

# Question-Order Effects in Social Network Name Generators<sup>1</sup>

**James E. Pustejovsky**  
**James P. Spillane**  
**Northwestern University**

Draft: May 7, 2008

Social network surveys are an important tool for empirical research in a variety of fields, including the study of social capital and the evaluation of educational and social policy. A growing body of methodological research sheds light on the validity and reliability of uni-dimensional social network survey data, but much less attention has been paid to the measurement of multi-dimensional networks and the validity of dimensional comparisons. In this paper, we identify ways that surveys designed to collect multi-dimensional social network data might be vulnerable to question-order effects, then test several hypotheses using a split-ballot experiment embedded in an online multiple name generator survey of teachers' advice networks. We conclude by discussing implications for the design of multiple-name generator social network surveys.

---

<sup>1</sup> An earlier version of this paper was presented at the 28<sup>th</sup> annual International Sunbelt Social Network Conference, January 22-28, 2008 in St. Pete Beach, Florida. Work on this project is supported by the Distributed Leadership Studies (<http://www.distributedleadership.org>), funded by research grants from the National Science Foundation (RETA Grant EHR-0412510), with support from Northwestern University's School of Education and Social Policy and Institute for Policy Research. We thank Michael Lach, Gina Grant, Bill Suttles, Ruth Heaton, and Jim Lewis for their assistance with data collection, and Ken Frank for many helpful comments on our analysis. We also gratefully acknowledge our research team for their continuous support, especially Enrique Orlina, Virginia Pitts, Cindy Sigal, Bijou Hunt, and Cassandra Chen. All opinions and conclusions expressed in this paper are our own and do not reflect the views of any funding agency. Please direct correspondence to James Pustejovsky at Northwestern University, School of Education and Social Policy, 2120 Campus Drive, Evanston, IL 60201 or [j-pustejovsky@northwestern.edu](mailto:j-pustejovsky@northwestern.edu).

Social network surveys are an important tool for empirical research in a variety of disciplines and applied fields, including the study of social capital and the evaluation of educational and social policy. A growing body of methodological research sheds light on the validity and reliability of uni-dimensional social network survey data—that is, measurements of a single relationship among a set of actors. However, much less attention has been paid to the measurement of multi-dimensional networks and the validity of dimensional comparisons, despite the fact that many research questions require attention to several types of relationships among a given set of actors. In this paper, we identify ways that surveys designed to collect multi-dimensional social network data might be vulnerable to question-order effects, then test several hypotheses using a split-ballot experiment embedded in an online multiple name generator survey.

Sociologists and organizational theorists recognize that social capital, broadly defined as resources for action that are attained through relationships (Coleman, 1988; Lin, 1982), is composed of various, qualitatively distinct types of relationships. One line of research on social capital attends to these multiple relationships by examining the structure of multi-dimensional social networks, or networks of heterogeneous content (Burt, 1997). In studies of personal social capital, multi-dimensional network data has been used to examine differences between emotional support networks and social support networks (Bernard, Johnsen, Killworth, McCarty, Shelley, and Robinson, 1990), to identify factors affecting reciprocal exchange of support (Plickert, Côté, and Wellman, 2007), and to validate widely-used name generator questions (Ruan, 1998). In the realm of organizational network analysis, multi-dimensional networks have been used to study patterns in the relational structure of a law firm spread across multiple offices (Lazega

and Pattison, 1999), to determine the dimensionality of advice seeking behavior (Cross, Borgatti, and Parker, 2001), and to identify factors related to the career advancement of managers in large corporate firms (Burt, 1997; Podolny and Baron, 1997).

While multi-dimensional network data allow opportunities for a wider range of analysis than do networks of a single relation, the measurement and analysis of multi-dimensional networks is also subject to a broader array of potential validity threats. Not only must each dimension of the network be a valid and reliable measurement in its own right, but the distinction between dimensions must itself be valid and reliably measured. Empirically, the relational dimensions that a researcher seeks to measure must be specified so that a survey respondent can distinguish among them.

Among the many design choices involved in constructing a multi-dimensional network survey, the researcher is faced with the question of how the measurement of several relational criteria should be arranged. Which relation should be measured first? The possibility that the order in which questions are posed could create bias, or what we term here a question-order effect, is an immediate concern for researchers relying on survey designs (Burt, 1997; Ruan, 1998; Straits, 2000).

In order to measure multiplex networks, survey methods for measuring single networks, which include roster-based methods and recall-based methods, are typically extended to cover multiple criterion relationships. In organizational network studies where the set of all relevant actors can be determined in advance, roster-based recognition methods can be applied. Roster-based surveys use name interpreter questions, which ask a respondent to characterize their relationship with each member of a group. However, roster methods may present a considerable reporting burden if participants are

asked to report on each member of a large organization. In some cases, the reporting burden may be lessened by limiting the survey to a sub-set of names, chosen based on the organizational structure (Reagans and McEvily, 2003).

To measure multiplex networks using rosters, name interpreter questions would be posed about each of several relationships. Name interpreter questions are posed either question-wise, where the respondent answers one criterion relationship question about the entire set of possible alters, then repeats the process with the next criterion relationship, or alter-wise, where a respondent characterizes her relationship with one possible alter in terms of all the criterion relationships of interest before moving on to the next possible alter (Kogovšek, Ferligoj, Coenders, and Saris, 2002).

Recall-based methods use name generator questions, which ask the respondent to name a set of people that fit a given criterion relationship. The criterion relationship can be formulated in various ways: by specifying a social role, a minimum frequency of contact, closeness, or a specific type of social exchange (Marin and Hampton, 2007).

Recall-based methods are often necessary because all of the relevant possible members of a network cannot be identified in advance, thereby preventing the use of roster methods. A name generator may be followed by a set of name interpreter questions that ask the respondent to provide additional information about some or all of the individuals they have named.

To measure multiplex networks using recall methods, a sequence of two or more name generators would be posed; each name generator would ask the respondent to list people that fit a specific criterion relationship. Name interpreter questions might also be posed about the contacts from each name generator.

In this paper we focus on such recall-based approaches for measuring multi-dimensional social networks, examining the effects of question order on responses to multiple name generator surveys. After reviewing previous research on measuring social networks, we outline several ways in which surveys that collect multi-dimensional social network data might suffer from measurement bias. We test our theories using a randomized experiment embedded in two studies of advice networks among elementary and middle school teachers. We conclude by drawing implications for the design of survey instruments to measure multi-dimensional social networks.

### **Name generator accuracy and reliability**

A sizable literature addresses the accuracy and reliability of name generator questions (Marsden, 2005 provides an excellent overview), but most of this work focuses only on networks defined on a single relationship, such as acquaintanceship. In an influential series of papers, Bernard, Killworth, and Sailer (1976, 1977, 1979/80, 1982) argued that network data collected through recall-based methods is inaccurate on a dyadic level. Later studies also demonstrated that respondent forgetting is prevalent in recall-based social network data (for a review, see Brewer, 2000). Forgetting increases with size of a person’s network but decreases with behavioral specificity (Bell, Belli-McQueen, and Haider, 2007). Split-ballot experiments have found that recall-based methods produce much smaller estimates of local network size than do roster-based methods (Sudman, 1985). There is also evidence that the names generated in recall-based instruments are affected by closeness and both frequency and recency of contact

(Hammer, 1984). Further, the salience of the criterion relationship specified in a name generator question affects recall accuracy (Bell et al., 2007).

Still, some scholars argue that while recall-based instruments generate an inaccurate record of specific interactions or events, especially under-estimating the number of interactions, they are a useful record of how a respondent thinks about past events (Freeman, Romney, and Freeman, 1987). Freeman and colleagues suggested that the accuracy of network data may be a function of the degree to which respondents have elaborated a mental framework for remembering the interactions or relationships that they are asked to recall.

Multi-dimensional network data present an additional set of validity concerns, particularly for studies comparing one dimension to another. There is a scarcity of work that studies multiple name generator social network survey designs. Ferligoj and Hlebec (1999) demonstrated question order effects on reliability, concluding that network data from later name generators is somewhat more reliable than data from initial name generators. Examining the drawbacks of a sequential, multiple-name generator survey design, Straits (2000) suggested three possible mechanisms that would produce question-order effects: priming, fatigue, or a reluctance to repeat alters for fear of being redundant. Below we elaborate on these mechanisms and several others, drawing from research on cognitive aspects of survey methodology.

### **Question-order effects in social network name generators**

Studies of question-order effects in behavioral and attitudinal surveys suggest specific mechanisms that may also be applicable to multiple name generator social

network surveys. Most research on the cognitive aspects of survey methodology focuses on attitude questions, which ask respondents to select an opinion from a list of options or evaluate their level of agreement with a statement, or on behavioral frequency questions, which ask respondents to report on how often they have engaged in certain behaviors (Sudman, Bradburn, and Schwarz, 1996). Though a social network name generator question presents a very different set of concerns than the Likert-scale item design that is typical in attitudinal surveys, we believe that some of the same cognitive principles may apply. In developing their model of cognitive responses to attitude questions, Tourangeau and Rasinski (1988) suggest that respondents access a mental network of associated beliefs and ideas in order to determine a summary attitude in response to a question. Social network name generators also ask respondents to access a mental network, but one composed of people and relationships, as distinct from abstract beliefs and ideas. To the extent that the cognitive process involved in accessing networks of beliefs and ideas is similar to that of accessing memory representations of social networks, similar principles of survey response might be applicable.

Available evidence suggests that question-order effects might appear as a result of at least five inter-related mechanisms: fatigue, satisficing, conversational norms of non-redundancy, cognitive priming, or question scope redefinition (Tourangeau and Rasinski, 1988; Straits, 2000). We take up each of these mechanisms in turn, examining conditions under which they might cause question-order effects in multiple name generator surveys, and reasoning about the resulting bias.

The following examination of question-order effects is limited to those that appear as a result of using sequential name generators in a survey. For explanatory

purposes, we imagine a multiple name generator survey with two name generators. Question-order effects are present to the extent that the network as measured by the second name generator differs in some dimension from the network that would have been produced by the second name generator, if the first name generator were not asked. By extension, in a survey containing three or more name generators, question-order effects are present to the extent that the network as measured by a given name generator differs from the network that would have been produced, were any of the preceding name generators not posed.

*Fatigue.* Respondent fatigue is perhaps the simplest mechanism that could create question-order biases. Fatigue effects create bias if, in response to the second name generator, a respondent names fewer alters than she otherwise would have, had the first name generator not been posed. In the extreme, fatigue might lead to non-response to later name generators. Fatigue effects may be particularly pronounced in surveys where the overall length depends on the number of items named in response to a question (Tourangeau and Rasinski, 1988). As discussed below, the name interpreter questions in our survey create such a situation.

Fatigue effects would lead to a diminished average out-degree, and consequently also diminished density. Such a pattern of bias would be particularly troublesome for multi-variate network studies seeking to compare the relative size or density of two networks. Beyond these basic measures, fatigue effects could create bias that are connected with patterns in the reported order of alters, if alters that would typically be reported further down the list are censored. For example, if alters that are encountered

less frequently tend to be reported later down the list in response to a given name generator, the respondent’s average frequency of interaction would be biased upward.

*Satisficing.* Satisficing is a more nuanced theory of how respondents behave when fatigued, bored, unmotivated, or confused. Satisficing effects occur when a respondent gives a response that she believe satisfies the request for information, but is not a complete, optimally considered response (Krosnick, 2000). The theory of satisficing as developed in survey research is typically applied to attitude questions, where it is used to explain primacy or recency effects, acquiescence bias, and status-quo or no-opinion bias (Krosnick, Narayan, and Smith, 1996). Satisficing behavior is thought to be regulated by task difficulty, respondent ability, and respondent motivation (Krosnick et al.,1996).

In the context of name generator prompts (or other factual/behavioral questions that ask the respondent to list an unspecified number of items), satisficing would play a role as a respondent decides how many alters to list in response to a name generator prompt. A satisficing respondent will take cues from the design of the survey to determine the number of names necessary for a sufficient response to a name generator. Such cues might include the number of lines provided after a name generator prompt, which a respondent takes as an indication of the researcher’s expectation about the range of items that will be listed. In a multiple name generator survey, the first prompt in a survey is a novel question, but the second prompt follows the same pattern as the first. When confronted with the second name generator prompt, a satisficing respondent could turn to the precedent that she herself set by responding to the first name generator. For example, she may stop searching her memory after listing three names in the second

generator, because the three names that she listed in the first generator seemed to be an adequate response.

Fatigue effects and satisficing effects are competing theories of how respondents answer survey questions when tired or under-motivated. Fatigue effects lead to downward bias in out-degree and network density, regardless of the relative density of the networks being measured. In contrast, satisficing leads to downward bias in measured out-degree and density only to the extent that a respondent has a higher (actual) out-degree in the second network than in the first. If the reverse is true, satisficing produces no bias in the network, or might even lead a respondent to list more names than she otherwise would have, so as to match the precedent set in response to the first name generator. Satisficing respondents are likely to list approximately equal numbers of alters in response to multiple generators, creating upward bias in the correlation between the out-degrees of each name generator.

*Non-redundancy.* Studies of attitudinal surveys have found that respondents sometimes interpret subsequent questions as requests only for new information rather than as independent questions, leading them to omit consideration of information that they have already offered (Schwarz, 1999). In the context of multiple name generator surveys, non-redundancy effects would appear if a respondent omits the names of certain alters in the second name generator because she has already listed the alters in the first name generator. The respondent might interpret the second name generator prompt as beginning with the qualification, “Aside from the people you have already named...”

Non-redundancy effects are difficult to observe, because in advance of measurement it is difficult to know how much overlap is present in a given pair of

networks or relationships (in fact, measuring network multiplexity is sometimes a substantive question for study). Non-redundancy effects are observed if a respondent’s relationship to a given alter fits the criteria specified for both name generators (i.e., the respondent has multiplex ties to the alter), but the respondent names the alter only in the first name generator. The bias created by non-redundancy effects therefore depends on the actual prevalence of multiplex ties. The overall effect of non-redundancy is to reduce the average out-degree in the second network, reduce the density of the second network, and reduce the in-degree of actors that are a part of both networks (thus biasing downward the observed level of multiplexity), though all of these effects will only be present to the extent that the two networks overlap.

*Cognitive priming.* Cognitive priming in an attitude question affects the retrieval of information from memory that is relevant to answering the question (Tourangeau and Rasinski, 1988). A previous question may have sub-consciously activated a set of relevant memories. These memories would not otherwise have been drawn upon in forming a judgment about the current question, but now their activation could cause a change in the respondent’s reported opinion.

Priming is just as relevant for behavioral frequency questions. For example, one experiment found that first answering a set of questions about one’s general opinions of crime and victimization led to increased reporting of victimization incidents in the past year (Cowan, Murphy, and Weiner, 1978).<sup>2</sup> In the context of a multiple name generator

---

<sup>2</sup> The social network questions in the survey described below were preceded by several banks of attitude questions that ask the respondent to assess the general social climate of her organization. Posing such questions may effectively prime a respondent’s memories of specific interactions with colleagues, making it more likely that those colleagues are considered in response to the name generators that follow.

survey, the process of retrieving names from memory for the first generator may start a sub-conscious activation process that brings certain names to the forefront for subsequent name generator questions. If not for the priming effect of the earlier name generator, a respondent might not have listed certain alters in the current name generator.

Priming does not necessarily bias the number of names listed in response to name generators. Instead, priming produces bias by changing the set of names that a respondent considers when determining which of her relationships fit the criterion specified in the generator. Priming would have a greater effect on the results of a second name generator to the extent that the second network is composed of a different set of actors than the first. If the central actors in both networks are largely distinct, then priming the actors in the first network would result in additions to (or perhaps replacement of) the set of alters named in the second network. If the actors in both networks are largely the same, then priming the actors in the first network might have little effect on, or might even increase the accuracy of, the alters named in the second network. Priming effects therefore create bias that is directionally opposed to non-redundancy effects, by increasing the similarity between networks measured with subsequent name generators.

Determining the dimensionality of bias caused by cognitive priming is difficult without a model of a respondent’s cognitive process for answering network questions. To understand specifically how priming might distort the measurement of a network, one would need to know how a respondent’s social relationships are organized in her memory. After the first name generator prompt is answered, the act of retrieving those names from memory could cause other related names to be primed. But what how are those names related? Within an organization, it may be that colleagues tend to be grouped

in a respondent’s memory according to the formal organizational chart (e.g., departments, sub-units, or work teams), so that naming one as an alter primes the names of others in the same department. Also plausible is that colleagues are grouped together by some informal organization, such as people who tend to go to lunch together.

Evidence from a number of studies has suggested that the social proximity of alters plays a role in how a respondent recalls them from memory when answering a name generator (Brewer, Rinaldi, Mogoutov, and Valente, 2005). The more that two alters interact with one another, the more they are perceived as socially proximate, and the more likely they are to be listed consecutively in response to a name generator. Brewer and colleagues assume that the report order of a group of alters is a list of free-associations, suggesting that sequentially listed alters are closely linked in a respondent’s memory. However, while social proximity provides an explanation for the report order of alters in a single name generator, it is uncertain whether social proximity affects recall across multiple name generators, or name generators that use criterion relationships other than acquaintance.

*Question scope redefinition.* Question context effects may result from the manner in which survey respondents use the wording of specific questions, the sequencing of questions (adjacent questions, in particular), and other facets of the instrument to infer the pragmatic meaning of a question (Schwarz, 1999). Social network name generators are no exception; a respondent must make some assumptions about the sort of names that the question is intended to produce, and will look for contextual clues in order to understand

the relationship being described (Bailey and Marsden, 1999).<sup>3</sup> If a respondent relies on contextual clues from the first name generator to understand the pragmatic meaning of the second generator, the alters that she names may be different from those she would have named in the absence of the first generator. For example, asking “Please list the names of five friends” as an initial name generator may produce a wide variety of responses, because the “friend” criterion is fairly ambiguous. If instead the “friends” name generator is preceded by questions about childhood experiences, the scope of the name generator may be implicitly re-defined to focus exclusively on childhood friends.

Question scope redefinition would produce empirical effects that are very similar to those produced by priming. To the extent that a respondent assumes that the meaning of the first name generator is similar to the meaning of the second, the results from the second name generator should more closely resemble the results from the first name generator. Conceptually, question-scope re-definition could be distinguished from priming based on the respondent’s cognitive process. Priming occurs at the level of sub-conscious memory processes, whereas question-scope redefinition has to do with

---

<sup>3</sup> McPherson, Smith-Lovin, and Brashears (2006) consider question-scope redefinition as a possible confounding effect in their comparison of ego-centric network size from 1985 and 2004. Network data was collected using the same name generator (using a criterion relationship of *people with whom you discuss important matters*), but the questions immediately preceding the name generator were different in the two years. In 1984, the name generator followed a group of questions about religion, whereas in 2004, the name generator followed a group of questions about voluntary group affiliations. Given that the “important matters” criterion relationship is fairly vague, it certainly seems possible that respondents may have relied on the preceding questions to infer the pragmatic meaning of the name generator. However, it is unclear how the change in context could have produced the reported large decrease in network size.

respondent’s understanding and interpretation of the question, something that they should be able to express.

The five possible sources of question-order effects that have been identified fall into three areas, which structure our empirical analysis below. Fatigue and satisficing effects act most directly on out-degree, the number of alters that a respondent lists. They offer competing hypotheses regarding the direction of bias. Non-redundancy and priming effects are directly related to the amount of overlap between networks defined on different criterion relationships. Our experimental design lets us say very little about the extent of these effects. Finally, question-scope redefinition can act on nearly any aspect of the measured network, potentially producing biases in both name generators and name interpreters. Below we examine empirical evidence of question-order effects, using a split-ballot experiment embedded in an online social network survey.

### **Survey design and research methods**

To understand the logic of our social network survey instrument, it is helpful to first explain the theory and prior research that informed our design. We begin with a brief overview of the context which guided the design of our multiple name generator survey, followed by a description of the instrument.<sup>4</sup> We then describe two studies that made use of the instrument and outline our approach to data analysis.

*Theory and prior research.* Our instrument was designed to study social capital in elementary and middle schools by measuring advice relationships among teachers and

---

<sup>4</sup> The full survey can also be accessed at the following website:

<http://www.sesp.northwestern.edu/Survey/SchoolNetworkSurvey.html>.

other school staff. Social capital is defined as resources for action, real or potential, that are attained *through relationships* (Bourdieu, 1981; Coleman, 1988; Lin, 1982). In schools and other organizations, relationships can be a source of resources such as trust (Bryk and Schneider, 2002) as well as knowledge and expertise (Coburn, 2001; Frank, Zhao, and Borman, 2004; Spillane, 2004). Most work on the school organization focuses on social capital at the level of the organization, with few exceptions (Penuel, Frank, and Krause, 2006; Coburn, 2005, 2007).

At the level of the individual, we operationalize social capital using out-degree and related measures of relationships upon which a teacher draws to accomplish their work. In the course of designing our instrument, prior research on social capital and previous findings regarding school organizations informed three important operational choices: that we measure an “advice or information” relation, that we differentiate among school subjects, and that we capture both ties within the organization and ties that span the organizational boundary.

We chose to use a criterion relationship defined by advice and information-seeking because the transfer of information is important for the creation of new knowledge and innovative practices in organizations such as schools. Prior research suggests that network structure is associated with the transfer and use of complex information as well as the diffusion of innovation in organizations (Frank, Zhou, and Borman, 2004; Reagans and McEvily, 2003; Uzzi, 1997; Uzzi and Lancaster, 2003). In schools, the transfer and development of complex and often tacit knowledge about teaching is essential to improving classroom teaching, which is the core work of schools.

We also chose to use a criterion relationship that differentiates among school subjects. Prior research suggests that the school subject is an important influence on teachers’ thinking, their classroom teaching, and their efforts to improve that teaching (Drake, Spillane, and Hufferd-Ackles, 2001; Spillane, 2000; Stodolsky, 1988). School staff members’ mental scripts for their work and the structure of relations among staff differ by school subject (Burch and Spillane, 2005; Hayton and Spillane, 2007; Spillane, 2006). Specifically, the frequency of interactions among school staff and the prominence of formally designated leaders in the networks differ depending on the school subject (Diamond and Spillane, 2006; Hayton and Spillane, 2007). Hence, the school subject is an important consideration in the structure of social relations among school staff, necessitating the collection of multi-dimensional network data.

Finally, we chose to use a name generator design that would capture both internal and external ties. Schools, like most organizations, are open systems that derive legitimacy, funding, ideas, and knowledge from their external environments (Meyer, Scott, and Deal, 1983; Scott, 1992). Further, prior research suggests that ties that span subgroups or organizational boundaries provide access to information that may not be within one’s immediate group (Coburn and Russell, 2007; Burt, 1992; Reagans and McEvily, 2003), and that external ties are important for creating and disseminating innovation and for productivity (Burt, 2000; Hansen, 1999). Arguing for attention to both internal and external ties, Adler and Kwon state:

Organizational research would benefit if we overcame the tendency to bifurcate our social capital research into a strand focused on external, bridging social capital and a strand focused on internal, bonding social

capital...Although the mechanics of research are simplified by restricting ourselves to a single level of analysis, the reality of organizations is shaped by the constant interplay of the individual, group, business unit, corporate, and inter-firm levels (2002, p. 35).

*Instrument design.* Our program of research has both substantive and methodological goals. One methodological concern is how to design a survey to measure social capital within schools, while differentiating between school subjects. To address this goal, we use a split-ballot experimental design to test whether the order of name generator prompts in the survey affects the validity of inferences made based on the resultant data.

Each name generator begins with the same wording: “In the past year, to whom have you gone for advice or information about [SUBJECT PROMPT]?” The flow of the name generator questions is tailored in two ways. The subject prompt is determined based on a respondent’s report of her role (contained classroom teacher, specialist teacher, or administrator), whether the respondent teaches mathematics, and whether the respondent teaches Reading/Writing/Language Arts.<sup>5</sup> Self-contained classroom teachers and administrators are asked about mathematics and reading, in random order. Specialist teachers are first asked about teaching their primary subject (if it is something other than mathematics or reading), then asked about both mathematics and reading, in random order. Specialist teachers whose primary subject is mathematics are always asked about mathematics first; likewise, specialist teachers whose primary subject is reading are always asked about reading first. Figure 1 depicts the randomized design of the name generators.

---

<sup>5</sup> In the remainder of this paper, we use the term reading to abbreviate Reading/Writing/Language Arts.

[Insert Figure 1 Here]

The precise wording of the subject prompt is also modified depending on role and whether or not a respondent teaches a particular subject. If a respondent teaches reading, they are asked about “teaching Reading/Writing/Language Arts,” whereas if the respondent does not teach reading, they are asked about “Reading/Writing/Language Arts as it relates to your classroom teaching.” A similar approach is used for mathematics. All administrators are asked about “Reading/Writing/Language Arts instruction” and “Mathematics instruction,” in random order.

Each name generator is followed by a series of name interpreter questions. For each alter that a respondent names, data is collected on the role or job description of the alter, the content of the advice interactions, the frequency of interactions between respondent and alter, and the respondent’s rating of the influence of the alter’s advice on her work.<sup>6</sup> Data on role or job descriptions is free-response text, and is collected for purposes of identifying unique individuals during data cleaning. Data on the content of advice interactions is broken into the following five categories, plus an “other” category for free-responses: deepening your content knowledge, planning or selecting course content and materials, approaches for teaching content to students, strategies specifically to assist low-performing students, and assessing students’ understanding of the subject. Data on frequency of interaction and influence are collected using multiple-choice scales.

The survey was designed to randomize the order of the math and reading prompts only in situations where the respondent either taught both subjects or taught neither. The analysis that follows is limited to those respondents and therefore excludes teachers and

---

<sup>6</sup> Note that the number of name interpreter items that a respondent is asked to answer depends on the number of names she lists in the name generator. The total length of the survey therefore depends in part on the length of a respondent’s list of alters.

specialists who teach only mathematics, teach only reading, or primarily teach other subject specific classes, but also teach math or reading (but not both). Limiting the analysis in this way ensures that, for each respondent, the only variation in wording between the name generators is in the academic subject. For example, comparisons are NOT made between responses to a name generator asking about “teaching Reading/Writing/Language Arts” and one asking about “Mathematics as it relates to your classroom teaching.”

*Data collection.* In the analysis that follows, we use results from two samples that were collected using our instrument. One sample consists of 15 public elementary schools and 4 Catholic elementary schools (most serving kindergarten through 8<sup>th</sup> grade) in a large U.S. city. School faculties vary in size from 14 to 69. All teachers, administrators, and school-level specialists were asked to complete the web-based survey during a six-week period in the Spring of 2007. In this sample, we received a full or partial response from 414 out of 544 staff (76%); of these, 264 contained randomized name generator prompts suitable for comparison.<sup>7</sup> The randomized portion of the sample is composed mostly of contained-classroom, primary grade teachers, with relatively few subject-area specialist teachers. Table 1 presents the number of respondents by sample, treatment group, and role.

[Insert Table 1 Here]

---

<sup>7</sup> School-level response rates range from 41% to 95%. The two treatment groups do not differ in mean age or sex composition. Some differences exist in racial composition; the M/R treatment group contains more African American, fewer Hispanic/Latino, and fewer White/Caucasian respondents.

The second sample consists of 10 public middle-schools in a mid-sized city in a different state, all serving grades 6 through 8. All teachers, administrators, and school-level specialists were asked to complete the web survey. School faculties range in size from 49 to 69 certified staff. We received a full or partial response from 548 out of 634 staff (87%); of these, 323 contained randomized name generator prompts suitable for comparison.<sup>8</sup> The randomized portion of the sample is composed mostly of subject-area teachers (teaching academic subjects other than mathematics and reading), and sixth grade teachers in self-contained classrooms. All of the subject-area teachers were first asked about teaching their primary subject, making the randomized prompts the second and third name generators in the survey. Sixth grade teachers were asked only about teaching math and teaching reading, in random order.

*Data analysis.* The survey design permits us to detect question order effects by comparing the two randomly assigned sub-sets of each sample. Approximately half of our respondents answered the math name generator and interpreter questions before the reading name generator; below we refer to this group as the M/R treatment. The remaining respondents answered the reading name generator and interpreter questions before the math name generator; we refer to them as the R/M treatment. For each subject area, we compare the data from the treatment where the name generator was posed first to the data from the treatment where the name generator was posed second.

We assume that assigning the two treatments groups creates a random partition of the out-degree distribution, so that any differences between these distributions are attributable to question-order effects. Further, taking column-totals of each treatment

---

<sup>8</sup> The two treatment groups do not differ in mean age, sex composition, or racial composition.

group produces a random partition of in-degree for each individual in the organization; in the absence of question-order effects, a given individual should receive an equal number of nominations from each treatment group.

Throughout, we use non-parametric tests of significance. Network degree distributions are typically very skewed; often one observes that a few individuals have many ties to other, while most other individuals have very few ties (Wong, Pattison, and Robins, 2006). Normality assumptions are likely to be invalid, making the use of t-tests inappropriate; Mann-Whitney tests provide an alternative that makes no distribution assumption.

## **Findings**

The two samples reveal a consistent pattern of question-order effects. We first examine effects on out-degree, discussing the evidence for fatigue effects versus satisficing effects. We then turn to non-redundancy and primacy effects, and conclude by examining evidence of question-scope redefinition.

For both samples, the reading name generator reveals significant differences between treatment groups in average out-degree; respondents in the M/R treatment list about one name fewer than respondents in the R/M treatment (see Table 2). This difference, which is on the order of a 40% decrease in the number of names generated, is large in magnitude. Results from the math name generator do not reveal a clear pattern. In the elementary school sample, respondents in the R/M treatment list an average of 0.24 fewer names for mathematics than respondents in the M/R treatment, a difference which is neither large in magnitude nor statistically significant. In the middle school sample,

respondents who received the math name generator second actually list more names, on average, than respondents who received the math generator first, though the difference is small and not statistically significant.

[Insert Table 2 Here]

The patterns of differences are largely consistent across groups of respondents with different formal roles. In the elementary school sample, differences between treatment groups in math out-degree are always small and insignificant; differences in reading out-degree are largest for special education teachers, followed by administrators, contained teachers, and subject-area teachers. This ordering seems intuitively reasonable, as subject-area teachers answer a name generator about their primary subject first, and might be expected to list fewer names regarding both reading and math.

In the middle school sample, the general pattern is followed by contained teachers, administrators, and subject-area teachers, but special education teachers display a perplexing pattern. Among the 40 special education respondents, the M/R treatment group lists many more names than the R/M treatment group in response to both the math name generator and the reading name generator. Checks for outliers among special education respondents did not reveal patterns in other variables that could explain the result. This perplexing pattern leads us to suspect that a problem occurred with randomization among special education teachers.

In both samples, question-order affects out-degree in a substantively meaningful way. If the purpose of our research were only to determine whether teachers sought more advice about reading or about math, we would reach opposite conclusions if we looked only at the R/M treatment or only at the M/R treatment. Data from the R/M treatment

suggests that the average reading out-degree (and therefore network density) is significantly larger than the average math out-degree. On the other hand, data from the M/R treatment suggests that the difference between reading and math is much smaller, and statistically insignificant. Pooling across the randomized treatment groups in the elementary school sample, we find an average reading out-degree of 2.5, significantly larger than the average math out-degree of 1.6. Pooling across treatment groups in the middle school sample, the difference between the average reading out-degree (1.9) and the average math out-degree (1.5) is smaller in magnitude but still statistically significant.

*Fatigue effects.* Recall that if fatigue effects are present in the survey, one would expect the number of names listed to decrease from the first name generator to the second. However, one would also expect that the distribution of the total number of alters named in both generators to be similar across treatment groups, because there is no reason for the two randomly-assigned treatment groups to differ in the amount of effort they are willing to exert. Such does not appear to be the case.

In the elementary school sample we observe significant differences between the two treatment groups in the total number of alters named. The M/R treatment group named 1.1 fewer alters, on average, than the R/M treatment group.

In the middle school sample, the M/R treatment group named 0.7 fewer alters, on average, than the R/M treatment group, though this difference is not significant for the sample as a whole. However, larger differences are observed in the subset of 6<sup>th</sup> grade contained-classroom teachers, the group of teachers whose role most closely resembles that of elementary teachers, and who answered only the math and reading name generators. The subject-area teachers answered a name generator regarding teaching their

primary subject before answering the randomized math and reading name generators, and the bulk of the names that these teachers listed came in response to this primary subject name generator. Perhaps because they were less salient for subject-area teachers, the subsequent randomized name generators produced too weak a response for meaningful differences to be detected between treatment groups.

*Satisficing effects.* A satisficing respondent chooses how much effort to exert in responding to the second name generator by using the precedent of her response to the first name generator. If the true size of the second network is larger than the reported size of the first network, the respondent will list only as many names as she did in response to the first name generator, because such a response seems sufficiently complete.

Effectively, the response to the first name generator creates a ceiling for the response to the second name generator.

Satisficing is consistent with the observed pattern of average differences in network size between the reading advice network and the math network. Across samples and treatment groups, the average reading out-degree is larger than the average math out-degree. The magnitude of the difference is much smaller when the math name generator is posed first, which is consistent with the hypothesis that respondents are limiting the number of names they list in response to a later name generator based on the number of names they list in the initial generator. If satisficing is creating bias in the number of names listed, as this evidence suggests, then one should look to the averages from the R/M treatment group only, rather than from the entire sample, to find the best estimate of the true difference between reading out-degree and math out-degree.

Satisficing effects also become evident by studying differences between reading out-degree and math out-degree at the level of the individual respondent. Respondents in the M/R treatment group are likely to have a difference close to zero, because their response to the first name generator creates a ceiling for their response to the second name generator. Only 33% of respondents in the elementary school M/R treatment group listed more names in the reading generator than in the math generator, whereas in the elementary school R/M treatment group, 69% of respondents listed more names in the reading generator than in the math generator (See Table 3). A similar pattern is observed in the middle school sample, though it is not as large in magnitude.

[Insert Table 3 Here]

Further, if responses to subsequent name generators are censored due to satisficing, one would expect an increase in the correlation of out-degrees. Table 3 reports that the reading out-degree and math out-degree are strongly correlated in each treatment group and each sample. In the elementary school sample, correlation in the M/R treatment group is higher than in the R/M treatment group. This is consistent with the theory that the M/R treatment group is affected by a satisficing ceiling, leading to increased correlation in out-degrees, while the R/M treatment group is not affected by the ceiling. In the middle school sample, a higher correlation appears in the R/M treatment group, which is not consistent with satisficing.

*Non-redundancy effects.* The design of our survey does not allow robust tests for non-redundancy effects because we have no way of estimating the true level of multiplexity in the two networks that we measured—the extent to which respondents seek advice about both math and reading from the same alters. For purposes of description, we

can only report on observed levels of overlap, in the form of Jaccard similarity coefficients. In the elementary school sample, 21% of all alters were named by the same respondent in both generators; in the middle school sample, 15% of alters were named by the same respondent in both generators.

*Cognitive priming.* A rough analysis of cognitive priming effects is possible by examining differences between treatment groups in the alters who are named. Cognitive priming tends to increase the likelihood that the alters or types of alters named in response to the first generator will also be named in the second name generator. Effects operate on the structure in which contacts are organized in a respondent’s memory. Teachers may think about their colleagues in terms of departmental affiliation, grade-level assignment, or formal role. Though other structures are certainly possible, such as classifying one’s colleagues by who socializes with one another, or who eats lunch together, our data only permit a test of priming by formal role, department, or grade-level.

In our split-ballot survey design, cognitive priming effects will be evident if the distribution of alters named for a given subject-area differs across treatment groups, and if within a given treatment group, the distribution of alters named in response to the second name generator resembles the distribution of alters named in response to the first name generator.

For each name generator, we calculate the percentage of respondents in each treatment group who list one or more alters with a given role. Though some differences between treatment groups are apparent, on the whole the composition of the networks is similar across treatment groups (see Table 4).

[Insert Table 4 Here]

In the elementary school sample, differences between treatment groups in the composition of the reading network could be explained in terms of cognitive priming. The largest difference is in the percentage of contained classroom teachers named in response to the reading name generator: fifty-nine percent of respondents in the R/M treatment named at least one contained teacher, compared to only 36% of respondents in the M/R treatment. Respondents who answered the math name generator first may be less primed to consider self-contained teachers in subsequent name generators, because the math network is less focused on self-contained teachers, and more focused on math teachers and specialists. Similar patterns of difference exist in the percentage of respondents naming alters who are Special Education teachers and alters who are administrators.

While the varying prominence of contained teachers may be explained in terms of priming effects, it is unclear why priming would affect results of the reading network but not the math network. This ambiguity, in addition to the lack of a consistent pattern in the middle school sample, leads us to conclude that evidence of cognitive priming based on staff’s formal roles is weak. Moreover, the data do not allow us to test for priming based on other possible organizations of alters in a respondent’s memory.

*Question scope redefinition.* Respondents understand the pragmatic meaning of questions by looking at the sequence of questions in a survey, as well other aspects of the design. In our survey, we observe how the scope of the second question could be redefined by the preceding name generator and name interpreter question. The name interpreter that follows the first name generator contains a bank of questions about the

dimension of instruction for which the respondent seeks advice from each alter, questions asking the respondent to rate the frequency of their contact with each alter, and questions asking the respondent to rate the influence of the alter’s advice on the respondent’s practice. We examine two different types of question-scope redefinition, one that has to do with the design of the survey and one that has to do with respondents’ understanding of the school subject areas that they teach.

First, the name interpreter questions provide additional, specific context that could influence the respondent’s understanding of subsequent name generators. The respondent may recall the descriptions of different dimensions of instruction as examples of issues about which they have sought advice, almost as if the second name generator read: “In the past year, to whom have you gone for advice about teaching Mathematics, *for example, about deepening your content knowledge, planning or selecting course content and materials, approaches for teaching content to students, strategies specifically to assist low-performing students, or assessing students’ understanding of the subject?*”

The five specific instructional dimension questions in the first name interpreter provide context that seems to be applied in answering the second name generator. In both treatment groups of both studies, the total number of the content areas checked increases from the first interpreter to the second interpreter. In the elementary school sample, the number of content areas per alter increases from 2.8 in the first name interpreter to 3.1 in the second interpreter (See Table 5). In the middle school sample, the average increases from 2.5 to 2.7.

[Insert Table 5 Here]

While these differences are not large in magnitude, the trend is clear. Across samples and treatment groups, nearly every specific category is checked with increased frequency in the second interpreter. Only the 6<sup>th</sup> “other” category is checked less frequently in the second interpreter. In the elementary school sample, the “other” category is checked 14% of the time in the first interpreter and 8% of the time in the second; in the middle school sample, the “other” category is checked 16% of the time in the first interpreter and 9% of the time in the second. The set of alters named in the second name generator appears to be a better fit for the categories of advice content, suggesting that the scope of the second question has been redefined by the content-area questions answered during the first name interpreter question.

The above analysis presents differences between the first and second sets of name interpreter data, averaging across treatment groups, but question-scope effects can also be analyzed by treatment groups. As table 6 reports, differences between treatment groups in the number of content-areas checked are observed for the math network questions, but not the reading network questions. In the elementary school sample, respondents in the R/M treatment group checked an average of 3.3 content areas (out of five) per alter, whereas respondents in the M/R treatment group checked only 2.6. In the middle school sample, the difference between treatment groups is smaller but still significant. Moreover, in both samples, each of the individual content areas in the math interpreter was checked more often by the R/M treatment group than by the M/R treatment group.

[Insert Table 6 Here]

The question-scope effects we observe here confound the possibility of comparing the different dimensions of instruction in the two subject-area networks. Suppose that we

are interested in learning whether teachers seek advice about a broader array of dimensions of instruction in reading or in math. Because of the question-scope redefinition created by the sequence of the survey (generator, interpreter, generator, interpreter), we would have reached different conclusions depending on the order in which the subject areas were measured.

Based on earlier theory building and hypothesis generating work (Diamond and Spillane, 2006; Hayton and Spillane, 2007), we believe that the differences between treatment groups are driven by respondents’ subject-specific thinking about advice-seeking. Question-scope redefinition effects would be observed if a respondent’s understanding of the scope of advice that is sought for a particular subject is carried over to the respondent’s interpretation of subsequent name generators. Our earlier work suggests that when elementary school staff interact about mathematics, conversations tend to focus on fewer dimensions of instruction compared to interactions about reading. If a respondent is asked about math first, she may apply this narrower understanding of math advice in responding to the reading name interpreter, and therefore check fewer dimensions of instruction. If she is asked about reading first, the broader understanding of reading advice is carried over to the math questions, so she checks more dimensions of instruction.

### **Discussion and design considerations**

In two samples collected using a multiple name generator survey that randomized the order of name generators, we find evidence of satisficing (rather than fatigue effects) and question scope redefinition. Evidence for non-redundancy effects and priming effects

is inconclusive. The effects for which we have found evidence are troubling because they are so closely related to the substantive questions that provoked our research. In particular, satisficing effects and question-scope redefinition effects create biases that would cause us to reach opposite conclusions depending on the order in which the name generators and interpreters were posed.

Our conclusions are limited in several of ways, some of which are suggestive of directions for further research. First, the design of our instrument does not allow us to test for non-redundancy effects, and provides for only a very rough test of cognitive priming. Both effects are strongly influenced by the amount of overlap between the networks being measured. Below, we suggest a survey design that would allow for tests of non-redundancy and priming.

Second, the analysis we have presented tests separately for the various types of question-order effects. In order to isolate the relative contribution of each effect (for example, to determine whether satisficing or cognitive priming is more important), an integrated model would be necessary.

Third, the validity and accuracy of any measurement depends not just on the instrument used to collect the data, but also on the particular statistic or metric that is applied to the raw data (Costenbader and Valente, 2003). We have focused only on very basic measures such as out-degree and in-degree. Whether more complex metrics such as closeness, betweenness, or transitivity indices could be affected by question-order remains a subject of future work.

Fourth, our results may not generalize to multiple name generator surveys that measure different sets of criterion relationships. We have measured and attempted to

compare two criterion relationships, advice about mathematics and advice about reading, that vary only in the school subject of interest. Both criterion relationships focus on the core work of school staff. Further methodological research is needed to examine question-order effects using criterion relationships that are less parallel. In the realm of organizational network analysis, sets of criterion relationships such as friendship, co-work, and advice-seeking should be tested. In the realm of personal network research, or studies of social capital, multiple name generators that measure instrumental support, emotional support, and social support should be tested.

Finally, we also urge caution in generalizing to other organizational settings. The cognitivist approach suggests that the accuracy of name generator recall depends on the degree to which respondents have a well-developed structure for storing memories of other people (Freeman et al., 1987). Biases created by question-order effects may be lessened to the extent that name generators specify criterion relationships in social systems for which respondents have good mental models. For example, corporate headhunters or community organizers might very likely have good mental models for their contacts, since they make constant use of them; teachers’ mental models of advice sources may be developed and accessed very differently than teenagers’ mental models of social support. Further research on cognitive models used in name generator recall should therefore be domain-specific, attending to the relationship between the research setting and instrument design.

Based on our findings, we conclude with some general suggestions about instrument design for capturing multidimensional network data, addressing the relative

merits of rosters versus name generators, considerations about the relationship between criterion relationships, and the sequencing of name interpreter questions.

For measuring single networks, others have recommended using complete roster methods whenever possible (Brewer, 2000). For multiple criterion relationships, roster methods would also seem to have an advantage; by not asking respondents to recall names from memory, attendant problems of fatigue, satisficing, non-redundancy, and priming effects could be avoided. However, roster-based methods may involve a considerable response burden, which must be weighed against the advantages of the design. In choosing roster-based methods over name generator surveys, the researcher might also be trading in one set of context effects for another. When posing a set of questions about each alter in an organization, response effects such as fatigue or satisficing might come into play based on the order of the alters in the roster. Further methodological research is needed to test the validity of roster-based designs for multiple name generator surveys.

In situations where roster-based methods are not feasible, multiple name-generator surveys should be designed with careful attention to the relationship between the criterion relationships of interest. The researcher should consider the relative density of the networks likely to be generated, the likely degree to which criterion relationships will overlap with one another, and how the criterion relationships may be perceived in relation to one another.

In studies where relative network size is of primary interest, minimizing the possibility of fatigue or satisficing effects is a key concern. If the mode of data collection

permits, one might consider randomizing the order in which the name generators are presented, so that the extent of fatigue or satisficing effects can be quantified.

In studies where the multiplexity of several criterion relationships is of primary interest, confounding processes such as non-redundancy effects or priming effects should be controlled. Interpreting the results of a multiple name generator survey, one might easily assume that the non-inclusion of a given alter in response to a name generator means that the alter does not fit the specified criterion relationship. However, if question-order effects are likely to bias the process of recalling names from memory, one should be wary of this assumption. Instead, the task of generating names should be separated from the task of interpreting information regarding those names.

To control for the possibility of non-redundancy or priming effects, name generators should all be run first, using specific criterion relationships or more general ones, and prompting the respondent to keep searching her memory if appropriate. Once a set of alter names has been generated, name interpreter questions could be posed that ask the respondent to classify the alter into one or more of the criterion relationships of interest. A similar approach has been applied in surveys that collect egocentric network data (see for instance Marin, 2004; Brewer, 2000 also cites a survey by L.M. Jones and C.S. Fischer that apparently uses a similar design).

In our own survey, incorporating this design would entail several steps. First, both the math and reading name generators are posed. Then, for each alter named only in response to the math name generator, the respondent is asked whether she also goes to that alter for advice about reading. For each alter named only in response to the reading

name generator, the respondent is asked whether she goes to that alter for advice about math.

Alternately, a name interpreter could be posed that consists of a grid of check-boxes, where each row contains the name of a unique alter and each column refers to a given criterion relation. In a web-based survey, it would be possible to seed the check-boxes so that alters named in response to the math name generator would be checked off on the math advice column of the name interpreter grid. The respondent would then be given the task of editing the grid to ensure that alters are appropriately classified.

Either of the above approaches would also allow a test of non-redundancy effects, which was not possible in the design presented in this paper. Non-redundancy effects would be present to the extent that respondents are named in only one subject-area name generator, but later re-classified in the name interpreter grid as giving advice in both subject areas.

## References

- Adler, P.S., Kwon, S, 2002. Social capital: Prospects for a new concept. *Academy of Management Review* 27, 17-40.
- Bailey, S., Marsden, P.V., 1999. Interpretation and interview context: examining the General Social Survey name generator using cognitive methods. *Social Networks* 21, 287-309.
- Bell, D.C., Belli-McQueen, B., Haider, A, 2007. Partner naming and forgetting: recall of network members. *Social Networks* 29, 279-299.
- Bernard, H.R., Johnsen, E.C., Killworth, P.D., McCarty, C., Shelley, G.A., Robinson, S., 1990. Comparing four different methods for measuring personal social networks. *Social Networks* 12, 179-215.
- Bernard, H.R., Killworth, P.D., 1977. Informant accuracy in social network data II. *Human Communication Research* 4, 3-18.
- Bernard, H.R., Killworth, P.D., Sailer, S., 1979/80. Informant accuracy in social network data IV: a comparison of clique-level structure in behavioral and cognitive network data. *Social Networks* 2, 191-218.
- Bernard, H.R., Killworth, P.D., Sailer, S., 1982. Informant accuracy in social network data V: an experimental attempt to predict actual communication from recall data. *Social Science Research* 11, 30-66.

- Bourdieu, P., 1981. Men and machines. In: Knorr-Cetina, K., Cicourel, A.V. (Eds.), *Advances in Social Theory and Methodology*. Routledge, London, pp. 304-317.
- Brewer, D.D., 2000. Forgetting in the recall-based elicitation of personal and social networks. *Social Networks* 22, 29-43.
- Brewer, D.D., Rinaldi, G., Mogoutov, A., Valente, T.W., 2005. A Quantitative review of associative patterns in the recall of persons. *Journal of Social Structure* 6.
- Bryk, A.S., Schneider, B., 2002. *Trust in Schools: A Core Resource for Improvement*. Russell Sage Foundation, New York.
- Burch, P., Spillane, J., 2005. How subjects matter in district office practice: instructionally relevant policy in urban school district redesign. *Journal of Education Change* 6, 51-76.
- Burt, R.S., 1992. *Structural Holes: The Social Structure of Competition*. Harvard University Press, Cambridge, MA.
- Burt, R.S., 1997. A note on social capital and network content. *Social Networks* 19, 355-373.
- Burt, R.S., 2000. The network structure of social capital. In: Staw, B.M., Sutton, R.L. (Eds.), *Research in Organizational Behavior*. JAI Press, Amsterdam, pp. 345-413.
- Coburn, C.E., 2001. Collective sensemaking about reading: how teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis*, 23, 145-170.

- Coburn, C.E., 2005. Shaping teacher sensemaking: school leaders and the enactment of reading policy. *Educational Policy* 19, 476-509.
- Coburn, C.E., Russell, J.L., 2007. District policy and teachers’ social networks. Working paper, Berkeley, CA.
- Coleman, J., 1988. Social capital in the creation of human capital. *The American Journal of Sociology* 94, Supplement: Organizations and Institutions: Sociological and Economic Approaches to the Analysis of Social Structure, S95-S120.
- Costenbader, E., Valente, T.W., 2003. The stability of centrality measures when networks are sampled. *Social Networks* 25, 283–307.
- Cowan, C.D., Murphy, L.R., Weiner, J., 1978. Effects of supplemental questions on victimization estimates from the National Crime Survey. In: *Proceedings of the American Statistical Association Section on Survey Research Methods 1978*. American Statistical Association, Alexandria, VA, pp. 277–282.
- Cross, R., Borgatti, S.P., Parker, A., 2001. Beyond answers: dimensions of the advice network. *Social Networks* 23, 215-235.
- Diamond, J.B., Spillane, J.P., 2006. High stakes accountability in urban elementary schools: challenging or reproducing inequality? *Teachers College Record* 106, 1145-1176.
- Drake, C., Spillane, J., Hufferd, K., 2001. Storied identities: teacher learning and subject matter context. *Journal of Curriculum Studies* 33, 1-23.

- Ferligoj, A., Hlebec, V., 1999. Evaluation of social network measurement instruments. *Social Networks* 21, 111-130.
- Frank, K.A., Zhao, Y., Borman, K., 2004. Social capital and the diffusion of innovation within organizations: the case of computer technology in schools. *Sociology of Education* 77, 148-171.
- Freeman, L.C., Romney, A.K., Freeman, S.C., 1987. Cognitive structure and informant accuracy. *American Anthropologist, New Series* 89, 310-325.
- Hammer, M., 1984. Explorations into the meaning of social network interview data. *Social Networks* 6, 341-371.
- Hansen, M.T., 1999. The search-transfer problem: the role of weak ties in transferring knowledge across organizational sub-units. *Administrative Science Quarterly* 44, 82-111.
- Hayton, P., Spillane, J., 2007. Professional community or communities? School subject matter and elementary school teachers' work environments. In: MacBeath, J., Chen, Y.C. (Eds.), *Leadership for Learning: International Perspectives*. SENSE, Rotterdam, Netherlands, pp. 59-71.
- Killworth, P.D., Bernard, H.R., 1976. Informant accuracy in social network data. *Human Organization* 35, 269–286.
- Kogovšek, T., Ferligoj, A., Coenders, G., Saris, W., 2002. Estimating the reliability and validity of personal support measures: full information ML estimation with planned incomplete data. *Social Networks* 24, 1-20.

- Krosnick, J.A., 2000. The threat of satisficing in surveys: the shortcuts respondents take in answering questions. *Survey Methods Centre Newsletter* 20, 4-8.
- Krosnick, J.A., Narayan, S.S., Smith, W.R., 1996. Satisficing in surveys: initial evidence. In: Braverman, M.T., Slater, J.K. (Eds.), *Advances in Survey Research*. Jossey-Bass, San Francisco, pp. 29-44.
- Lazega, E., Pattison, P., 1999. Multiplexity, generalized exchange and cooperation in organizations: a case study. *Social Networks* 21, 67-90.
- Lin, Nan, 1982. Social resources and instrumental action. In: Marsden, P.V., Lin, N. (Eds.), *Social Structure and Network Analysis*. Sage, Beverly Hills, CA, pp. 131-145.
- Marin, A., 2004. Are respondents more likely to list alters with certain characteristics? Implications for name generator data. *Social Networks* 26, 289–307.
- Marin, A., Hampton, K.N., 2007. Simplifying the personal network name generator: alternatives to traditional multiple- and single-name generators. *Field Methods* 19, 163-193.
- Marsden, P.V., 2005. Recent developments in network measurement. In: Carrington, P.J., Scott, J., Wasserman, S. (Eds.), *Models and Methods in Social Network Analysis*. Cambridge University Press, New York, pp. 8-30.
- McPherson, M., Smith-Lovin, L., Brashears, M.E., 2006. Social isolation in America: changes in core discussion networks over two decades. *American Sociological Review* 71, 353-375.

Meyer, J.W., Scott, W.R., Deal, T.E., 1983. Institutional and technical sources of organizational structure: explaining the structure of educational organizations. In: Meyer, J.W., Scott, W.R. (Eds.), *Organizational Environments: Ritual and Rationality*. Sage, Beverly Hills, CA, pp. 45-70.

Penuel, W. R., Frank, K. A., Krause, A., 2006. Distributed expertise and the implementation of schoolwide reform initiatives: a network approach to examining the effects of leadership practice. Working paper, Michigan State University.

Plickert, G., Côté, R.R., Wellman, B., 2007. It’s not who you know, it’s how you know them: who exchanges what with whom? *Social Networks* 29, 405-429.

Podolny, J.M., Baron, J. N., 1997. Resources and relationships: social networks and mobility in the workplace. *American Sociological Review* 62, 673-693.

Reagans, R., McEvily, B., 2003. Network structure and knowledge transfer: the effects of cohesion and range. *Administrative Science Quarterly* 48, 240-267.

Ruan, D., 1998. The content of the General Social Survey discussion networks: an exploration of General Social Survey discussion name generator in a Chinese context. *Social Networks* 20, 247-264.

Schwarz, N., 1999. Self-reports: how the questions shape the answers. *American Psychologist* 54, 93-105.

Schwarz, N., Oyserman, D., 2001. Asking questions about behavior: cognition, communication and questionnaire construction. *American Journal of Evaluation* 22, 127-160.

Scott, W.R., 1992. *Organizations: Rational, Natural, and Open Systems* (3rd Edition). Prentice-Hall, Englewood Cliffs, NJ.

Spillane, J.P., 2000. Constructing an ambitious pedagogy in fifth grade: the mathematics and literacy divide. *Elementary School Journal* 100, 307-330.

Spillane, J.P., 2004. *Standards Deviations: How Local Schools Misunderstand Policy*. Harvard University Press, Cambridge, MA.

Spillane, J.P., 2005. Primary school leadership practice: how the subject matters. *School Leadership and Management* 25, 383-397.

Stodolsky, S.S., 1988. *The Subject Matters: Classroom Activity in Math and Social Studies*. University of Chicago Press, Chicago.

Straits, B.C., 2000. Ego’s important discussants or significant people: an experiment in varying the wording of personal network name generators. *Social Networks* 22, 123-140.

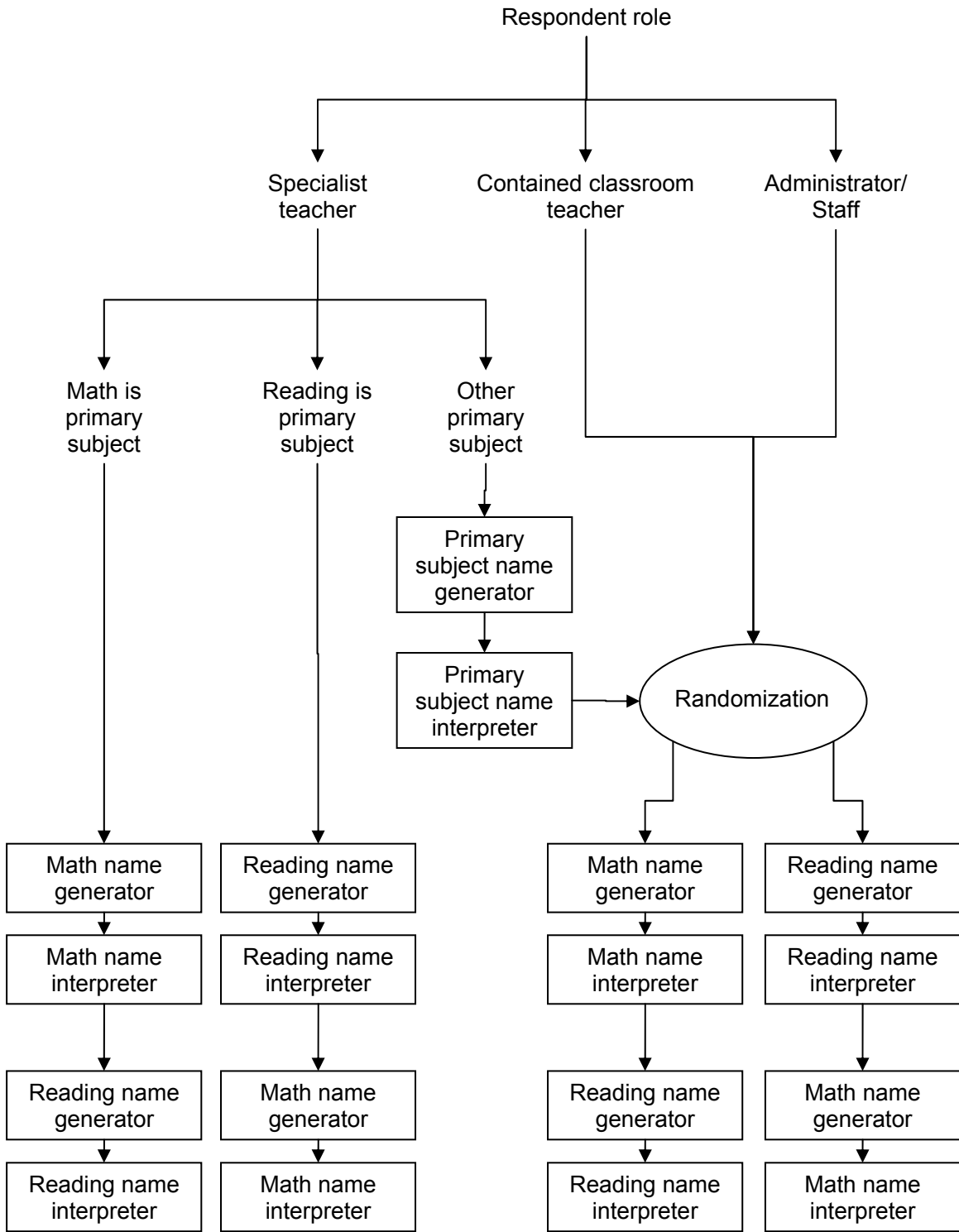
Sudman, S., 1985. Experiments in the measurement of the size of social networks. *Social Networks* 7, 127-151.

Sudman, S., Bradburn, N.M., Schwarz, N., 1996. *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. Jossey-Bass, San Francisco.

Tourangeau, R., Rasinski, K.A., 1988. Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin* 103, 299-314.

- Uzzi, B., 1997. Social structure and competition in interfirm networks: the paradox of embeddedness, *Administrative Science Quarterly* 42, 35–67.
- Uzzi, B., Lancaster, R., 2003. Relational embeddedness and learning: the case of bank loan managers and their clients. *Management Science* 49, 383-399.
- van Duijn, M.A.J., Snijders, T.A.B., Zijlstra, B.J., 2004. p2: a random effects model with covariates for directed graphs. *Statistica Neerlandica* 58, 234-254.
- Wong, L.H., Pattison, P., Robins, G., 2006. A spatial model for social networks. *Physica A: Statistical Mechanics and its Applications* 360, 99-120.

Figure 1. Name generator survey design



**Table 1. Number of respondents by treatment**

<i>Sample</i>	<i>Respondent role</i>	<i>Treatment</i>		<i>Total</i>
		<i>R/M</i>	<i>M/R</i>	
Elementary (K-8) Schools	Contained teachers	98	106	204
	Subject-area teachers	12	13	25
	Special education teachers	6	5	11
	Administrators	10	14	24
	Total	126	138	264
Middle Schools	Contained teachers	48	37	85
	Subject-area teachers	88	98	186
	Special education teachers	16	24	40
	Administrators	7	5	12
	Total	159	164	323

**Table 2. Average out-degree**

<i>Sample</i>	<i>Statistic</i>	<i>Treatment</i>		<i>Difference</i>
		<i>R/M</i>	<i>M/R</i>	
Elementary (K-8) Schools	Respondents	126	138	-12
	Reading Out-Degree	3.21	1.91	1.29*
	Math Out-Degree	1.46	1.70	-0.24
Middle Schools	Respondents	159	164	-5
	Reading Out-Degree	2.35	1.47	0.88*
	Math Out-Degree	1.64	1.32	0.32

\* Difference is significant at the 5% level according to a Mann-Whitney test.

**Table 3. Difference between reading out-degree and math out-degree**

<i>Reading out-degree minus math out-degree</i>	<i>Elementary (K-8) Schools</i>		<i>Middle Schools</i>	
	<i>M/R group</i>	<i>R/M group</i>	<i>M/R group</i>	<i>R/M group</i>
<=-3	2%	1%	5%	1%
-2	3%	0%	5%	5%
-1	13%	5%	10%	8%
0	49%	26%	51%	45%
1	20%	21%	16%	14%
2	10%	20%	9%	12%
>=3	3%	28%	5%	14%
Correlation between reading out-degree and math out-degree	0.75	0.59	0.69	0.78

**Table 4. Percentage of respondents naming an alter in a given role**

<i>Sample</i>	<i>Alter role</i>	<i>Math</i>			<i>Reading</i>		
		<i>M/R</i>	<i>R/M</i>	<i>Difference</i>	<i>M/R</i>	<i>R/M</i>	<i>Difference</i>
Elementary (K-8) Schools	Contained teachers	37	37	0	36	59	-23*
	Reading	1	1	0	14	13	1
	Math	21	21	0	3	8	-5
	Special Subjects	1	2	0	2	6	-4
	SPED	4	4	1	6	17	-11*
	Administrators	11	6	5	11	25	-15*
	Specialists	34	36	-2	55	59	-4
	Other	3	0	3	1	3	-2
External contacts	17	6	12*	16	16	0	
Middle Schools	Contained teachers	20	30	-11*	23	32	-10
	Reading	1	1	0	34	31	3
	Math	28	26	2	0	0	0
	Special Subjects	2	1	1	2	4	-2
	SPED	10	9	1	18	18	0
	Administrators	4	5	-1	5	13	-8*
	Other	1	0	1	2	1	1
	External contacts	11	14	-3	8	21	-13*

\* Difference is significant at the 5% level according to a Mann-Whitney test.

**Table 5. Percent of alters with specific instructional dimension checked**

<i>Sample</i>	<i>Content area</i>	<i>First interpreter</i>	<i>Second interpreter</i>	<i>Difference</i>
Elementary (K-8) Schools	Deepening your content knowledge	46%	51%	4%
	Planning or selecting course content and materials	66%	68%	2%
	Approaches for teaching content to students	62%	71%	9% <sup>2</sup>
	Strategies specifically to assist low-performing students	60%	63%	3%
	Assessing students’ understanding of the subject	51%	58%	7% <sup>2</sup>
	Other	17%	9%	-8% <sup>2</sup>
	Total number of content-areas checked per alter <sup>1</sup>	2.85	3.10	0.25 <sup>3</sup>
Middle Schools	Deepening your content knowledge	46%	40%	-6%
	Planning or selecting course content and materials	50%	54%	4%
	Approaches for teaching content to students	59%	65%	5%
	Strategies specifically to assist low-performing students	50%	56%	6% <sup>2</sup>
	Assessing students’ understanding of the subject	43%	53%	10% <sup>2</sup>
	Other	19%	10%	-9% <sup>2</sup>
	Total number of content-areas checked per alter <sup>1</sup>	2.48	2.67	0.19 <sup>3</sup>

Notes:

1. Excludes the non-specific “other” category.
2. Difference is significant at the 5% level according to a Fisher exact test.
3. Difference is significant at the 5% level according to a Mann-Whitney test.

**Table 6. Average number of instructional dimensions checked per alter,  
by treatment group**

<i>Sample</i>	<i>Name interpreter</i>	<i>Treatment</i>		<i>Difference</i>
		<i>R/M</i>	<i>M/R</i>	
Elementary	Reading	3.0	2.8	0.2
(K-8) Schools	Math	3.3	2.6	0.8*
Middle Schools	Reading	2.6	2.5	0.1
	Math	2.6	2.3	0.3*

\* Difference is significant at the 5% level according to a Mann-Whitney test.