The following manuscript represents the accepted version of the article published as:

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.*

Please note that, as a result of copy editing, minor differences may be evident between this manuscript draft and the final published article.

On the Empirical Identification and Evaluation of “Expert Nominators”

Peter E. L. Marks

Austin College

Ben Babcock

The American Registry of Radiologic Technologists

Antonius H. N. Cillessen

Radboud Universiteit Nijmegen

Author Note

Peter E. L. Marks, Department of Psychology, Austin College; Ben Babcock, American Registry of Radiologic Technologists (ARRT); Antonius H. N. Cillessen, Behavioural Science Institute, Radboud Universiteit Nijmegen.

The views and discussions presented in this research are not necessarily the official views of ARRT. Data presented in this article was collected with support from a grant awarded by the University of Connecticut Research Foundation to Antonius H. N. Cillessen.

Correspondence should be addressed to Peter E. L. Marks, Department of Psychology, Austin College, 900 N. Grand Ave., Sherman, TX 75090. Phone: 903-813-2017. Email: pmarks@austincollege.edu

Abstract

The current study aims to evaluate and empirically investigate Prinstein’s (2007) conclusions regarding “expert nominators,” a subsample of individuals in a peer group whose peer nominations might substitute for nominations from the full sample. The current study empirically identified experts based on *comparative accuracy*, the extent to which each participant’s nominations matched full sample nominations across items. 8th graders at two schools (*N*s = 273 and 334) completed 16 nomination items. Participants were labeled “experts” if they showed above-average comparative accuracy on at least 75% of items; however, expert nominations were neither highly internally reliable nor valid. Nominations from experts would not have adequately substituted for collecting full-sample nominations. Future research may possibly benefit from identifying more limited, single-item experts.

Keywords: sociometric methods, sociometry, expert nominators, peer nominations

On the Empirical Identification and Evaluation of “Expert Nominators”

In peer nomination research with adolescents, obtaining high participation rates can be a challenge. Unfortunately, there are several problems associated with low participation rates. First, participants who do participate may systematically differ from nonparticipants in terms of peer status (Fournier, 2009) or demographics (Brown & Larson, 2009), which may decrease the validity of the data. Second, even random nonparticipation can undermine the validity of nominations (Crick & Ladd, 1990). Third, the internal reliability of sociometric measures is substantially lower when few adolescents provide nominations (Babcock, Marks, Crick, & Cillessen, in press; Marks, Babcock, Cillessen, & Crick, 2013). Thus, high participation rates are vital for sociometric research, but often difficult to obtain. A promising solution is to identify a subgroup of adolescents who supply reliable and valid sociometric data. They will be referred to as “expert nominators,” due to their presumed ability to provide accurate peer group data.

The use of expert nominators was originally investigated by Prinstein (2007). Prinstein used correlations to conclude that data derived from teacher-identified experts (approximately 9% of all participants) was similar to data derived from the full sample for popularity and preference nominations. For example, correlations between full-sample and expert-sample nomination counts were .89 for popularity and .63 for peer acceptance/liking. Unfortunately, the degree of correspondence between subsample and full-sample data may have been inflated (Marks et al., 2013), due to the item-overlap correlation (i.e., the fact that the subsample participants were also included in the full sample; Guilford, 1936; Hsu, 1992). Even if participants had been responding *randomly*, item overlap would have yielded a correlation of over .30 between full sample and expert nominations (based on Equation 7 in Hsu, 1994). Thus, more research is needed to determine whether data from a subsample of expert nominators can accurately substitute for data derived from a large proportion of participants.

If expert nominators exist, their identification would constitute a massive boon to peer relationships research. The use of expert nominators would make it possible to obtain valid data from everyone in the reference group via participation of a minority of individuals in the group. However, if expert nominators are illusory, previous research supporting the utility of experts might inaccurately justify small participation rates in sociometric research (e.g., Fournier, 2009; Walcott, Upton, Bolen, & Brown, 2008). The large potential benefits of verifying the expertise of some nominators, as well as the large potential costs of inappropriately using expert nominators in sociometric research, motivated the current investigation.

The Current Study

Identifying experts. The overall aim of this study was to replicate and expand the findings of Prinstein (2007) by identifying expert nominators and evaluating the reliability and validity of their nominations. Although we operationalized expertise differently than Prinstein (based on empirical comparison of all participants, rather than teacher identification), we conceptualized experts similarly—adolescents who can accurately report on peers in terms of reputation, behavior, and affective information (e.g., liking and disliking).

Evaluating expert reliability. Once experts have been identified, the next goal of this study was to assess the internal reliability of nominations by experts. Internal reliability (measured by Cronbach's α) indicates the extent to which nominators' choices agree. We expected that expert nominators would provide reliable nomination data for a variety of peer nomination constructs. Optimal reliability levels for the expert sample should approach $\alpha = .80$. We selected .80 because it generally yields good measurement across numerous research contexts (Crocker & Algina, 1986); lower reliabilities imply that the variation in scores is greater than 20% error, which is unacceptable in many research contexts. The current study's use of internal reliability to evaluate nominations differs from Prinstein's (2007) strategy, which involved evaluating test-retest reliability (for expert nominations only) over a two-week period. In Prinstein's study, same-construct test-retest correlations were below .65 for both liking and disliking, indicating substantial inconsistency over this fairly short time frame.

Although we sought a minimum alpha of .80 for optimal reliability, previous findings (Babcock et al., in press; Marks et al., 2013) imply that this cutoff might be unrealistic for some variables. Marks et al. (2013) found that data from small proportions of participants (10-20%) were unreliable when measuring some constructs (i.e., friendship and peer acceptance), but reliable for other constructs (i.e., popularity and overt aggression) in *some* of the sampled schools. Thus, we also compared alphas between the expert sample and the full sample. Even if the experts did not reach the .80 cutoff, the use of experts might be supported if their reliabilities were not significantly lower than those for the whole group and certainly significantly higher in reliability than a randomly chosen group of raters of the same size.

Evaluating expert validity. The final goal of this study was to determine the validity of nominations derived from expert nominators. We correlated nomination counts from the expert sample with those from the full sample (adjusting for item overlap); these correlations should be high (greater than .85) if expert nominators are valid substitutes for large samples of adolescents. Recall that a correlation can be obtained as the cosine of an angle, $r = \cos(\theta)$, where θ is the angle between vectors representing the variables in space. Suppose that a peer nomination item and an external variable correlated .95. If the expert nominators correlated .85 with the full sample item, the range of correlations between the expert nominations and that external variable using the cosine relationship is $.64 \leq r \leq .97$. Even with high correlations, the range of potential correlations with an external variable is over .3 wide on a correlation scale. We determined that the potential correlation range should not be wider than that, so we settled on a correlation of .85 for this validity criterion. Our method for establishing validity is similar to Prinstein (2007), in that we are correlating the expert nominations to the whole group; however, our criterion is more conservative and we will statistically correct for item overlap. Although Prinstein did not specify an *a priori* cutoff for validity correlations, coefficients of around .60 were concluded to be valid. We felt that these correlations were too low to indicate validity, and the potential range of correlations with an external variable should be minimized for experts to replace the full sample.

To summarize, we believe that expert nominators should provide comparatively accurate nominations across a variety of social constructs, and that these nominations should be internally reliable and overlap substantially with full-sample nomination counts. For the purposes of this study, all three criteria (comparative accuracy, reliability, validity) must be met in order to support the utility of expert nominators as sole sources of nomination data in future research.

Method

Participants

The sample included 607 students participating in the 8th grade phase of a longitudinal study of peer relationships. The sample was 49.1% female, and was primarily White/Caucasian (68.9%), with minorities of Black/African-American (17.6%) and Hispanic/Latino (11.5%) students. The 607 participants represented the entire 8th grade population of a medium-sized town in the Northeastern United States and came from two middle schools; School 1 enrolled 273 8th graders; School 2 enrolled 334. Both schools were mostly lower- and middle-class, and the gender and ethnic breakdowns of the two samples were similar. Completion rates of peer nomination measures were 91.9% (N = 251) at School 1 and 90.7% (N = 303) at School 2.

Procedure and Measure

Students were provided with a roster of all same-grade peers of both genders for each item, and were asked to circle an unlimited number of names of peers who fit the description provided. Schools were structured such that 8th graders rotated among different classrooms throughout the day, and most students within each school had been attending that school since the 6th grade. Thus, although participants could not have been expected to personally know all of their peers, they had experiences and interactions with a large proportion of their grade-mates.

Sociometric items included (in order presented to participants): *friendship* (best friends), *acceptance/liking* (like most), *rejection/disliking* (like least), *received attention* (others pay attention to or go along with), *overt aggression* (start fights, pick on, tease), *withdrawal* (stay by themselves), *relational victimization 1* (get left out of group activities when others are mad at them), *overt victimization 1* (get picked on and teased), *relational aggression 1* (ignore others or spread rumors about others), *relational aggression 2* (try to keep others they don't like out of their group), *popularity* (most popular), *unpopularity* (least popular), *leadership* (leaders), *relational victimization 2* (has lies, rumors, and mean things said about them), *overt victimization 2* (get pushed and kicked by others), and *cooperation* (share and help others).

Data Preparation

For analyses of validity and characteristics of experts, nomination counts received by each participant were summed; as is standard in sociometric research (Cillessen & Marks, 2011). As in previous research (Babcock et al., in press; Marks et al., 2013), the data were prepared for reliability analyses by creating a binary 1/0 matrix with nominators as columns and nominees as rows. From a classical measurement perspective, we treated nominators as "items" and nominees as "participants." This allowed us to calculate inter-rater correlations for each nominator and

Cronbach’s α for each variable. Conceptually, Cronbach’s α used for peer nominations indicates the extent to which the nomination count from any given half of the nominators agrees with the nomination count from the other half (Marks et al., 2013).

Results

Our analysis strategy involved two major steps. First, we determined the best method of identifying expert nominators who would provide accurate nomination data across a range of items. Once identified, we compared the expert sample to the full sample. Second, we evaluated the accuracy of the expert nominations by their internal reliability and external validity.

Identifying Expert Nominators

To address the ontology of expert nominators, one must first settle on a criterion of what constitutes an expert. Most peer nomination research has participants nominate their peers on a wide variety of variables. If one were to use a limited number of expert nominators, one would most effectively seek a single set of nominators skilled at nominating their peers on most variables. We deduced that an overall expert nominator must be a good nominator on 75% or more of the variables in a study for multiple reasons. First, an expert peer rater should be able to rate more than a majority of items well; a person with poor ratings on half or more of the variables certainly is not a true expert. Second, requiring that experts be a good nominator on every single variable seemed unrealistically high. Third, there must be a sufficiently large pool of experts so that an entire grade is not being rated by just a handful of raters. Finally, experts should be a small enough pool so that they satisfy the goal of making a study easier to conduct. The 75% criterion seemed to fulfill these criteria, and we confirmed this expectation empirically.

Assuming this definition, we used two methods to determine whether or not expert nominators existed. First, we looked at nominators’ accuracy in an absolute sense, using the mean inter-rater correlation between a nominator’s ratings with all other nominators’ ratings for a single item. We labeled a person as an expert for an item if the person had a mean inter-rater correlation of .10 or higher. We picked .10 because of how this value influences Cronbach’s α. If every nominator had a mean inter-rater correlation of.10, Cronbach’s α (Lattin, Carroll, & Green, 2003) would meet or exceed .80 if 15% of the students were labeled as experts in the smaller of the two schools. These key reliability values go from good to acceptable in terms of the amount of measurement error that many nomination researchers are willing to accept.

The second method for identifying expert nominators was a relative definition. Instead of focusing on the inter-rater correlation and resulting reliability estimate, a researcher may be satisfied with having the best nominators among the available participants. To this end, we re-examined the relative position of each nominator in terms of the mean inter-rater correlation. If the mean inter-rater correlation was in the top *x* percent of nominators, we labeled the person as an expert for that particular nomination item.

Table 1 contains the proportion of students from each school with a mean inter-item correlation of .1 or greater. As the table shows, some variables had few or no expert nominators according to this statistical criterion. Friendship, acceptance, rejection, and both relational aggression items had fewer than 10% expert nominators for both schools. [Table 1 about here.]

Given that several individual items had few experts, it is not surprising that few participants were experts on most of the variables. Fewer than 20% of students at both schools (*N*s = 43 at School 1 and 23 at School 2) had a mean inter-rater correlations above .1 for 7 or more variables. Only 10 students at School 1 and none at School 2 had high inter-rater correlations for 9 or more variables. From a correlational perspective, few students were experts.

A more optimistic picture resulted using a relative criterion. For this, we looked at the number of students in the top 50, 40, or 30 percent of inter-rater correlations for 12 or more of the measured variables. Using the top 30% criterion was still overly conservative (only six students at School 1 and three at School 2 met the criterion). Using the top 40% criterion was not much better (20 students at each school). However, the top 50% criterion identified 42 students at School 1 and 50 at School 2. We felt that this was a reasonable proportion (just over 16.5% of both samples), and labeled these nominators as "expert nominators" in all further analyses.

While the results for the top 50% criterion were promising, it was important to confirm that they were not a statistical artifact (i.e., were not in the top 50% on several variables by chance). Therefore, we compared the distribution of the number of times adolescents were labeled as experts according to the 50% criterion to the binomial distribution, which would reflect a random process. The empirical distributions in both schools were significantly different from the binomial distribution (χ^2 goodness-of-fit test, $p < .01$), indicating that the experts identified reflected systematic differences among nominators.

Expert Versus Non-Expert Nominators

A series of *t*-tests were conducted to investigate differences between expert nominators and non-experts in terms of all sociometric variables in each school. To avoid confounding non-expertise with non-participation, only non-experts who provided nominations were included. In order to provide a conservative estimate of differences, we made the *a priori* determination to set α to .01 and to only accept differences that were significant in *both* schools.

Results showed significant differences between experts and non-experts across both schools for four study variables. Experts were significantly higher on popularity, received attention, leadership, and cooperation (ts > 3.22; ps < .01).

χ^2 tests were conducted to determine whether experts and non-experts differed on the demographics of gender and race. Because some ethnic groups were small, the race variable was dichotomized into “majority group” (White) and “minority group.” The gender composition did not differ between experts and nonexperts at either School 1 ($\chi^2 = .07$, $p = .799$) or School 2 ($\chi^2 = .00$, $p =.980$). The racial makeup of experts differed from that of nonexperts at both schools. At School 1, the proportions of minority group members were 40% for the nonexpert sample and 9.5% for the expert sample ($\chi^2 = 13.34$, $p < .001$). At School 2, the proportions of minority students were 30% for the nonexpert sample and 12% for the expert sample ($\chi^2 = 6.39$, $p = .018$).

Statistical Performance of Experts

Cronbach’s α was estimated for each variable for the full sample and the expert subsample. We then compared these estimates to the mean reliability estimates across 1,000 randomly selected samples of nominators with the same nominator sample size at each school.

As shown in Table 2, the expert sample α was substantially lower than the full sample α for all variables. In every case, the expert subsample was significantly lower than the full sample α using a 1-tailed *z*-test (see Iacobucci & Duhachek, 2003) with a critical p-value of 0.0032 (familywise Type I error rate of .05 for 16 tests). The mean reliability drop was .29 for School 1 and .37 for School 2. The minimum reliability drop for a single variable was .10. In other words, using experts instead of a full group led at least 10% more error variance, with a conservative average of 30% additional error variance. Using an expert sample clearly has a negative impact on the reliability of peer nominations. [Table 2 about here]

Beyond the drop in reliability when moving from the full sample to the expert subsample, the absolute values of expert αs for several variables was quite low. While some reliabilities were high (e.g., most popular), many would be considered inadequate (e.g., Crocker & Algina, 1986; Lattin et al., 2003). The low reliability variables also tended to be the variables upon which experts were low compared to a random sampling of students.

Table 2 also contains mean statistics for the groups of randomly picked nominators. The omnibus expert nominators performed about equally as well or worse on average when compared to a randomly picked sub-group of nominators.

It is also important to examine how the experts performed in terms of the external validity of their nominations. As the goal of using expert nominations is to approximate full sample, the relation between expert and full-sample nominations can be considered a measure of external validity (i.e., the extent to which our experts are nominating what they are supposed to be). To this end, we correlated the experts' nominations with the nominations of all nominators. The correlation between all nominators and the expert nominators will be biased due to the fact that the experts' measurement error is in both groups; it is conceptually equivalent to presenting participants with a test, and then correlating the results of the full test with the results of a subsample of the items. This shared measurement error inflates the correlation (Hsu, 1994). We corrected for this statistical artifact using Hsu's (1992) item overlap correction equation.

Table 3 contains the overlap-corrected correlations for each variable. The mean correlation for School 1 was .74, while the mean for School 2 was .69. Many of these correlations are quite low, considering that the correlations represent students from the same school nominating other students on the same variables. In fact, many of the expert correlations were lower than the mean correlations from a random sample of the whole grade using the same number of nominators, and none of the expert nominator correlations were significantly larger than the random samples. For only one variable did correlations in both schools exceed .85 (popularity). The only variables with correlations even above .80 for both schools were overt victimization, popularity, and unpopularity. Thus, for most variables, expert nominators did not meet our *a priori* criterion for acceptable validity. [Table 3 about here]

The expert nominations were not reliable enough in many cases to allow for a high correlation. Unreliability limits the size of the correlation between two measured variables (Liu & Salvendy, 2009; Spearman, 1907). With expert nominators, the nominations for numerous variables had too much measurement error to produce consistently valid measurements.

In order to investigate why the reliabilities and correlations to the full sample were often equal to or less than for the random subsamples of raters, we examined how many experts were raters with the highest inter-rater correlations for each individual variable. Both schools had about 16.5% of the full group labeled as experts. We analyzed the original inter-correlations to see what proportion of the top 16.5% of raters (best raters) were from the omnibus experts group. Across variables, the mean proportion of best raters who were also omnibus experts was 33%. In other words, 2/3 of the best raters for each variable were not omnibus experts. The raters who were the best at rating any individual item varied among the different rating items.

Discussion

The purpose of this study was to investigate the feasibility of using "expert nominators" in peer nomination research. Expert nominators are individuals within a peer group who can, ideally, provide peer nominations that will accurately substitute for nominations from the full sample of nominators. We used data from full samples of participants (8^{th} graders in two schools) and empirically determined whether (a) some participants were higher in nomination accuracy than others; (b) identified expert nominators provided reliable nominations; and (c) identified expert nominators provided valid nominations.

As we investigated ways of identifying expert nominators, it quickly became apparent that there were no easily identifiable participants who could provide highly accurate nominations across constructs. After utilizing multiple approaches for identifying "experts" based on comparative accuracy, the measure that ultimately provided a sufficient number of experts required nominators to be in the top 50% of comparative accuracy for 12 or more variables. This is a fairly liberal requirement for defining "experts," reflecting the fact that there were few (if any) participants who were dramatically more accurate than others across all variables.

Once expert nominators were identified, we compared the internal reliabilities of the full sample and the expert nominations. Expert nominations were below our *a priori* cutoff levels for most variables, and were less reliable than full-sample nominations. This fits with Marks et al. (2013), who showed that removing nominators reduced the reliability of peer nominations. Surprisingly, agreement among experts for a given peer nomination was not higher than the agreement among random, similarly sized subsamples of participants. It seems that those who were somewhat above average nominators for multiple variables were not necessarily the very best nominators for any one variable, indicating that selecting omnibus experts cuts out nominators who are extremely good on just one or two variables but not the other variables.

In order to test the external validity of expert nominations, we correlated nominations from the full sample of nominators with those from the expert subsample; correlations were statistically adjusted to correct for the fact that participants in the expert subsample were also included in the full sample. We set the *a priori* cutoff for adequate validity at r = .85, but only one variable (popularity) showed correlations of this magnitude in both schools. Four other variables showed correlations above .85 in one school. These four variables were, not coincidentally, also the most reliable variables in the study. These data reflect the well-known fact that validity-type correlations, or any correlational measures, are limited in how high they can be by the reliability of the scores involved (see Crocker & Algina, 1986; Liu & Salvendy, 2009). This psychometrically consistent finding also supports the use of psychometric reliability indices, such as Cronbach's α, to study the upper bounds for potential correlations between sociometric variables (Marks et al., 2013).

As with reliability, the omnibus experts generally did not perform much better than a random subsample of raters in terms of validity, because most of the very best raters were not in the experts group. The strong effect of missing the very top individual variable raters is due to the positive skewness of the distributions of person inter-correlations. In fact, 11 out of 16 inter-correlation distributions were positively skewed in School 1; 15 out of 16 distributions were positively skewed in School 2, where the effect was larger. The very top raters by variable strongly pulled the distribution means up, and tended to have a disproportionate contribution toward more valid scores compared to other raters. Although the overall expert raters were in the 50th percentile or higher on 75% of variables, the fact that the overall expert group lacked most of the very top individual variable raters meant that it would be difficult for the experts to beat a random sample, which could include some of the overall experts also by random chance.

Interestingly, the relative reliability and validity across constructs in this study were in line with previous investigations of the internal reliability of peer nomination measures (Babcock et al., in press; Marks et al., 2013), with peer nominations of status and concrete behavior (e.g., popularity, overt aggression, prosocial behavior) tending to be more reliable and valid than affective nominations (see also Prinstein, 2007). In other words, participants showed higher levels of agreement for variables for which higher peer consensus was expected.

Contrasts with Prinstein (2007)

In addition to investigating expert nominators more broadly, a goal of our study was to examine Prinstein's (2007) findings. We conducted this investigation because we understand the appeal of reducing peer nominators by 80 to 90%. Prinstein’s work has been cited to support use of low participation rates in sociometric research (e.g., Walcott et al., 2008) and has been noted as a promising methodological option in two frequently cited handbooks (Brown & Larson, 2009; Cillessen, 2009). If Prinstein’s conclusions are problematic (as we argue) and are taken at face value, they may lead researchers not well versed in sociometric methodology to either (a) publish unsound research using Prinstein (2007) to justify a low participation rate or (b) collect data from a small subset of participants, believing that those participants are providing valid data.

It is worth noting that some correlations between the expert and the full samples were similar in the current study to those reported by Prinstein (2007). This similarity is likely due to psychometric properties of the variables themselves, rather than to actually finding similar levels of validity. Given the information provided by Prinstein (and using Hsu, 1994 Equation 7), we can estimate that correlations between “expert” and full group measures would have been around .33 due to item overlap if participants had responded *randomly*—this correlation would be significant at the .001 level. Given that our study included a higher proportion of nominators and that both of our full samples were slightly larger than Prinstein’s, we would posit that the correlations appear numerically similar to Prinstein’s only because Prinstein’s were inflated.

Our method of identifying experts also stacked the deck in their favor. Prinstein’s (2007) experts were identified by English teachers. Although Prinstein’s method allows for *a priori* identification of experts, we cannot know whether the chosen participants will be accurate across constructs (or highly accurate for individual constructs). In contrast, the current study’s method of identifying experts required that experts provide above-average nominations across items. We were unable to identify an adequate set of broad omnibus expert nominators in an existing dataset with a high participation rate using the actual nomination data to guide the expert selection decision. As teachers would have neither this much relevant data nor a way to use the data systematically, it is difficult to argue that teachers could nominate experts that are more effective than those used in this study. In this regard, this study created unrealistically favorable conditions to find experts. Our results indicate highly accurate omnibus expert nominators may not exist, so it is impossible to be more effective in choosing such experts.

Even if we ignore the methodological and analytical differences between the current study and Prinstein (2007) and focus only on the results, our conclusions differ from Prinstein’s, especially with regard to the utility of using experts for affective nominations. We would argue that correlations between the full and expert samples in the range of .6 to .7 are far too low to indicate that experts can be substituted for full samples. Experimenting with the cosine angle of correlation equation from above, expert to full correlations of .6 can yield correlations with external variables that span more than half the correlation range (i.e., correlates positively with the full group, correlates negatively with the experts). As we can see from the full sample reliability values (Table 2), peer nominations are inherently quantified with measurement error. Using an expert subsample whose correlations do not strongly correspond to full-sample nominations will greatly compound this measurement error. As such, we would argue that experts who deliberately choose to use small subsamples of participants to evaluate multiple peer nomination variables (particularly if those variables include affective nominations) are risking the introduction of considerable error. As noted above, even an expert/full sample correspondence of .85 will cause major fluctuations in correlations between those variables.

In addition to the large differences in potential correlations with external variables, the increase in measurement error when just using experts will affect virtually every statistical procedure, not the least of which is substantially lowering the size of correlations between variables. This and numerous other effects have already been well documented in the literature (Liu & Slavendy, 2009). At times, using only experts increased error variance by 30%. Considering the host of correlational and null hypothesis significance testing problems that measurement error can cause, this increase in error when using experts only is unacceptable.

Assumptions and Limitations

The conceptualization of “experts” in this study assumed a traditional use of peer nominations in which nominations received are totaled for each participant. This is a variable-oriented view of peer nominations (as opposed to person-oriented; Rodkin, Farmer, Pearl, & Van Acker, 2000), and does not reflect dyadic and group processes that are increasingly being studied by sociometric researchers (e.g., using friendship nominations to identify and study dyadic relationships or social networks). Although issues of participation rates and the quality of informers are important in dyadic and network analyses, the current conceptualization and evaluation of experts is not directly relevant to these types of analyses.

Our method of determining external validity was to compare the nominations of our subsample of “experts” with those of the full sample of nominators. Thus, validity was established by comparing two related sets of data collected with the same method, rather than comparing nominations to an alternate criterion such as teacher reports or behavior observations. We used this method because our goal was to show whether nominations from experts could substitute nominations from a full sample; however, we acknowledge that more studies must be conducted to establish criterion validity as well as external validity across multiple contexts.

One major assumption of this research is that all variables assessed by peer nominations are unidimensional latent variables that can be measured based on consensus (Marks et al., 2013). Although this assumption has received empirical support (Terry, 2000), it may contrast with the way many researchers conceive of "affective" variables like acceptance. Indeed, asking participants whom they like (assessing individual feelings toward each peer) is qualitatively different than asking who is popular (assessing reputation) or aggressive (assessing observable behavior). However, researchers treat affective variables as latent variables both practically and conceptually. In practice, most researchers analyze nomination data using statistical techniques which assume that constructs are normally distributed in the population. Conceptually, since each participant's criterion for liking is different, liking nominations are inherently assumed to include measurement error; researchers also assume that this error will balance out across nominators and provide us with an understanding of which participants are liked within a classroom or grade. This assumed measurement error is exactly why Moreno (1934) argued that nominators should *never* simply be asked whom they like or dislike (see Polansky, Lippitt, & Redl, 1950). From a measurement perspective, the consensus regarding a variable like acceptance is important because it is calculated as a composite of nominations. If we expect low agreement among nominators, we acknowledge that acceptance is measured with non-trivial amounts of error. Expecting *no* agreement is tantamount to saying that the data are random. Saying that liking nominations should correlate with (for example) prosocial nominations is essentially saying that there is systematic variance in peer acceptance that is systematically related to variance in prosocial behavior. This systematic variance determines reliability.

The current study also used only one statistical measure of measuring reliability: Cronbach's alpha. Future studies may want to use other indices of reliability or rater agreement. Examples of indices for use in future studies are Guttman's λ_2, one of the most popular reliability estimation alternatives to Cronbach's alpha, or Fleiss' kappa coefficient, which is an index of rater agreement. It is possible that the absolute quantities reported in this study could change somewhat using alternative measures. The authors doubt, however, that these alternative indices would change the main message of the findings here: a small group of experts is generally not a good substitute for obtaining most people in a grade for a peer rating study.

An additional limitation of this study is that we used only one operational definition of peer nomination experts, which related to a nominator’s relative correlation with other peers’ ratings for all variables. This is not the only possible definition for an expert. Alternative definitions of could also be used, such as using only students with very high absolute values of correlation with others’ nominations, using only students with the very highest correlation with all students’ ratings for individual variables, using students nominated by teachers, or any number of other measures. If the purpose of research was only to study a very limited number of nomination variables, a defining experts as those only who do a very good job nominating on a few variables could lead to more positive results than those found in the current study. The current research suggests that more conservative definitions of experts, such as only using nominators with extremely high inter-rater agreement across many variables, would lead to having so few expert nominators that the experts’ results would not be reliable. Whatever definition that future research uses, it is important that, even though it is possible for fewer nominators to potentially yield reliable results, there are enough of the experts to have nomination scores suitable for research (Marks, Babcock, Cillessen, & Crick, 2013). Definitions of experts that would include too many students violate the pragmatic goal of having fewer students in the study, which researchers of this subject must also keep in mind.

Finally, it is important to note that our analyses and conclusions are relevant to peer nominations only. There are other sociometric methods that utilize information from peers (e.g., rating and ranking methods) that might be feasible using experts, particularly for smaller classroom-based contexts (since these alternatives to nominations become logistically onerous with large samples; Terry, 2000). In addition, any use of expert nominators can only be relevant to peer nomination studies in which nominators are allowed to identify non-participating class- or grade-mates. Indeed, one of the primary reasons that researchers might want to use expert nominators is to avoid the need to obtain consent from a full classroom or grade of students. However, ethical review boards (at research institutions or the schools in which data are being collected) may object to the collection of nominations regarding non-consenting individuals. In this case, there could be no advantage to collecting data from expert nominators, because those nominators would be unable to provide information about their non-expert peers.

Conclusions and the Future of Expert nominators

Ultimately, the results of this study paint a fairly pessimistic picture regarding the ability of expert nominators to substitute for full samples of peer nominators. Our findings indicate that it might be unrealistic to expect to find expert nominators who can provide accurate information across behavioral, reputational, and affective criteria.

Even if it is not feasible for researchers to use general sociometric experts in future research, one promising avenue of study may be the identification of more limited, criterion-dependent experts. Whereas the current study categorized sociometric experts based on their accuracy across 16 items, it may be more practicable to find, for example, “popularity experts” for a study of popularity or “aggression experts” for a study of aggression. This may be more likely for some variables than for others. As demonstrated (see Table 1), more than half of the students in both schools demonstrated adequate comparative accuracy for popularity nominations. In contrast, none of the students showed adequate comparative accuracy for acceptance nominations. These results mirror those of recent reliability analyses (see Babcock et al., in press; Marks et al., 2013), implying that it may be more feasible to identify experts for sociometric criteria that are internally reliable overall. Such methods would be most useful when researchers are only investigating a small number of peer-nominated variables.

The approach of the current investigation was similar to the approach used by Robert Terry to identify the *least* accurate nominators in a sample. Terry (2000) argued that identifying the least accurate nominators would allow researchers to exclude (or reduce the statistical weight of) participants who either were nominating peers randomly or were less conscious of the social consensus. These goals further highlight the importance of future research investigating the relative accuracy of different sets of nominators. Beyond the general methodological and psychometric questions presented by such investigations, it would be important to know why some adolescents may be systematically more or less accurate in their nominations than others. The fact that popular, well-liked adolescents were more likely to be identified as experts in the current investigation is unsurprising—we would expect these participants to be more "tuned in" to the peer consensus (Prinstein, 2007). The fact that ethnic minority adolescents were significantly less likely to be identified as experts is more interesting, and could suggest that the "peer consensus" measured by sociometric nominations is a reflection of the majority ethnic group's views. This may indicate that the social views of the ethnic majority conflict with those of the minority, or it may indicate that individuals in the minority are part of different social networks within the school than members of the majority.

In conclusion, this study supports the recommendation that researchers use the maximum amount of available data (in this case, the maximum number of nominators) in sociometric research. At its basic level, this study investigated a very specific type of systematic missingness, and showed that even retaining the most broadly accurate nominators in a sample increases error and decreases accuracy. Future research may reveal exceptions to this rule or show ways in which expert nominators might be identifiable (particularly for studies focusing on one or two sociometric variables), and there will always be an ongoing debate regarding cut-offs for “good enough” reliability and validity. Currently, however, the results of this study and related studies (Babcock et al., in press; Marks et al., 2013) suggest that deliberately excluding nominators in sociometric research is, psychometrically speaking, imprudent.

References

Bracken, B. A. (1992). *Multidimensional Self Concept Scale*. Austin, TX: PRO-ED.

Brown, B., & Larson, J. (2009). Peer relationships in adolescence. In R. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (3rd ed., Vol. 2, pp. 74-103). New York: Wiley. doi: 10.1002/9780470479193.adlpsy002004

Babcock, B., Marks, P. E. L., Crick, N. R., & Cillessen, A. H. N. (in press). Limited nomination reliability using single- and multiple-item measures. *Social Development*.

Cillessen, A. H. N. (2009). Sociometric methods. In K. H. Rubin, W. Bukowski, & B. Laursen (Eds.), *Handbook of peer interactions, relationships, and groups* (pp. 82-99). New York: Guilford Press.

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: Guilford.

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.

Crocker, L., & Algina, J. (1986). *Introduction to classical & modern test theory*. Belmont, CA: Wadsworth.

Fournier, M. A. (2009). Adolescent hierarchy formation and the social competition theory of depression. *Journal of Social and Clinical Psychology, 28,* 1144-1172. doi: 10.1521/jscp.2009.28.9.1144

Guilford, J. P. (1936). *Psychometric methods.* New York: McGraw-Hill.

Hsu, L. M. (1992). Correcting correlations of personality scales for spurious effects of shared items. *Multivariate Behavioral Research, 27,* 31-41. doi: 10.1207/s15327906mbr2701_3

Hsu, L. M. (1994). Item overlap correlations: Definitions, interpretations, and implications. *Multivariate Behavioral Research, 29,* 127-140. doi: 10.1207/s15327906mbr2902_1

Iacobucci, D., & Duhachek, A. (2003). Advancing alpha: Measuring reliability with confidence. *Journal of Consumer Psychology, 13,* 478-487.

Lattin, J., Carroll, J. D., & Green, P. E. (2003). *Analyzing multivariate data*. Pacific Grove, CA: Brooks/Cole-Thompson Learning.

Liu, Y., & Salvendy, G. (2009). Effects of measurement errors on psychometric measurements in ergonomics studies: Implications for correlations, ANOVA, linear regression, factor analysis, and linear discriminant analysis. *Ergonomics, 52*, 499-511. doi: 10.1080/00140130802392999

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. doi: 10.1111/j.1467-9507.2012.00661.x

Moreno, J. L. (1934). *Who shall survive? A new approach to the problem of human relations*. Washington D. C.: Nervous and Mental Disease Publishing Company.

Polansky, N., Lippitt, R., & Redl, F. (1950). The use of near-sociometric data in research on group treatment process. *Sociometry, 13*, 39-62.

Prinstein, M. J. (2007). Assessment of adolescents' preference- and reputation-based peer status using sociometric experts. *Merrill-Palmer Quarterly, 53,* 243-261. doi: 10.1080/00140130802392999 10.1353/mpq.2007.0013

Rodkin, P. C., Farmer, T. W., Pearl, R., & Van Acker, R. (2000). Heterogeneity of popular boys: Antisocial and prosocial configurations. *Developmental Psychology, 36*, 14-24. doi: 10.1037/0012-1649.36.1.14

Spearman, C. (1907). Demonstration of formulæ for true measurement of correlation. *American Journal of Psychology, 18*, 161-169.

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* (Vol. 88, pp. 27-53). San Francisco: Jossey-Bass. doi: 10.1002/cd.23220008805

Walcott, C. M., Upton, A., Bolen, L. M., & Brown, M. B. (2008). Associations between peer-perceived status and aggression in young adolescents. *Psychology in the Schools, 45*, 550-561. doi: 10.1002/pits.20323

Table 1

Proportion of Nominators with Inter-item Correlation ≥ .1 for Each Item

Variable	School 1	School 2
Friendship	.00	.00
Acceptance	.00	.00
Rejection	.00	.00
Received Attention	.33	.08
Overt Aggression	.04	.22
Withdrawal	.35	.43
Relational Victimization 1	.11	.18
Overt Victimization 1	.56	.45
Relational Aggression 1	.00	.00
Relational Aggression 2	.07	.00
Popularity	.60	.65
Unpopularity	.43	.42
Leadership	.31	.00
Relational Victimization 2	.29	.24
Overt Victimization 2	.52	.19
Cooperation	.11	.00

Table 2

Internal Reliability of Full Sample, Identified Experts, and Random Subsamples of Nominators

	School 1			School 2		
Variable	Full Sample α	Expert Subsample α	Random Subsamples α	Full Sample α	Expert Subsample α	Random Subsamples α
Friendship	.72	.17	.29	.72	.17	.28
Acceptance	.85	.32	.47	.89	.64	.55
Rejection	.85	.42	.46	.86	.43	.49
Received Attention	.94	.68	.70	.92	.61	.65
Overt Aggression	.91	.46	.59	.94	.72	.69
Withdrawal	.93	.72	.68	.96	.57	.78
Relational Victimization 1	.77	.57	.33	.88	.30	.52
Overt Victimization 1	.96	.81	.78	.95	.65	.75
Relational Aggression 1	.80	.11	.36	.83	.39	.42
Relational Aggression 2	.88	.65	.51	.82	.27	.41
Popularity	.97	.85	.85	.98	.86	.87
Unpopularity	.94	.79	.72	.95	.63	.73
Leadership	.92	.80	.66	.85	.45	.47
Relational Victimization 2	.90	.72	.59	.91	.57	.61
Overt Victimization 2	.97	.87	.82	.91	.43	.60
Cooperation	.88	.65	.54	.86	.60	.50

Note. Values listed for the random subsamples indicate the mean Cronbach's α across 1000 randomized subsamples which included the same number of individuals as the expert sample within each school.

Table 3

Correlations between Full Sample Nominations and Both Expert and Random Subsample Nominations, Corrected for Rater Overlap

Variable	School 1		School 2	
	Expert Subsample	Random Subsamples	Expert Subsample	Random Subsample
Friendship	.45	.47	.45	.47
Acceptance	.63	.64	.78	.71
Rejection	.59	.65	.65	.66
Received Attention	.82	.82	.75	.79
Overt Aggression	.69	.75	.82	.82
Withdrawal	.84	.81	.73	.88
Relational Victimization 1	.64	.52	.62	.69
Overt Victimization 1	.88	.87	.81	.86
Relational Aggression 1	.44	.56	.56	.61
Relational Aggression 2	.71	.69	.49	.60
Popularity	.93	.91	.93	.92
Unpopularity	.87	.83	.80	.84
Leadership	.86	.78	.66	.64
Relational Victimization 2	.78	.74	.70	.76
Overt Victimization 2	.93	.90	.55	.76
Cooperation	.77	.69	.73	.66

Note. Random samples of equal number of nominators as experts obtained by resampling the full groups without replacement; correlations indicate means across 10,000 replications.