry, 2000, for a complete discussion of these methods). Compared to peer nominations, each of these methods provides more nuanced and detailed data regarding each participant (e.g., Thompson & Powell, 1951). However, these alternative methods are practical only within small, well-acquainted groups (Keislar, 1957), and each of these methods is substantially more cumbersome than peer nominations. In a classroom with 30 students, for example, participants could provide answers to several peer nomination items in the amount of time it would take to rate all 29 peers on a 1–5 scale for a single item (Parkhurst & Asher, 1992). Given that results of peer nominations are often similar to those derived from other methods (Maassen, van Boxtel, & Goossens, 2005; Terry & Coie, 1991) but are more easily obtained, peer nominations are most commonly used today.

Peer Nominations as "Direct" Measures. One long-standing criticism of peer nomination research is that nominations are indirect measures of social behavior because they reflect peer perceptions of behaviors rather than measuring the behaviors directly. This is accurate only insofar as all methods of data collection assess perceptions; certainly, teacher reports, parent reports, and self-reports do so as well. Naturalistic observations arguably are more "direct" but are difficult to conduct accurately beyond early childhood and unlikely to adequately assess low-visibility and low-frequency behaviors. In addition, they are still limited by the researchers' conceptualization of constructs and by observers' evaluations. Overall, peer nominations are no less direct than other methods when assessing overt behaviors.

Of course, whether there can be a direct measurement or not will depend on the nature of the construct and whether peers have made direct observations of the construct. Researchers expect peer nominations to provide accurate measurements of social behaviors such as aggression or prosocial behavior, for example, but would not necessarily expect them to accurately assess intelligence or physical strength. At times, it is more important for researchers to determine which participants are seen by peers as the "best students" or "best athletes," whether or not these perceptions reflect more

objective measures such as grade point average or average number of points scored (Coleman, 1961).

Furthermore, we argue that affective and reputational items can be directly assessed only with peer informant measures. Liking and disliking, of course, are emotional reactions toward individual peers and cannot be externally observed in any direct manner. Popularity, in turn, is a reputation—it is a consensus among the peer group and therefore can be directly measured only by assessing the peer group's consensus (Cillessen & Marks, 2011). Although teachers may guess which students are well liked (Babad, 2001) and observers may record by-products of status such as peer visual regard (Pellegrini et al., 2007), these are clearly indirect measures.

Specific Methodological Choices

The goal of this article is to address issues and provide recommendations that will help researchers make informed decisions when conducting peer nominations research or when considering to include peer nomination methodology in a new study. Our goal is to facilitate the sociometric data collection for both participants and researchers, to maximize the quality of peer nomination data, and to optimize the psychometric properties of sociometric data collected with youth. Toward this end, a number of specific methodological choices were identified.

Single vs. Multi-item Measures. One important choice regards the sociometric criteria, that is, the actual questions that are given to participants and in response to which they nominate peers. A question often asked by reviewers of manuscripts that include peer nominations is whether single-item measures of sociometric constructs are acceptable. For example, peer acceptance is typically measured with one item ("Who do you like most?"). For researchers not familiar with peer nomination research, the use of single-item measures may seem unusual and unreliable. In questionnaire research, it is typical that constructs are assessed by multiple items. Of course, the latter refers to self-report measures where there is only one respondent (the participant herself), whereas in the case of peer-report measures, reliability is obtained not from the fact that a construct is represented by multiple items but from the fact that there are multiple reporters (e.g., all classroom peers minus the participant herself, as self-nominations typically are discouraged in the instructions).

Nevertheless, a valid question is whether sociometric constructs should be assessed with more than one question (or sociometric criterion). The respective benefits and limitations associated with single- and multi-item peer nomination measures have been theoretical and have not been widely studied. Researchers have, however, demonstrated that intercorrelations between single- and multi-item measures are similar across studies, and some single-item measures have been shown to be both reliable and valid (Becker & Luthar, 2007; Parkhurst & Asher, 1992).

On a theoretical level, the main problem associated with single-item measures is one of content validity; that is, can we assess every necessary aspect of a construct within a single item? Because the most common behavior sets assessed by peer nominations (prosocial behavior, overt aggression, and relational aggression) involve multiple individual behaviors, single-item measures often use double- or triple-barreled questions (e.g., an overt aggression item asking participants who "starts fights, picks on, and teases" other kids is drawing for both physical and verbal behaviors). Additionally, as demonstrated by Babcock, Marks, Crick, and Cillessen (2014), individual nomination items vary considerably in terms of internal reliability; for constructs in which individual item reliability may be moderate (e.g., prosocial behavior and relational aggression), combining multiple items can greatly improve the reliability of measurement. For other constructs, such as overt aggression, the addition of a second item hardly contributes to the data because overt aggression is highly visible and noticed, and consensus within and between sources of information about overt aggression is high. In that case, adding a second item will only take additional data collection time that could better be devoted to the measurement of another construct. That construct could be relational aggression—as this is a more heterogeneous construct that is more difficult to observe. It is wise, if possible, to assess relational aggression with more than one peer nomination, whereas this may not be necessary for overt aggression.

Indeed, the major disadvantage of multi-item measures is that peer nominations, in general, are logistically challenging. For large-scale studies, single-item measures make it feasible for researchers to collect data on many constructs in a reasonable amount of time and with minimized fatigue for the respondents. From a psychometric perspective, multi-item measures are preferable, because they allow greater reliability and content validity than single-item alternatives. However, single-item measures have shown excellent utility (Becker & Luthar, 2007) and satisfactory psychometric properties (Babcock et al., 2014; Parkhurst & Asher, 1992; van den Berg & Cillessen, 2013) in previous research and can be used effectively in studies where logistical limitations restrict the use of multi-item measures.

Choice of the Reference Group. As indicated, the reference group is the peer group or social network within which status is determined. The choice for the reference group for a study is usually driven by considerations of ecological validity. What is the peer group within which the participants spend most of their time? Who are the peers they know well and relative to whom we would like to know their status given the goals of the study? Which peer group is most meaningful based on developmental considerations for the age group of the participants? In most applications in peer relations research, the reference group is the classroom or grade, but other possibilities are sports teams, hobby clubs, or all children in an after-school program. It is important to realize that peer status is always relative to the reference group in which it is assessed. In North America, the

reference group for kindergarten and elementary-school children is typically the classroom, for early adolescents in middle school it is all peers in their grade, and for adolescents in high school it is also the entire grade level or even all peers in school across grades. Although the latter may be highly ecologically valid, it is only infrequently applied because of the practical implications and difficulties (see, for an exception, Franzoi, Davis, & Vasquez-Suson, 1994). In other cultures with a different structure of secondary education, the proper reference group often continues to be the classroom even beyond elementary school. This is often the case, for example, in European secondary schools (see, e.g., van den Berg et al., 2015). Of course, a very important emphasis for future and further research continues to be the study of peer relationships in ecological contexts other than the classroom or school.

Inclusion vs. Exclusion of Nonparticipants as Nominees. In peer nomination research, maximizing the number of nominators is central to collecting reliable and valid data (Babcock et al., 2014; Crick & Ladd, 1989; Marks, Babcock, Cillessen, & Crick, 2013). Participants may be missing from data collection for a variety of reasons, including school absence, participation in other school activities, or lack of parental consent. Many constructs assessed by nominations are specific to the reference group of peers in which the data are being collected, and even random nonparticipation reduces the amount of data collected regarding each nominee. Unfortunately, nonparticipation is often systematic, with nonparticipants showing lower levels of status, social adjustment, and academic adjustment than participants in peer nomination research (Noll, Zeller, Vannatta, Bukowski, & Davies, 1997) as well as in school-based research more generally (Detty, 2013). Recently, Babcock, Marks, van den Berg, and Cillessen (2016) simulated systematic nonparticipation by removing nominators based on popularity and preference and found that systematic nonparticipation can have a major impact on the validity of peer nomination data, even at relatively low levels of missingness.

On the one hand, one question that researchers must ask when conducting peer nomination research is whether to include or exclude nonparticipants from peer nomination rosters. In instances of school absence or other cases in which participants miss the data collection sessions themselves, there is little reason that they should be excluded as nominees. However, when nonparticipants lack parental consent or individual assent, the inclusion of their names on rosters raises ethical concerns (Mayeux, Underwood, & Risser, 2007).

On the other hand, peer nominations are valid only insofar as the reference group is valid (Bronfenbrenner, 1943). Excluding nonparticipants as nominees changes the reference group and undermines both the external and internal validity of nominations. This is particularly true when nonparticipation is systematic; how can we trust nominations of social preference, status, and social behavior when the most rejected or least

popular or most aggressive participants are ineligible to receive nominations? Nominators are typically asked to make judgments about who "most" or "least" fits given criteria; when individuals are systematically excluded from consideration, nominators no longer have an adequate basis of comparison to make such judgments. Indeed, although published empirical data on this topic are lacking, preliminary data (using two large samples, which includes both simulated and nonsimulated participant missingness) indicated that excluding nonparticipants from rosters under conditions of systematic nonparticipation can greatly undermine the reliability and validity of nomination measures (Marks, Babcock, van den Berg, & Cillessen, 2017, April).

Ultimately, excluding nonparticipants as nominees violates a fundamental tenet of peer nomination research—that is, the assumption that nominators are choosing from, and making comparisons between, all of their peers within a given social context (the reference group). The ethical considerations are compelling and should not be ignored, but researchers (and ethical review boards) may need to balance ethical ambiguity with the cost of seriously compromising the validity of the data (see Mayeux et al., 2007).

Limited vs. Unlimited Nominations. Another issue to consider when using peer nominations is whether to collect limited or unlimited nominations. In the original sociometric studies with children in the 1940s and 1950s, both limited and unlimited nominations were used in different studies. Limited nominations were more common when conducting quantitative analyses, usually restricted to three of five, in part because of computational limitations that made "the mathematical manipulation of [unlimited nomination data] impossible" (Lemann & Solomon, 1952, pp. 15–16). The later studies by Coie et al. (1982) and Newcomb and Bukowski (1983) set the stage for asking exactly three nominations; this practice was then followed in many studies.

More recently, this issue has been revisited for multiple reasons. First, during the practice of sociometric data collection, participants often have indicated that they wanted to name fewer or more peers for a certain question. For example, to the "best friend" question, participants might indicate that they had more than three friends in their classroom and did not want to make a choice among them. In these cases, the nomination of an individual who appropriately fits an item's criterion is dependent on whether other students have been nominated for that criterion (Parkhurst & Asher, 1992). For more difficult and low-frequency behaviors such as peer victimization or social withdrawal, it is not uncommon for participants to indicate that they know only one or two children in their classroom to whom the question applies and cannot come up with a third. In those cases, allowing the number of nominations to be unlimited ("You can name as many or as few peers as you want for each question") rather than requiring a fixed number seems more ecologically valid. Second, limited nominations

inherently require that the average number of given nominations is the same for each construct within each reference group, which can obscure important differences in terms of the relative frequencies of characteristics and behaviors across different reference groups (Parkhurst & Asher, 1992). Finally, researchers increasingly are collecting sociometric data in larger reference groups, such as middle and high school grades. In those contexts in particular, nominations fixed to a small number can be problematic for participants. Based on these considerations, the use of unlimited nominations has increased, especially with adolescents in middle and high schools (see, e.g., Cillessen & Mayeux, 2004; Poulin & Dishion, 2008).

From a psychometric perspective, providing the opportunity to collect a higher number of nominations is clearly preferred over a smaller number. The classical measurement perspective, for example, would conceptually frame nominees as participants and nominators as items; as such, a small number of given nominations means that items have high statistical difficulty (Marks et al., 2013). As discussed in Babcock et al. (2014), a class of 31 children who each name three peers per question will have an average item difficulty of 10%; such difficult items generally have low variability, which results in reduced reliability and increased chance of Type II error.

There is also some empirical evidence that unlimited nominations have advantages. Indirect comparisons have shown that the stabilities of sociometric scores derived from unlimited nominations across multiple years (Cillessen & Borch, 2006; Cillessen & Mayeux, 2004) are at the high end of the distribution of stability coefficients found in a meta-analysis (Jiang & Cillessen, 2005). In direct comparison studies, unlimited nominations also yielded higher stability correlations with measures of social behavior than limited nominations (Terry, 2000), and unlimited friendship nominations were found to show higher internal reliability than limited friendship nominations did (Marks et al., 2013).

Recently, Gommans and Cillessen (2015) conducted a direct comparison study of limited and unlimited nominations with elementary-school children. They compared limited and unlimited nominations directly in a counterbalanced design with 8- to 12-year-old elementary school children. When examining the descriptive statistics, correlates, and predictors of sociometric status (acceptance and rejection), they found very similar results for nominations derived from either method. In general, unlimited nominations yielded better psychometric properties, especially for questions about sociometric status. However, for questions about bullying and victimization, limited nominations had some advantages. This may be caused by the fact that bullying and victimization are more "selective" questions, in the sense that there are fewer bullies and in the classroom than, for example, classmates children like or are their friends. This, in addition to the size of the reference group, the nature of the criterion (type of question) may also have an impact on whether a limited or an unlimited procedure is better.

Practical experience with sociometric data collection over many years and in many settings suggests that the benefits of unlimited nominations are particularly pronounced when the reference group is large, as in middle or high school grades, and that the benefits are smaller when the reference group is smaller, as in classrooms. However, this has not yet been directly investigated empirically and it should be noticed that the difference in reference group size (large middle or high school samples vs. smaller elementary school classrooms) is confounded with the age and development of the participants, and developmental differences in (social) cognitive capacities as well as relationships confound the comparison of smaller versus larger reference groups. This issue should be examined empirically before further recommendations can be made. One hypothesis for such research is that the correlation between sociometric scores derived from limited and unlimited nominations should be high in studies that collect nominations within elementary school classrooms.

The use of unlimited nominations when collecting sociometric data in middle or high schools is now very common (e.g., Poulin & Dishion, 2008). When using unlimited nominations, it may be wise to cap them to a maximum to prevent students becoming very unselective in their choices and simply naming "everyone" in response to a question. Franzoi et al. (1994) used this practice when they asked high school students to name peers from any grade in their school using unlimited nominations but capped them at a maximum of 10.

A theoretical advantage of using unlimited nominations is that they make it possible to study individual differences in what Moreno (1934) called "social expansiveness," that is, individual differences between persons in the number of peers they choose as friends, which might be an indicator of their peer sociability, social competence, social orientation, or the size of their social network. The other approach is to see individual differences in voter selectivity as a source of noise in the data that should be controlled. As a general rule, a vote by a less selective voter makes a smaller contribution to the data than a vote by a more selective voter. This is true even when simply counting nominations received. Differences in voter selectivity can be modeled with item response theory (Terry, 2000) or other mathematical procedures (DeRosier & Thomas, 2003).

How Choices Are Recorded. Another decision made by researchers collecting peer nomination data is the action that nominators are required to perform in identifying nominees. When collecting nominations using traditional, pen-and-paper questionnaires, these actions fall into three categories: (a) directly identifying peers on rosters, by marking checkboxes, circling names, etc.; (b) writing the full names of peers; or (c) writing out a code number associated with each peer's name. In the first of these methods, participants typically receive a roster of peers for each item. In the latter two methods, participants typically receive a single roster to use across items

(although the second option allows nominators to choose peers without a roster; e.g., Farmer, Hall, Petrin, Hamm, & Dadisman, 2010).

The choice of how nominators identify peers has both logistical and ethical implications. Providing a roster for each item necessarily increases paper use and printing costs. Asking participants to provide code numbers, on the other hand, maximizes the confidentiality of pencil-and-paper nominations by making it much more difficult to determine which peers are being nominated for each item. Code numbers also allow for easier data entry and lower the likelihood of transcription errors during the data entry process (although they could potentially increase the possibility of transcription errors by participants themselves during the nomination procedure).

One question that, to our knowledge, has never been investigated is the extent to which the method for identifying participants may affect the nominations provided. Certainly, each method listed previously requires a different amount of effort on the part of participants—circling names on a roster is quick and easy, whereas finding a name on a list, noting the accompanying code number, and then writing that code number by hand require more time and effort. Particularly when completing peer nomination questionnaires that contain large numbers of items, it is plausible that participants might name fewer peers when the method for identifying those peers is more onerous.

To investigate this possibility, we analyzed data from a longitudinal study that collected data from participants in 6th grade through 12th grade (more than 450 participants per year; Marks et al., 2013; Marks, Babcock, & Cillessen, 2015). In the middle school phase of this study, participants were provided with a roster of grademates for each item and asked to circle the names of nominees. In the high school phase, participants were provided with a single roster of grademates for all items and were asked to write out each nominee's name. The average number of nominations provided for friendship, peer acceptance, popularity, and overt aggression is presented for each year in Table 1.1. As the table shows, although given nomination counts were relatively consistent within the middle school years and within the high school years, participants provided substantially more nominations (over three times as many, in the case of overt aggression) in middle school than in high school. Clearly, the nomination methodology may not have been entirely responsible for this difference; importantly, participants were nested into two middle schools and then combined into a single high school, which may have affected the pattern of responses (although it seems unlikely that having longer rosters in high school would lead to fewer nominations). However, the dramatic difference in given nomination counts following a change in data collection methodology provides a preliminary indication that researchers should consider (and study) such methodological differences in the future. Considering that the number of nominations collected for a given item will affect that item's reliability, there is a strong psychometric case for making the data collection process as easy on

Table 1.1. Average Numbers of Given Nominations for Four Items in Middle School and High School

Grade	*Friends*	*Peer Acceptance*	*Popularity*	*Overt Aggression*
Middle School				
6th	13.0	13.9	8.3	6.8
7th	9.9	16.0	9.3	6.1
8th	10.3	17.6	10.3	6.4
High School				
9th	6.7	6.9	2.7	2.0
10th	6.1	6.4	3.1	2.2
11th	6.0	6.6	3.4	2.1
12th	5.7	6.8	3.1	1.5
MS Average	11.1	15.8	9.3	6.3
HS Average	6.1	6.7	3.1	2.0

Note: MS = middle school. HS = high school.

participants as possible (Marks et al., 2013). For pencil-and-paper measures, this may mean printing out a roster for each item and asking participants to circle names, but new technologies may offer even easier avenues for participants to provide peer nominations.

Computer-Based Data Collection. Among school-aged children, researchers increasingly collect sociometric nominations via computerized assessments (see also van den Berg & Gommans, this volume). This is less time consuming, costly, and error prone than paper-and-pencil assessments and makes it easier to store data confidentially (see, e.g., van den Berg & Cillessen, 2013). Computerized data collection of sociometric data affects several elements of the sociometric method: data collection, organization of the data collection, and data processing. Computerized assessment also facilitates randomization, which may aid in avoiding the name order effects (discussed later) found by Poulin and Dishion (2008) and Marks, Cillessen, and Babcock (2016). Computerized assessment makes it possible to fully randomize the order in which peers are presented to the participants.

As shown by Endedijk and Cillessen (2015), a further advantage of computerized assessment is its suitability for young children. Most children see a computer task as a game. This helps to sustain their interest and young children might find it easier to answer nonverbally (on the computer) than verbally (to the experimenter). The attractiveness of a computer task and the possibility to answer via a computer may make the assessment easier for young children.

An organizational advantage of computerized assessment is the efficiency of administration because data are stored automatically and printed questionnaires and manual data entry are no longer needed. Moreover, laptops and software can be used for multiple projects over multiple years.

Therefore, computerized assessment is efficient, especially for longitudinal or large-scale studies.

Computerized assessment also has advantages for data processing. Because the data are automatically stored as they are being collected, errors during manual data entry are eliminated. There is also an ethical advantage. In standard assessments, researchers typically maximize the confidentiality of children's answers by replacing names with code numbers and by keeping names and codes on separate sheets. On a computer, it is easy to directly save answers by codes only. For parents who are asked permission for participation it may help to know that code numbers and not names are saved in data files. Further, computerized assessment will offer new possibilities for the analysis of sociometric data that have not yet been considered. Computerized assessment yields the same data as standard assessment but will also afford new analyses of such response variables as response times, order of nominations (e.g., in what order do voters nominate peers?), and name/item order effects.

Open Questions and Challenges

The methodological issues presented above are actively investigated in ongoing studies. In addition, there are a number of other methodological issues that have not yet been resolved and should be addressed in further research.

Importance of Item Order. An interesting question that is occasionally asked is whether it matters in what order the questions on a sociometric test are asked and whether perhaps that order should be randomized between participants. The answer is that we do not really know and that sociometric questions are typically presented to all participants in a fixed order that is determined by researchers before the study. Researchers seem to use a few unwritten rules: place the sociometric questions at the beginning of the instrument to make sure all participants have the time to answer them, and place the reputational and behavioral items (for which a missing vote is less critical) after the sociometric questions; intersperse positive and negative questions; and make sure to end the list of sociometric questions with a positive question.

Other considerations may be made by researchers on an ad-hoc basis. Some researchers might choose to avoid putting similar questions close together because these questions may be confusing to a quick reader. On the other hand, item order could be used to emphasize differences between items. For example, because the term *popularity* is typically not defined for nominators but may be confused with liking or likability (Cillessen & Marks, 2011), including popularity and unpopularity questions immediately following acceptance and rejection questions may imply to participants that popularity is being assessed separately from individual feelings of liking and disliking.

Historically, the order of questions has remained uniform within a study and with paper versions of sociometric instruments, creating random orders of questions is essentially impossible for logistical reasons. Of course, with the availability of computer technology and the possibilities of online or other forms of computerized assessments, it would now be possible to experiment with order effects. This is interesting and may reveal important insights into peer nomination methodology, but possible order effects of the questions simply have not been addressed empirically in the literature.

Avoiding Name Order Effects. Two studies thus far (Marks et al., 2016; Poulin & Dishion, 2008) have demonstrated that individuals listed earlier on alphabetized rosters of same-grade peers were more likely to be nominated for various peer nomination items. In both studies, the effects of name order were small and significant for only a subset of variables, but the idea that rosters can introduce bias in the data collection, even when it is small, deserves our attention.

Clearly, further research is needed to clarify the extent of name order effects and to determine whether they affect peer nominations in smaller reference groups than have been considered so far. (Both Marks et al., 2016, and Poulin & Dishion, 2008, analyzed nominations based on rosters containing at least 190 names.) In the interim, however, researchers should take care to avoid order effects if possible. Randomizing name order on rosters (van den Berg & Cillessen, 2013) or eliminating rosters entirely (Farmer et al., 2010) are options but would make it more onerous for participants to provide nominations, particularly when rosters are large and many items are being assessed. Poulin and Dishion (2008) suggested counterbalancing name order on rosters and recommended that participants be directed to think about relevant peers for each item before consulting the roster; both approaches would be expected to mitigate, but not eliminate, name order effects.

Most important, any future researchers using alphabetized rosters should take care to assess possible name order effects. They should report the results of such assessments (whether significant or nonsignificant), and (if they are significant) control for name order in subsequent analyses (Marks et al., 2016).

Issues in Non-English-Language Item Construction. Peer relationships research has become increasingly globalized in recent decades, with peer nomination research programs being explored for the first time in countries in Africa, South America, and Southeast Asia. Unfortunately, this research has not seemed to lead to a corresponding increase in cross-cultural considerations of item translation, and methods sections of most peer nomination studies in non-English samples provide English wordings of nomination items without providing any discussion of how the items were translated.

One exception to this trend has followed the concurrent explosion of research on popularity, which is seen as having a very specific, ecologically

valid meaning among English-speaking children and adolescents. Interestingly, some translations of the term popularity in other languages, countries, and cultures do not have the same denotations and connotations as the English term. Dutch adolescents, for example, use two different terms to denote different types of popular peers (de Bruyn & Cillessen, 2006). Research in Chinese populations has utilized at least three different translations for popularity, none of which entirely corresponds to the English-language construct (Niu, Jin, Li, & French, 2016).

Translations of "popularity" are particularly worth problematizing because the term has such a specific meaning and usage among English-speaking child and adolescent populations. But even direct translations of items measuring discrete, overt behaviors may lack ecological validity in other cultures. For example, Hecht and colleagues qualitatively asked Ugandan adolescents how individuals of each sex might aggress against peers (Hecht, Ralston, Crick, & Cicchetti, 2013). Results included typical Western examples of physical and relational aggression but also acts such as convincing them to take drugs, burning their homes, and (in the case of male aggression against females) giving them HIV or making them pregnant so they have to drop out of school. Such responses have not been seen in qualitative assessments of aggression in Western samples; as such, direct translations of aggression questionnaires in Uganda may not be providing ecologically valid measures.

These sorts of translation problems can be serious. If a study in China finds that the association between popularity and aggression differs between Chinese adolescents and American adolescents, is it because the association between status and aggression is different in China than in America or is it because the Chinese study's items were inadvertently measuring different constructs? There are important issues to be addressed in what it is hoped will be a growing number of cross-cultural studies using peer nomination methodology.

Assessment and Reporting of Psychometric Properties

Because sociometric and peer assessment methods do not follow the usual format of traditional multi-item self-report scales, it is often unclear how to determine their psychometric properties, and how they should be reported. However, reliability, validity, and other statistical properties can be reported easily for peer nomination data, and should be reported more consistently than is currently the case.

Reliability. Peer nomination measures are somewhat unique among pencil-and-paper measures with standardized scores in that researchers have generally not been able to report associated required reliability metrics. Many peer nomination constructs have been found to be moderately or highly stable (e.g., Jiang & Cillessen, 2005), but even recent longitudinal

studies using nominations have typically reported stability only incidentally as part of longitudinal analyses, if at all.

Although investigations of internal reliability of peer nominations were commonplace half a century ago (see Gronlund, 1959), an unspoken consensus seems to have emerged among researchers that reporting internal reliability is either inappropriate or unimportant. However, our research team has recently been using internal reliability metrics to evaluate the psychometric properties of peer nomination measures, including the statistical impact of low participation rates (Marks et al., 2013), the advantage of multi-item nomination measures (Babcock et al., 2014), and the potential utility of using a small subsample of "expert nominators" for providing sociometric data (Marks et al., 2015). For these types of analyses, peer nomination data are arranged in a binary matrix with nominators on the columns (acting as items, psychometrically speaking) and nominees on the rows (acting as participants). Thus, internal reliability metrics (e.g., KR-20 or Cronbach's alpha) indicate the degree of consensus among the peer group (nominators) regarding the sociometric criteria (Perry, Kusel, & Perry, 1988), and internal reliability can be calculated even for single-item nomination measures.

Our results also show that internal reliability and stability of peer nominations are highly related. For example, using data from the 6th-grade through 12th-grade waves of the same longitudinal study used for Marks et al. (2013, 2015), we conducted a series of multiple regression analyses, each using item stability across 2 consecutive years as the criterion variable and the items' KR-20 values of those 2 years as the predictor variables (mean KR-20 values ranged from .77 to .88 across different years). For the middle school years (grades 6 to 7 and 7 to 8), separate analyses were conducted for both middle schools. Results of these regression analyses are presented in Table 1.2. Across variables, internal reliability consistently accounted for a substantial proportion of the variance in stability (between 42% and 89%).

Although reliability can be assessed for multi-item nomination measures by comparing computed scores across items, the resulting correlation coefficients or alphas are using the nomination question as the unit of analysis rather than the participant (see Babcock et al., 2014, for a discussion). This is akin to assessing parallel forms reliability when interitem reliability can be easily assessed. In general, reliability should be assessed at the finest level possible.

Both stability and internal reliability metrics provide an indication of the amount of systematic variance and measurement error in the data. Ultimately, there is no reason that peer nomination researchers should be exempt from reporting reliability metrics in their reports. Indeed, reporting reliability information allows researchers (and reviewers) to evaluate factors such as the strength of items or potential problems with measure administration; such evaluations are nearly impossible when researchers fail to report the psychometric properties of their items.

Table 1.2. Descriptive Statistics for Stability and Results of Regressions Predicting Stability from KR-20

		Stability		
	N of items	*M*	*SD*	*Regression R^2 Predicting Stability from KR-20*
6th → 7th				
School 1	16	.63	(.21)	.42
School 2	16	.67	(.16)	.44
7th → 8th				
School 1	20	.78	(.12)	.73
School 2	20	.80	(.07)	.68
8th → 9th	20	.69	(.13)	.76
9th → 10th	18	.75	(.13)	.89
10th → 11th	20	.72	(.14)	.77
11th → 12th	21	.74	(.13)	.77

Note: Stability was measured as the correlation between nomination counts across consecutive years. Multiple regression analyses were conducted by predicting stability between any 2 consecutive years from the KR-20 internal reliability values in both years. Mean KR-20 values across items varied from .77 to .88 (*SD*s ranged from .06 to .15) across years.

Validity. Several peer nomination measures that were developed decades ago, such as the Revised Class Play (Masten, Morison, & Pellegrini, 1985) and the Peer Nomination Inventory (Perry et al., 1988), are still being used today and have been well validated over the years (although both have been occasionally criticized; e.g., Parkhurst & Asher, 1992). Nomination items for some newer constructs, such as relational aggression (Crick & Grotpeter, 1995) and popularity (Parkhurst & Hopmeyer, 1998), were also established with careful consideration and assessment of validity; in fact, Crick and Grotpeter's measure of physical aggression, relational aggression, and prosocial behavior is one of the most common methods for assessing those constructs today.

On the other hand, some research is conducted with central variables measured by items that do not appear to have been validated. For example, recent studies have used items assessing gender noncomformity (Toomey, Card, & Casper, 2014), electronic aggression (Badaly, Kelly, Schwartz, & Dabney-Lieras, 2013), and academically oriented behavior (Torrente, Cappella, & Neal, 2014) without demonstrating that the items are assessing the targeted constructs or providing references to previous nomination studies using the items. This is not to say that these items are not valid (indeed, asking participants to identify "students that use the internet or text messages to insult or be mean" certainly appears to be a face-valid way of measuring electronic aggression; Badaly et al., p. 895) but rather that face validity alone does not sufficiently justify the use of measures in any study. Although there is a long history of accepting peer nomination items on the basis of face validity (see Lindzey & Borgatta, 1954), this convention allows peer

nomination researchers to publish results based on looser psychometric constraints than in other fields of behavioral research. In conducting studies with other pencil-and-paper measures, researchers are typically required to establish empirical evidence of validity; peer nomination researchers should hold ourselves to the same standard.

Statistical Properties. Peer nomination measures frequently violate the assumptions of the most commonly used statistical tests. In particular, peer nomination data are commonly positively skewed and may be markedly zero-inflated, especially when assessing negative traits (see Avant, Gazelle, & Faldowski, 2011), when using single-item measures, and when using limited nomination procedures. It is vital that researchers assess, report, and (if possible) statistically account for violations of statistical assumptions when conducting research using peer nomination data. In particular, the issue of skewness of sociometric scores deserves further attention, as does the issue of the nestedness of sociometric data. So far, the nestedness of sociometric data has been treated as noise to be controlled for through sociometric standardization, but it could be studied as an important issue in and of itself as researchers are increasingly interested in group-level effects (in particular grade or school, e.g., Boor-Klip, Segers, Hendrickx, & Cillessen, 2016; Hendrickx, Mainhard, Boor-Klip, Cillessen, & Brekelmans, 2016). In particular, the use of multilevel modeling instead of sociometric standardization is worth exploring and further developing. Of course, when applying such advanced statistical methods, the distributional properties of the underlying data become even more central and deserving of careful attention.

Differences Between Participants and Nonparticipants. As discussed earlier, systematic nonparticipation can undermine the reliability and validity of peer nomination measures, and research has shown that nonparticipants are often lower in status and adjustment than participants (e.g., Noll et al., 1997). Peer nomination procedures are unique among methodologies in that, assuming nonparticipants are included as potential nominees on rosters, researchers can make valid comparisons between participants and nonparticipants on any construct assessed through nominations (rather than simply being limited to comparing demographic characteristics of the sample to those of the population).

Despite the convenience of comparing participants to nonparticipants, however, most peer nomination researchers either fail to assess these comparisons or fail to report them. Given the potential problems associated with systematic nonparticipation, and given the ease of conducting and reporting the results of comparisons (assuming no differences, the results can be summed up in a single sentence), this information should be included in published peer nomination research, so that it can be subjected to systematic (meta-analytic) review that can form the basis for further recommendations.

Conclusion

In this article, we have attempted to break down the major choices facing researchers interested in conducting peer nomination research, and we have attempted to provide a sense of (a) what researchers commonly do when collecting data, (b) "best practices" for optimal data collection, and (c) other justified choices in data collection. In many cases, the "best practice" from a psychometric perspective may not be the most logistically practical option; for example, although reliability and validity of measures are increased by using multiple items to assess each construct, single-item measures may be more realistic in studies that are collecting data on a large number of variables. In other cases, the "best practice" may conflict with ethical considerations. For example, using passive consent procedures may increase participation rates, and allowing nonparticipants to be included on nomination rosters may increase validity, but both methodological practices can raise ethical concerns (and may be directly prohibited by schools or ethical review boards). In the cases in which there are conflicts between ethical considerations, logistical issues, and the maximization of data quality, researchers must be especially intentional about their methodological choices and should acknowledge any limitations of the methods they have chosen.

In looking at the recommendations we have provided in this article, two major themes come to the fore. The first is the need to maximize the number of meaningful nominations collected from participants. This involves maximizing participation rates and particularly reducing the extent to which nonparticipation is systematic. It also involves increasing the number of nominations that each nominator provides, through methodological choices such as allowing unlimited nominations, making the nomination procedure itself less onerous for participants, and choosing the largest ecologically valid reference group possible. Obviously, not all methods for increasing nomination counts are appropriate—requiring each nominator to name 10 peers for each item would be unlikely to increase the quality of the data. However, increasing the number of meaningful nominations increases reliability and validity and will generally improve the quality of statistical analyses by decreasing the degree to which nomination data are skewed or zero-inflated.

The second major theme in this article concerns reporting information about peer nominations. Currently, most journal articles include details of the data collection procedure itself but often exclude information about reliability and validity, comparisons between participants and nonparticipants, and the degree to which data are normally distributed. Reporting this information allows for better comparisons across studies and a better understanding of limitations within our field. Furthermore, requiring ourselves to provide such information is a matter of holding ourselves to the same standards as those required of researchers in other areas of psychology.

Sociometric methods continue to change with a focus on different types and numbers of peer nominations and the increasing use of information technology to assess peer status and the affiliations of individuals in social groups and networks. Sociometry may be seen as a tool toward an end but also offers many opportunities to address important substantive questions about the nature of child and adolescent peer relationships. Sociometric methodology in itself is a fascinating area of research and a developing field with continuous improvements. Much work remains to be done in this area that can lead to the further improvement of our methods and to the further development of best practices and methodological recommendations.

References

Avant, T. S., Gazelle, H., & Faldowski, R. (2011). Classroom emotional climate as a moderator of anxious solitary children's longitudinal risk for peer exclusion: A child × environment model. *Developmental Psychology*, *47*, 1711–1727. https://doi.org/10.1037/a0024021

Babad, E. (2001). On the conception and measurement of popularity: More facts and some straight conclusions. *Social Psychology of Education*, *5*, 3–29.

Babcock, B., Marks, P. E. L., Crick, N. R., & Cillessen, A. H. N. (2014). Limited nomination reliability using single- and multiple-item measures. *Social Development*, *23*, 518–536. https://doi.org/10.1111/sode.12056

Babcock, B., Marks, P. E. L., van den Berg, Y. H. M., & Cillessen, A. H. N. (2016). Implications of systematic nominator missingness for peer nomination data. *International Journal of Behavioral Development*. https://doi.org/10.1177/0165025416664431

Badaly, D., Kelly, B. M., Schwartz, D., & Dabney-Lieras, K. (2013). Longitudinal associations of electronic aggression and victimization with social standing during adolescence. *Journal of Youth and Adolescence*, *42*, 891–904. https://doi.org/10.1007/s10964-012-9787-2

Becker, B., & Luthar, S. S. (2007). Peer-perceived admiration and social preference: Contextual correlates of positive peer regard among suburban and urban adolescents. *Journal of Research on Adolescence*, *17*, 117–144. https://doi.org/10.1111/j.1532-7795.2007.00514.x

Boor-Klip, H. J., Segers, E., Hendrickx, M. M. H. G., & Cillessen, A. H. N. (2016). Development and psychometric properties of the Classroom Peer Context Questionnaire. *Social Development*, *25*, 370–389. https://doi.org/10.1111/sode.12137

Bronfenbrenner, U. (1943). A constant frame of reference for sociometric research. *Sociometry*, *6*, 363–397.

Bukowski, W. M., Laursen, B., & Rubin, K. H. (Eds.). (in press). *Handbook of peer interactions, relationships, and groups* (2nd ed.). New York: Guilford.

Cillessen, A. H. N., & Borch, C. (2006). Developmental trajectories of adolescent popularity: A growth curve modelling analysis. *Journal of Adolescence*, *29*, 935–959.

Cillessen, A. H. N., & Bukowski, W. M. (in press). Sociometric perspectives. In W. M. Bukowski, B. Laursen, & K. H. Rubin (Eds.), *Handbook of peer interactions, relationships, and groups* (2nd ed.). New York: Guilford.

Cillessen, A. H. N., & Marks, P. E. L. (2011). Conceptualizing and measuring popularity. In A. H. N. Cillessen, D. Schwartz, & L. Mayeux (Eds.), *Popularity in the peer system* (pp. 25–56). New York: Guildford Press.

Cillessen, A. H. N., & Mayeux, L. (2004). From censure to reinforcement: Developmental changes in the association between aggression and social status. *Child Development*, *75*, 147–163.

Clarke, K. E., & Ladd, G. W. (2000). Connectedness and autonomy support in parent–child relationships: Links to children's socio-emotional orientation and peer relationships. *Developmental Psychology*, *36*, 485–498. https://doi.org/10.1037/0012-1649.36.4.485

Coie, J. D., & Dodge, K. A. (1988). Multiple sources of data on social behavior and social status in the school: A cross-age comparison. *Child Development*, *59*, 815–829. https://doi.org/10.2307/1130578

Coie, J. D., & Dodge, K. A., & Coppotelli, H. (1982). Dimensions and types of social status: A cross-age perspective. *Developmental Psychology*, *18*, 557–570. https://doi.org/10.1037/0012-1649.18.4.557

Coleman, J. (1961). *The adolescent society*. New York, NY: Free Press.

Crick, N. R., & Grotpeter, J. (1995). Relational aggression, gender, and social-psychological adjustment. *Child Development*, *66*, 710–722. https://doi.org/10.2307/1131945

Crick, N. R., & Ladd, G. W. (1989). Nominator attrition: Does it affect the accuracy of children's sociometric classifications? *Merrill-Palmer Quarterly*, *35*, 197–207.

de Bruyn, E. H., & Cillessen, A. H. N. (2006). Popularity in early adolescence: Prosocial and antisocial subtypes. *Journal of Adolescent Research*, *21*, 607–627. https://doi.org/10.1177/0743558406293966

DeRosier, M. E., & Thomas, J. M. (2003). Strengthening sociometric prediction: Scientific advances in the assessment of children's peer relations. *Child Development*, *74*, 1379–1392.

Detty, A. M. R. (2013). School-based survey participation: Oral health and BMI survey of Ohio third graders. *Maternal and Child Health Journal*, *17*, 1208–1214. https://doi.org/10.1007/s10995-012-1107-7

Endedijk, H. M., & Cillessen, A. H. N. (2015). Computerized sociometric assessment for preschool children. *International Journal of Behavioral Development*, *39*, 383–388. https://doi.org/10.1177/0165025414561706

Endedijk, H. M., Cillessen, A. H. N., Cox, R. F. A., Bekkering, H., & Hunnius, S. (2015). The role of child characteristics and peer experiences in the development of peer cooperation. *Social Development*, *24*, 521–540. https://doi.org/10.1111/sode.12106

England, E. M., & Petro, K. D. (1998). Middle school students' perceptions of peer groups: Relative judgments about group characteristics. *Journal of Early Adolescence*, *18*, 349–737. https://doi.org/10.1177/0272431698018004002

Farmer, T. W., Hall, C. M., Petrin, R., Hamm, J. V., & Dadisman, K. (2010). Evaluating the impact of a multicomponent intervention model on teachers' awareness of social networks at the beginning of middle school in rural communities. *School Psychology Quarterly*, *25*, 94–106. https://doi.org/10.1037/a0020147

Franzoi, S. L., Davis, M. H., & Vasquez-Suson, K. A. (1994). Two social worlds: Social correlates and stability of adolescent status groups. *Journal of Personality and Social Psychology*, *67*, 462–473.

Gommans, R., & Cillessen, A. H. N. (2015). Nominating under constraints: A systematic comparison of unlimited and limited peer nomination methodologies in elementary school. *International Journal of Behavioral Development*, *39*, 77–86. https://doi.org/10.1177/0165025414551761

Gronlund, N. E. (1959). *Sociometry in the classroom*. New York: Harper & Brothers.

Hecht, K., Ralston, P., Crick, N. R., & Cicchetti, D. (2013, April). Peer relationships and psychopathology in war-affected Ugandan youth. In K. A. Dodge (Chair), *Honoring the legacy of Nicki R. Crick*. Paper symposium presented at the biennial meeting of the Society for Research in Child Development, Seattle, WA.

Hendrickx, M. M. H. G., Mainhard, M. T., Boor-Klip, H. J., Cillessen, A. H. N., & Brekelmans, M. (2016). Social dynamics in the classroom: Teacher support and conflict and the peer ecology. *Teaching and Teacher Education*, *53*, 30–40.

Jiang, X. L., & Cillessen, A. H. N. (2005). Stability of continuous measures of sociometric status: A meta-analysis. *Developmental Review*, *25*, 1–25. https://doi.org/10.1016/j.dr.2004.08.008

Keislar, E. R. (1957). The validity of the guess-who technique with large adolescent populations. *Journal of Genetic Psychology*, *91*, 131–135. https://doi.org/10.1080/00221325.1957.10533037

Lansu, T. A. M., & Cillessen, A. H. N. (2012). Peer status in emerging adulthood: Associations of popularity and preference with social roles and behavior. *Journal of Adolescent Research*, *27*, 132–150.

Lemann, T. B., & Solomon, R. L. (1952). Group characteristics as revealed in sociometric patterns and personality ratings. *Sociometry*, *15*, 7–90.

Lindzey, G., & Borgatta, E. (1954). Sociometric measurement. In G. Lindzey (Ed.), *Handbook of social psychology* (pp. 405–448). Cambridge, MA: Addison-Wesley.

Maassen, G. H., van Boxtel, H W., & Goossens, F. A. (2005). Reliability of nominations and two-dimensional rating scale methods for sociometric status determination. *Journal of Applied Developmental Psychology*, *26*, 51–68. https://doi.org/10.1016/j.appdev.2004.10.005

Marks, P. E. L., Babcock, B., & Cillessen, A. H. N. (2015). On the empirical identification and evaluation of "expert nominators." *International Journal of Behavioral Development*, *39*, 186–193. https://doi.org/10.1177/0165025414556518

Marks, P. E. L., Babcock, B., Cillessen, A. H. N., & Crick, N. R. (2013). The effects of participation rate on the internal reliability of peer nomination measures. *Social Development*, *22*, 609–622. https://doi.org/10.1111/j.1467-9507.2012.00661.x

Marks, P. E. L., Babcock, B., van den Berg, Y. H. M., & Cillessen, A. H. N. (2017, April). Effects of including versus excluding nonparticipants as potential nominees in peer nomination measures. In P. E. L. Marks (Chair), *Topics in sociometric methodology and analysis*. Paper symposium presented at the biennial meeting of the Society for Research in Child Development, Austin, TX.

Marks, P. E. L., Cillessen, A. H. N., & Babcock, B. (2016). Relations between alphabetized name order and nomination counts in peer nomination measures. *Social Development*, *25*, 866–874. https://doi.org/10.1111/sode.12163

Masten, A. S., Morison, P., & Pellegrini, D. (1985). A revised class play method of peer assessment. *Developmental Psychology*, *21*, 523–533. https://doi.org/10.1037/0012-1649.21.3.523

Mayeux, L., Underwood, M. K., & Risser, S. D. (2007). Perspectives on the ethics of sociometric research with children: How children, peers, and teachers help to inform the debate. *Merrill-Palmer Quarterly*, *53*, 53–78. https://doi.org/10.1353/mpq.2007.0002

Moreno, J. L. (1934). *Who shall survive?* Washington, DC: Nervous and Mental Disease Publishing Company.

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.

Niu, L., Jin, S., Li, L., & French, D. C. (2016). Popularity and social preference in Chinese adolescents: Associations with social and behavioral adjustment. *Social Development*, *25*, 828–845. https://doi.org/10.1111/sode.12172

Noll, R. B., Zeller, M. H., Vannatta, K., Bukowski, W. M., & Davies, W. H. (1997). Potential bias in classroom research: Comparison of children with permission and those who do not receive permission to participate. *Journal of Clinical Child Psychology*, *26*, 36–42. https://doi.org/10.1207/s15374424jccp2601_4

Parkhurst, J. T., & Asher, S. R. (1992). Peer rejection in middle school: Subgroup differences in behavior, loneliness, and interpersonal concerns. *Developmental Psychology*, *28*, 231–241. https://doi.org/10.1037/0012-1649.28.2.231

Parkhurst, J. T., & Hopmeyer, A. (1998). Sociometric popularity and peer-perceived popularity: Two distinct dimensions of peer status. *Journal of Early Adolescence*, *18*, 135–144. https://doi.org/10.1177/0272431698018002001

Pellegrini, A. D., Roseth, C. J., Mliner, S., Bohn, C. M., Ryzin, M. V., Vance, N., … Tarullo, A. (2007). Social dominance in preschool classrooms. *Journal of Comparative Psychology*, *121*, 54–64. https://doi.org/10.1037/0735-7036.121.1.54

Perry, D. G., Kusel, S. J., & Perry, L. C. (1988). Victims of peer aggression. *Developmental Psychology*, *24*, 807–814. https://doi.org/10.1037/0012-1649.24.6.807

Poulin, F., & Dishion, T. J. (2008). Methodological issues in the use of peer sociometric nominations with middle school youth. *Social Development*, *17*, 908–921. https://doi.org/10.1111/j.1467-9507.2008.00473.x

Rubin, K. H., & Cohen, J. S. (1986). The revised class play: Correlates of peer assessed social behaviors in middle childhood. *Advances in Behavioral Assessment of Children and Families*, *2*, 176–206.

Terry, R. (2000). Recent advances in measurement theory and the use of sociometric techniques. In A. H. N. Cillessen & W. M. Bukowski (Eds.), *New Directions for Child and Adolescent Development: No. 88. Recent advances in the measurement of acceptance and rejection in the peer system* (pp. 27–53). San Francisco: Jossey-Bass. https://doi.org/10.1002/cd.23220008805

Terry, R., & Coie, J. D. (1991) A comparison of methods for defining sociometric status among children. *Developmental Psychology*, *27*, 867–880. https://doi.org/0012-1649.27.5.867

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. https://doi.org/10.1177/001316445101100315

Toomey, R. B., Card, N. A., & Casper, D. M. (2014). Peer perceptions of gender nonconformity: Associations with overt and relational peer victimization and aggression in early adolescence. *Journal of Early Adolescence*, *34*, 463–485. https://doi.org/10.1177/0272431613495446

Torrente, C. E., Cappella, E., & Neal, J. W. (2014). Children's positive school behaviors and social preference in urban elementary classrooms. *Journal of Community Psychology*, *42*, 143–161. https://doi.org/10.1002/jcop.21599

van den Berg, Y. H. M., & Cillessen, A. H. N. (2013). Computerized sociometric and peer assessment: An empirical and practical evaluation. *International Journal of Behavioral Development*, *37*, 68–76. https://doi.org/10.1177/0165025412463508

van den Berg, Y. H. M., Lansu, T. A. M., & Cillessen, A. H. N. (2015). Measuring social status and social behavior with peer and teacher nomination methods. *Social Development*, *24*, 815-832. https://doi.org/10.1111/sode.12120

ANTONIUS H. N. CILLESSEN is Professor and Chair of Developmental Psychology in the Behavioural Science Institute at Radboud University, The Netherlands. He received his Ph.D. in developmental psychology from Radboud University.

PETER E. L. MARKS is Associate Professor in the Department of Psychology at Austin College. He received his Ph.D. from the Institute of Child Development at the University of Minnesota.