

THE DELPHI TECHNIQUE AS A METHOD FOR INCREASING INCLUSION IN THE EVALUATION PROCESS

Christina A. Christie
Claremont Graduate University
Claremont, California

Eric Barela
Los Angeles Unified School District
Los Angeles, California

Abstract: The question of how best to integrate the views of underrepresented and marginalized groups in the evaluation process is of critical importance to many evaluation theorists and practitioners. In this article the Delphi technique, a method used to achieve consensus on a set of issues with the participation of all interested parties without incident or confrontation that could compromise the validity of collected data, is offered as a procedure for enhancing marginalized group participation in the evaluation process. Demonstrated by a case example, the Delphi technique is used to help ensure that all relevant stakeholders have a voice and that sometimes-silenced voices have equal influence. As a result, it is suggested that this technique lends itself to implementation with social justice evaluation models. The benefits of and lessons learned when using the Delphi technique to promote marginalized group participation and representation in evaluations are discussed.

Résumé: La meilleure façon d'intégrer les points de vue des groupes sous-représentés et marginalisés au processus d'évaluation est une question d'une grande importance pour de nombreux théoriciens et praticiens de l'évaluation. Cet article présente la méthode Delphi — une méthode utilisée pour atteindre un consensus sur un certain nombre de questions avec la participation de toutes les parties intéressées et sans qu'il y ait d'incident ou de confrontation risquant de compromettre la validité des données recueillies — comme procédure pour rehausser la participation des groupes marginalisés au processus d'évaluation. Illustrée à l'aide d'une étude de cas, la méthode Delphi contribue à permettre l'expression de l'opinion de tous les intéressés et à s'as-

Corresponding author: Christina A. Christie, Claremont Graduate University, School of Behavioral and Organizational Sciences, 123 E. 8th St., Claremont, CA, USA 91711; <tina.Christie@cgu.edu>

surer que les voix parfois silencieuses aient une influence égale. Les auteurs préconisent cette technique particulièrement pour les modèles d'évaluation de la justice sociale. On discute également des avantages et des leçons apprises concernant la méthode Delphi pour promouvoir la participation et la représentation des groupes marginalisés dans les évaluations.

When studying social programs, evaluators are often concerned with how best to capture and include the views of underrepresented groups. Historically, the Delphi technique has been used by evaluators and program stakeholders to determine group consensus on issues related to the importance of program components and effectiveness through questionnaires and focused feedback. In this article, we offer the Delphi technique as a method for enhancing the level of involvement of marginalized program stakeholders in the evaluation process. Using the Delphi technique increases participation by including all relevant stakeholders and ensuring that they all have a voice, and that previously silenced voices have equal influence. Thus, we argue that this technique lends itself to implementation with social justice oriented evaluation approaches.

UNDERREPRESENTED GROUP PARTICIPATION IN THE EVALUATION PROCESS

When discussing underrepresented group participation in evaluation, we are referring to two broad categories of people: those who are marginalized because of their position in society (i.e., ethnicity, race, gender, sexual orientation) and those who are marginalized because of their position in the program (program participants/recipients, less powerful program staff). And, although it is seemingly apparent, we want to note an intersection of these two groups; those who are marginalized because of their proximity of the program are often also members of groups that are marginalized by society. It is the intersection of these groups that we refer to in this article.

The inclusion of traditionally marginalized groups contributes to the evaluation in important ways. Madison (1992) outlines three for us. First, when underrepresented groups are included in the evaluation, the potential for achieving an accurate understanding of the social benefits of the program is increased. Second, stakeholders from traditionally marginalized groups are sometimes in the best position to assess the aspects of the program that will work for a similar group of participants. And third, the inclusion of these perspectives

may lead to the development of strength-based program models, rather than ones based on a perceived deficit of such groups. Additionally, there are instances when evaluators know little about the program participants' world views (Lincoln, 1991). Therefore, including stakeholders from underrepresented groups in the evaluation process is often the best way to account for and include the views and perspectives of such groups.

It is the responsibility of the evaluator to determine how and when marginalized groups are included in the evaluation process as key program stakeholders. When included, the evaluator must use methods that are appropriate for determining the interests of underrepresented stakeholders, as well as methods that help to ensure that information collected from *all* stakeholders has equal influence. Determining how to integrate the views of underrepresented groups in the evaluation process can be challenging—our repertoire of methods (particularly survey methods) seemingly limited. In our case example, we turn to an established method, the Delphi technique, to help promote the inclusion of underrepresented groups and determine group consensus on issues related to the impact of the program process and outcomes through the use of anonymous questionnaires and focused feedback.

SOCIAL JUSTICE IN EVALUATION

Much attention has been paid to social justice issues in evaluation. Modern commentary about the need for evaluations that promote social justice began as early as the 1960s, as the field of evaluation was emerging (e.g., House, 1972). As the field developed so too did models of evaluation, which offer approaches that are intended to foster and support social justice. The goal of these approaches is to employ program evaluation to empower the disenfranchised (Stufflebeam, 2000).

Two of the more notable scholars on the subject are House and Fetterman. Each offers a different perspective with respect to the importance of promoting social justice in evaluation, as well as different models for practice. Fetterman is concerned with empowering the program community, which he defines as program staff and clients. His philosophy of social justice involves teaching people to define program goals, gathering information to help understand if the goals are being met, and using that information in ways that improve the program. The program community, in effect, conducts

its own evaluation. He calls this approach *empowerment evaluation*, the goal of which is to foster self-determination (1996).

In comparison, House argues that evaluation should advance the interests of the least privileged (1991). He is less concerned with empowering program people, but rather ensuring that those who are often unjustly ignored are given a voice. House asserts that evaluation establishes “who gets what” (House, 1980) and identifies race, ethnicity, and gender as areas where evaluation has served to promote social justice or, at the very least, has recognized injustices. Thus, any underrepresented group in an organization can be considered a group without equal voice. However, marginalized groups are defined in many ways, not just by race, ethnicity, or gender.

House claims that every evaluation has a value slant from its inception that includes motivations, biases, values, attitudes, and political pressures. He also recognizes the inevitable impact of an evaluator’s subjectivity, emphasizing the need to employ scientific rigour and ensure validity. Accordingly, evaluators are encouraged to explicitly state their valuations [judgements], verify their findings, and pinpoint data-based conclusions. However, House states, “It is hidden unseen valuation that is damaging and that leads to opportunistically distorted research findings, for covert valuations allow us to pursue our base interests at the expense of proper justification. We trick ourselves as well as others” (1972, p. 413).

To promote social justice through the evaluation process, evaluators need tangible, practical methods. House offers a model for doing so, as do others such as Lincoln (1991); however, most require the use of intensive qualitative interviewing methods. Such methods require stakeholders from underrepresented groups to reveal their identity. In some instances, this may dissuade participants from being entirely truthful or perhaps from taking part in the evaluation. The Delphi technique is an alternative method for involving participants from traditionally marginalized groups, which allows for representation while protecting the identity of those involved. As a result, we suggest using the Delphi as a method for promoting social justice in evaluation by increasing the involvement and including the voice of those from underrepresented groups.

THE DELPHI TECHNIQUE

Dalkey and Helmer of the RAND Corporation developed the Delphi technique in the 1950s. The purpose of this technique is to achieve

consensus on a set of issues with the participation of all interested parties without incident or confrontation that could compromise the validity of collected data. It was first used to determine group consensus on sensitive, often classified, military projects. Due to the nature of these projects, it was determined by senior military officials that group meetings might not yield the most accurate results about project effectiveness (Helmer, 1967). They felt the validity of the data, if collected in a group setting, would be threatened by the “bandwagon effect.” They were also quite concerned that those in less powerful positions would fear punitive action if they held views that differed from those in more powerful positions and that stakeholders, particularly those from marginalized groups (however defined), may not be willing to divulge candid information at a group meeting.

Since its development the Delphi technique has become a tool commonly employed by researchers and evaluators in business and industry (Fitch et al., 2001). Often it is used to distill knowledge and determine consensus from a group of experts who are often separated by geography (Adler & Ziglio, 1996). The Delphi technique is also used widely to forecast innovations, for example, in the area of technology. In this context, an expert panel is convened to determine consensus on the likelihood of developing and utilizing the proposed innovation as well as on a proposed timeline for development (Levary & Han, 1995).

There are relatively fewer examples of the use of the Delphi technique in social and educational program evaluations. Because program staff can often be found in the same facility, it is often more cost-effective to convene a group meeting. However, convening a group meeting can sometimes be unadvisable in the context of a program evaluation, particularly when relationships between program staff and program managers could lead to collecting bias consensus data (Adler & Ziglio, 1996). As Martino (1970) reminds us, when using the Delphi technique, the study outcome is solely based on opinion. Thus, the validity of the study is inextricably linked to the validity of the opinions of the stakeholders. Therefore, if stakeholder opinions are biased as a result of the “bandwagon effect” or insufficient representation of all relevant stakeholder opinions, results of the study will be of little use.

Benefits of the Delphi Technique

Using the Delphi technique is desirable when there is a lack of agreement or incomplete knowledge concerning either the nature of the

program or the components that must be included in a successful program (Delbecq, Van de Ven, & Gustafson, 1975), for example, if program managers do not know what components of the program are being implemented by staff. It is also meant to eliminate personal conflicts and allow for a better opportunity to reach consensus about program effectiveness (Delbecq et al.). Additionally, it allows for a representation of diverse views and interests based on experience and expertise (Linstone & Turoff, 1975).

Differences in perception and opinion are the basis for using the Delphi technique. Each stakeholder brings a different reality and a diverse set of experiences to the program. Stakeholder perceptions are important because they drive program implementation (Scheele, 1975). If a stakeholder perceives a component of a program as unnecessary for the program's success, it is very likely that the stakeholder will not support the implementation of that particular component. The Delphi technique operates on the assumption that if all stakeholders are included and can respond anonymously, accurate data can be gathered to determine where perceptions are helping and where they are hindering program implementation.

While it is desirable to achieve consensus, it is unlikely that enough members of a group will agree on every point raised in a Delphi study. Therefore, another purpose of the Delphi technique is to clarify the "dissensus" within a group. This clarification is important because it allows group members to figure out ways to address the differences in perception and opinion as a group.

Frequently, consensus is obtained by holding a round-table discussion with all parties present and is usually reached through compromise, with the "weaker" parties yielding to the influence of the more prominent parties (Helmer, 1967). The Delphi technique can be an important planning aid for educational policy-makers because it eliminates the bias of overpowering voices and the "bandwagon effect" (Adelson, Alkin, Carey, & Helmer, 1967). When the voices of a few dominate the overall evaluation effort, true program effectiveness is difficult to determine due to the fact that many of those involved in the program implementation are being silenced.

Lastly, the Delphi technique disguises individual answers. Because stakeholder responses cannot be identified, there is very little chance that responses on a Delphi questionnaire can be used punitively (Linstone & Turoff, 1975). It is for these reasons that we believe the

Delphi method is effective for increasing the representation of individuals from traditionally marginalized and underrepresented groups.

Steps of the Delphi Technique

Before an evaluator begins a Delphi study, three questions should be asked: (a) What kind of group communication process is desirable in order to explore issues at hand? (b) Who possesses expertise about the issues that are to be explored? (c) What alternative techniques are available for data collection and what results can be reasonably expected from their application? (Adler & Ziglio, 1996). These questions can be addressed by meeting with a group of stakeholders (this group may not include all stakeholders, but should include representatives from all constituencies) or by interviewing all relevant stakeholders about what they consider to be the pertinent issues. The evaluator must keep in mind that stakeholders from marginalized groups may want to address a different set of issues, some of which may not have been previously considered. Thus, it is imperative that all relevant perceptions from stakeholders be examined.

Once stakeholders have identified substantive program issues, the evaluator develops a Delphi questionnaire. Each item of the questionnaire should correspond to a specific issue that was discussed in the initial conversation with program staff and should be scored using a Likert scale. The evaluator works with stakeholders to determine the structure, content, and format of the scale. Scale development is often a process of negotiation between the evaluator and stakeholders. For example, the evaluator may opt for a 10-point scale to assess the degree of program effectiveness while the stakeholders may prefer a 3-point scale that seeks to identify whether the program is working, is intermittently working, or is not working.

As with the development of any survey instrument, the evaluator must also keep questionnaire length and wording in mind. A lengthy questionnaire is more likely to elicit inaccurate data. In an effort to complete a long survey, responses might be selected without full consideration. Or, if items are vaguely worded, stakeholders may inadvertently respond inaccurately. Thus, as with other instrument development, it is recommended that the questionnaire be piloted on a small group of stakeholders (Fowles, 1978). Evaluators should also be aware that when obtaining data at the feedback stage (de-

scribed below), respondents may change their decisions and potentially render questionable results (Linstone & Turoff, 1975).

Once the issues of item selection, overall length, response scale, and proper wording are resolved, the questionnaire is written and disseminated to all stakeholders. Instruments are disseminated to stakeholders who helped to identify the program issues from which the questionnaire items were developed as well as those who did not participate in this process. It then becomes the responsibility of the evaluator to ensure that every potential respondent completes the questionnaire.

Respondent dropout should be of concern when considering the implementation of a Delphi study (Sackman, 1975). Considering Scheele's (1975) assertion that perception can drive program implementation, potential dropout and non-participant rates must be considered when using this technique. Thus, we recommend that the evaluator systematically document who declines to participate and report reasons for non-participation. A relevant stakeholder may choose not to participate for a variety of reasons including lack of time, the belief that the technique may not yield accurate data on program effectiveness, lack of confidence in program effectiveness, or fear of reprisal (Sackman, 1975). Evaluators should attempt to interview non-respondents formally or informally whenever possible because, as we well know, stakeholders who do not participate often yield data just as important as those who do. Explanation for non-participation should be documented in the final report, as this can inform perceptions of program effectiveness. This is especially important for programs seeking to increase marginalized stakeholder participation.

After all questionnaires have been collected, consensus for each item must be determined. The criteria for consensus can be determined either by the evaluator alone or with stakeholder input. Typically, consensus is said to be reached on a given item when a certain percentage of respondents are within a specified distance from the mean or a specified standard deviation value. For example, an evaluator may determine that consensus is reached when at least 75% of participants' responses fall between two points above and below the mean on a 10-point scale or when the standard deviation is less than 1.5. Once a criterion for consensus has been developed, the evaluator calculates item means and frequencies. If enough respondents scored within the specified distance from the mean, consensus has

been reached for that item. For example, if an item mean is 3 and 90% of respondents score the item between 1 and 5 or has a standard deviation of 1.25, consensus has been reached on that item based on the criteria set. It is important to consider how consensus will be defined for each program being evaluated. Some clients may want 100% agreement on every item. Others may feel that 50% or 75% agreement is sufficient. The evaluator and the client should consider the importance of each item when considering the consensus criterion (Helmer, 1967) for a particular program. While some items may be more important than others, they should all be held to the same consensus criterion in order to maintain equal representation of all issues being considered.

Consensus is not reached if an insufficient number of respondents scored within the specified distance from the mean on a particular item. In order to reach consensus and clarify dissensus, it is important for respondents to consider alternate viewpoints and rethink their opinions on the controversial items (Helmer, 1967). To accomplish this, the evaluator selects at least one respondent who scored the item above the specified distance from the mean and at least one respondent who scored the item below. The evaluator interviews each respondent to determine the rationale for his or her response. The evaluator should intentionally select a diverse group of interviewees to ensure the inclusion of participants from all constituencies. This interviewing process is often referred to as the “feedback” stage of the Delphi technique. All feedback statements must remain confidential and the identity of the speaker must be withheld.

Once the feedback statements have been gathered, the evaluator develops a second iteration of the questionnaire. This version of the questionnaire includes the items on which consensus was not reached, accompanied by the corresponding (anonymous) feedback statements. Also included is the mean score for the item. The objective of this format is to provide additional information so that stakeholders can better determine their own position on the issue.

After the second-round questionnaire, the evaluator, using the same criterion from the first-round questionnaire, determines consensus on the items. For those items where consensus was not reached, the evaluator interviews stakeholders to gather additional feedback statements. Wherever possible, the respondents selected for the second-round interviews should be different from those selected for the first round of interviews. A third-round questionnaire is then devel-

oped using the feedback statements from both the first and second rounds, and the process is repeated.

The Delphi technique is flexible with respect to the number of iterations. Consensus for many items may be reached after the first round; however, a second round is almost always necessary. Indeed, with larger samples (e.g., 30 or more), three rounds is often recommended before concluding the Delphi study (Dalkey, Rourke, Lewis, & Snyder, 1972; Helmer, 1967; Linstone & Turoff, 1975). By the end of the third round, the evaluator should be able to determine those items where consensus has been reached and explain the nature of the dissensus on items where consensus was not reached. The purpose of multiple iterations is to clarify the group position on the effectiveness of a program component or the importance of an issue. If a dissensus exists between relevant stakeholders, it will emerge in the Delphi study. It is not the goal of this technique to force all respondents to feel one way about every program component or issue, but to give all stakeholders the opportunity to determine consensus and clarify dissensus truthfully and without fear of reprisal.

It is important to note that the Delphi technique is only appropriate for implementation in situations where the evaluator can first identify issues that are most important to staff and where data related to perception are desired. It should be noted that this technique is not appropriate when program outcome data are desired. Perceptions may relate to summative issues of program effectiveness, but they are not a substitute for reliable outcome data (Helmer, 1967). By generating data that adequately represent how all relevant stakeholders perceive issues such as program implementation and effectiveness, the Delphi technique can help the evaluator provide formative conclusions. This method is grounded in the assumption that program perception drives program implementation and, as such, data yielded from a Delphi study can provide program administrators with information pertaining to current program implementation, effectiveness, and areas of improvement as they are defined by all stakeholders, not just those who have had traditional representation in the evaluation process.

With respect to stakeholder involvement when using the Delphi method, a few points should be clarified. Stakeholder participation is a required condition for a Delphi study, which takes place at various stages of the evaluation; however, the evaluation is conducted primarily by the evaluator. At the beginning stages of the evalua-

tion, stakeholders collaborate with the evaluator to determine relevant programmatic issues to be studied and to develop study instruments. Stakeholders must also participate as the source for data; not only do stakeholders provide initial data, but they are also required (by definition) to provide feedback on item consensus and dissensus once data have been collected and summarized. Stakeholders do not, however, participate in the actual collection and analysis of study data; this is left to the evaluator. Thus, the Delphi method is not a participatory approach to evaluation as described by evaluation theorists such as Cousins and Earl (1995), but rather a method that can increase participation and representation in the evaluation process.

CASE EXAMPLE

Many of the academic outreach efforts at the University of California, Los Angeles (UCLA) resulted from the passage of Proposition 209, which abolished California's affirmative action policy in 1996. From the fall of 1997 to the fall of 1998, UCLA saw large declines in the admissions of underrepresented students, and in an effort to increase the number of educationally disadvantaged students admitted to the university, UCLA developed and instituted a vast network of activities with local school districts centred on educational reform. These efforts can be seen as having two objectives: (a) to foster systemic change related to college readiness (school-centred programs), and (b) to implement and sustain programs that assist prospective students in becoming UC eligible (student-centred programs). The student-centred outreach program, which is the focus of the study described in this article, is designed to increase the likelihood of select groups of promising middle and high school students becoming competitive for UC admissions through an intensive intervention delivered by UCLA's Early Academic Outreach Programs (EAOP) office.

Researchers and school district personnel developed a conceptual model for UCLA's academic outreach programs, identifying six conditions necessary for meeting the program objectives: the creation of a college-going *culture*, the building of a multicultural college-going *identity* amongst students, a rigorous academic *curriculum*, high-quality *teaching*, curricular and extracurricular academic *support* for students, and the fostering of parent and community *connections* with each school. One of our first roles as evaluators was to determine the validity of the academic outreach program theory by

examining the activities implemented. We focused first on the student-centred program theory and activities.

When contemplating the study design, there were two important issues to consider. First was the organizational structure of the program. The majority of the program staff were in the schools delivering services (activities). As a result, program staff were often left out of program planning meetings, which were run by program administrators and typically held during school hours. Second was the ethnic composition of the outreach staff. Program administrators and planning staff, predominantly university faculty and administrators, were mostly members of the dominant culture. The majority of the program staff delivering the student-centred activities in the schools, however, were from ethnically diverse groups that are typically thought to be underrepresented. We were concerned that the organizational structure of the program as well as the difference in ethnic composition between program administrators and staff could potentially result in a perceived discrepancy in power between the two groups. But we knew that honest opinions of program staff regarding program implementation were vital to the fidelity of the evaluation. After much thought, we chose to design a Delphi study because of its sensitivity to anonymity, which would help address potential power discrepancies and allow for participation of all respondents without fear of punitive action, and because of its survey format, which allows for program participants to participate with relatively little inconvenience and reduce the gathering of irrelevant information by controlling and focusing feedback (Dalkey et al., 1972).

To begin, we interviewed program staff and participants to determine the exact set of activities that were being delivered. All student-centred program staff were interviewed, and collectively they noted 13 different program activities that were being delivered. Once the set of activities was compiled, we developed an instrument to measure staff beliefs about how each activity related to each of the six conditions. Respondents were asked to indicate whether the activity (a) had little or no impact; (b) had some impact; (c) had great impact; or (d) was not applicable for each condition. Thirteen activities were being considered in relationship to 6 conditions; thus, the first-round Delphi questionnaire consisted of 78 items (13 x 6). This questionnaire was mailed to 19 UCLA Early Academic Outreach Program staff members, including the program director and administrators. Fifteen questionnaires were returned.

Consensus was determined item by item. If greater than 50% of respondents answered that an activity had a particular impact on a specific condition, we determined that to be consensus. With the four-choice response scale, our view was that to obtain a majority on any single item constituted substantial agreement. After the first round, 29 of 78 items did not reach consensus. We determined that there were enough unresolved items for a second-round questionnaire.

For items where consensus was not reached, we gathered feedback on the extreme responses through individual interviews. During the interviews, respondents were asked to justify and clarify their points of view related to the item's impact on each of the six conditions. We deliberately interviewed a diverse set of participants to ensure that the perspective of traditionally underrepresented stakeholders was represented. The second-round questionnaire consisted of the feedback statements for the extreme responses and descriptive statistics for each of the 29 items. Fourteen of the 15 first-round respondents completed the second-round questionnaire. After the second round, there were only two items for which consensus had not been met; thus, it was decided that another iteration of the Delphi study would not be necessary.

DISCUSSION

The Delphi method was originally conceived as a technique to gain consensus on sensitive issues. Our translation of the Delphi technique as a method to promote the inclusion of those that may not typically be involved in an evaluation presents a new interpretation of the technique. As demonstrated through the case example presented in this article, it is our view that the Delphi technique can serve as a vehicle for involving minority stakeholders and eliminating the power imbalances in stakeholder groups when attempting to evaluate perceptions of program effectiveness. As such, it is a tool that can be used to promote social justice in evaluation.

In our Delphi study we determined consensus and clarified dissensus, while considering all relevant stakeholders and not granting particular perceptions more weight than others. By most standards, this was a successful Delphi study. We attribute this success to study participation (almost all staff members participated), questionnaire length (questionnaires were not excessively long), and respondents' timely completion of both questionnaires and their provision of thoughtful feedback statements. In addition, when data were

analyzed, we were able to determine the similarities between participants' perceptions with regard to the overall impact of each activity on the overall program theory, and we were able to clarify the dissensus.

It is worth noting that in a Delphi study it is the evaluator's responsibility to secure near-complete participation in the study—because without participation there can be no representation. Indeed, we worked diligently to encourage participation. Respondents were sent multiple copies of each questionnaire, and the evaluators followed up by e-mail and with phone calls to ensure that questionnaires had been received. The evaluators also met individually and at locations convenient for respondents to gather interview data (feedback) after the first round. Diligence and flexibility helped to ensure participation.

With respect to increasing the representation of underrepresented groups in the evaluation, our Delphi study, we believe, transcended the organizational and power structures of the group. EAOP staff members, for various reasons including availability, were excluded from discussions pertaining to outreach program planning and effectiveness. By using the Delphi technique to determine which activities were most strongly related to program effectiveness, all respondents were represented in the process with an equal voice. Thus, the observations of the program director, program staff, and the associate vice chancellor were all given equal weight in our study findings and recommendations. Additionally, program staff expressed appreciation for the opportunity to participate in the evaluation and to have their views documented.

Further, we would like to suggest that the method chosen to study program activity implementation and effectiveness also promoted evaluation use. In subsequent informal conversations, program staff indicated that because study findings were perceived to be an accurate reflection of the program as they experienced it, they were more inclined to implement the program changes suggested by the program director and administrators. This is important to note, because program changes are difficult to implement without the support from those responsible for realizing them — the program staff.

Connecting evaluation theory to practice. We would like to suggest that the Delphi technique can be used to promote the social justice ideals consistent with both House and Fetterman's theoretical approaches by increasing participation and representation of under-

represented groups in the evaluation process. The technique, as described in our case study, gives equal weight to the perceptions of all stakeholders, while determining the patterns of group consensus and dissensus. By maintaining respondents' anonymity, the possibility of maintaining equality and ensuring representation of all voices increases. Let us conclude by connecting a bit more pointedly the relationship between House and Fetterman's models and the Delphi technique.

Fetterman's approach — empowerment evaluation — is a collaborative group activity. It is democratic in that it invites the participation of whoever has a vested interest in the program, and examines issues of concern to the entire community (1996). Program participants, including staff and clients, determine the type of evaluation they desire and their own goals and strategies for accomplishing such an evaluation. Thus, Fetterman's approach supports the use of methods that promote the inclusion of all stakeholders, and that gives equal voice to all involved in the process. Consequently, the Delphi technique offers the empowerment evaluator a method for accomplishing social justice when seeking to clarify consensus on an issue.

Different from Fetterman both theoretically and in practice (Christie, 2003), House promotes an evaluation approach that stresses the importance of plural values when rendering judgements about a program's worth. He contends that it is "morally correct for evaluators to represent within their evaluations the interests and needs of those unjustly ignored" (1991, p. 245). House, with Howe, offers a framework that links evaluation to the larger sociopolitical and moral structure, called *deliberative democratic evaluation* (House & Howe, 1999). House and Howe describe deliberative democratic evaluation as a theory that advocates for evaluation studies that are objective and impartial regarding fact and value claims. They state:

If we look beyond the conduct of individual studies by individual evaluators, we can see the outlines of evaluation as an influential societal institution, one that can be vital to the realization of democratic societies. Amid the claims and counterclaims of the mass media, amid public relations and advertising, amid the legions of those in our society who represent particular interests for pay, evaluation can be an institution that stands apart, reliable in the accuracy and integrity of its claims. (1999, p. 4)

The authors maintain that when

only a few people decide social policy, an aristocracy, plutocracy, or technocracy exists, depending on whether talent, money, or expertise is the source of authority.... If the relevant interests are not considered, we have only a sham deliberation from which some voices have been excluded. (House & Howe, 1999, pp. 5–6)

Given the theoretical and practical purposes of House's social justice evaluation, the Delphi technique, as described in the context of our case example, can be used to promote these ideals. House argues that all stakeholders, but particularly those who are typically marginalized, not only have the right, but are entitled to representation in an evaluation. As demonstrated in our study, the Delphi technique accomplishes this, while ensuring that the voices of those from underrepresented groups have equal power and expression.

REFERENCES

- Adelson, M., Alkin, M., Carey, C., & Helmer, O. (1967). Planning education for the future. *American Behavioral Scientist*, 10(7), 8–19.
- Adler, M., & Ziglio, E. (1996). *Gazing into the oracle*. Bristol, PA: Jessica Kingsley.
- Christie, C.A. (2003). What guides evaluation? A study of how evaluation practice maps onto evaluation theory. *New Directions for Evaluation*, 97, 7–35.
- Cousins, J.B., & Earl, L.M. (Eds.). (1995). *Participatory evaluation in education: Studies in evaluation use and organizational learning*. London: Falmer Press.
- Dalkey, N., Rourke, D.L., Lewis, R., & Snyder, D. (1972). *Studies in the quality of life: Delphi and decision-making*. Lexington, MA: Lexington Books.
- Delbecq, A.L., Van de Ven, A.H., & Gustafson, D.H. (1975). *Group techniques for program planning: A guide to nominal group and Delphi processes*. Glenview, IL: Scott, Foresman.

- Fetterman, D.M. (1996). Empowerment evaluation: An introduction to theory and practice. In D.M. Fetterman, S.J. Kaflarian, & A. Wandersman (Eds.), *Empowerment evaluation* (pp. 3–46). Thousand Oaks, CA: Sage.
- Fitch, K., Bernstein, S.J., Aguilar, M.D., Burnand, B., LaCalle, J.R., Lazaro, P., van het Loo, M., McDonnell, J., Vader, J.P., & Kahan, J.P. (2001). *The RAND/UCLA appropriateness method user's manual*. Santa Monica, CA: RAND Corporation.
- Fowles, J. (1978). *Handbook of futures research*. Westport, CT: Greenwood Press.
- Helmer, O. (1967). *Analysis of the future: The Delphi method*. Santa Monica, CA: RAND Corporation.
- House, E.R. (1972). The conscience of educational evaluation. *Teachers College Record*, 73, 405–414.
- House, E.R. (1980). *Evaluating with validity*. Beverly Hills, CA: Sage.
- House, E.R. (1991). Evaluation and social justice. Where are we? In M. McLaughlin & D. Phillips (Eds.), *Evaluation and education at quarter century* (pp. 233–247). Chicago: University of Chicago Press.
- House, E.R., & Howe, K.R. (1999). *Values in evaluation and social research*. Thousand Oaks, CA: Sage.
- Levary, R.R., & Han, D. (1995). Choosing a technological forecasting method. *Industrial Management*, 37(1), 14–18.
- Lincoln, Y.S. (1991). The arts and sciences of program evaluation. *Evaluation Practice*, 12, 1–7.
- Linstone, H.A., & Turoff, M. (1975). *The Delphi method: Techniques and applications*. Reading, MA: Addison-Wesley.
- Madison, A. (1992). Primary inclusion of culturally diverse minority program participants in the evaluation process. *New Directions for Evaluation*, 53, 35–43.
- Martino, J. (1970). The precision of Delphi estimates. *Technological Forecasting and Social Change*, 1, 293–299.

- Sackman, H. (1975). *Delphi critique*. Lexington, MA: Lexington Books.
- Scheele, D.S. (1975). Reality construction as a product of Delphi interaction. In H.A. Linstone & M. Turoff (Eds.), *The Delphi method: Techniques and applications* (pp. 215–219). Reading, MA: Addison-Wesley.
- Stufflebeam, D. (2000). *Evaluation models* (New Directions for Program Evaluation, No. 89). San Francisco: Jossey-Bass.

Christina A. Christie is an assistant professor in the School of Behavioral and Organizational Sciences, Claremont Graduate University.

Eric Barela is a doctoral candidate at the University of California, Los Angeles, and a research fellow at the Program Evaluation and Research Branch, Los Angeles Unified School District.