

Delphi Panels: Research Design, Procedures, Advantages, and Challenges

Jay R. Avella

University of Phoenix, Tempe, Arizona, USA

JayAvella@email.Phoenix.edu

Abstract

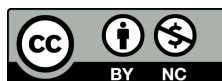
Among the typical dissertation research designs, one particular design that is slowly gaining acceptance is that of the Delphi Method. Using a panel of experts to achieve consensus in solving a problem, deciding the most appropriate course of action, or establishing causation where none previously existed, particularly in areas of business or education research, are uniquely ideal to employment of the design. This article reviews the origins of the method, provides detail on assembling the panel and executing the process, gives examples of conventional and modified Delphi designs, and summarizes the inherent advantages and disadvantages that the design brings. The article closes with some advice for those contemplating its use in their dissertations.

Keywords: Delphi, consensus, dissertation research, problem-solving, expert panel, critique

Introduction

Qualitative research designs come in many forms, with the case study perhaps the most frequently employed in doctoral dissertations. (Calendar year 2014/2015 abstract statistics reflecting dissertation research designs indicated almost 9,500 case studies, with the over 1,250 grounded theory studies the next most frequently indicated.) Others employed less frequently include phenomenology, ethnography, and narrative inquiry. One design that is becoming increasingly popular among student qualitative researchers pursuing their dissertations is that of the Delphi Method. This article traces how it came about, discusses elements of the design and its execution, provides examples of different types of the design, and articulates advantages and disadvantages.

The method takes its name from the ancient Greek city that housed the “oracle.” There, a priestess (called the “Pythia”) purportedly communicated directly with the gods and would answer questions (deBoer & Hale, 2002). This more recent Delphi was developed by the Rand Corporation under U.S. government contract in the 1950s as a method to forecast likely outcomes from nuclear weapons usage in war. Norman Dalkey and Olaf Helmer, two Rand mathematicians, developed the method and their approach was based



This article has been copyrighted by the Informing Science Institute on behalf of its authors. We license it to you under a Creative Commons Attribution-NonCommercial 4.0 International License. When you copy and redistribute this paper in full or in part, you need to provide proper attribution to it to ensure that others can later locate this work (and to ensure that others do not accuse you of plagiarism). You may (and we encourage you to) adapt, remix, transform, and build upon the material for any non-commercial purposes. This license does not permit you to use this material for commercial purposes.

on two key principles and the need to compensate for both of them (Dalkey & Helmer, 1963).

First, when it came to forecasting, there were two extremes on which individual predictions were typically based. On the one end was *knowledge*, which was based on evidence, and the other was *speculation*, which lacked any evidence and was basically an “educated guess.” In between was *opinion*, which was the result of an individual’s integration of the two extremes (Dalkey & Helmer, 1963).

Editor: Matthew Kemp

Submitted: January 12, 2016; Revised: May 22, June 28, July 26, August 12, 2016;

Accepted: September 9, 2016

Second, face-to-face group discussions where participants voiced individual opinions and arrived at a collective conclusion were often less accurate than the opinions of individuals when averaged without discussion (Dalkey & Helmer, 1963). Scholars and students of group dynamics know that dominant individuals can control the conversation and, in that way, eventually the outcomes. The one who speaks out loud and often can prevail over the group, even though that individual may not be the most knowledgeable (Fischer, 1978). When one combines that with the pressure for conformity that often exists within a group, characterized as *groupthink* by Janis (1982), it is no wonder this difference in accuracy resulted. Neither would have affected an average without discussion outcome (Fischer, 1978).

Dalkey and Helmer's method and approach provided a foundation for what has since become known as "futures research" (Von der Gracht, 2008). By using subject matter experts in an anonymous environment and bypassing those weaknesses found in meetings and conferences, researchers have been able to accurately forecast the development of a number of things that have since become components of everyday life. Among these were oral contraceptives, organ transplants, synthetic proteins, ultralight materials, and the economic desalinization of sea water (Amant, 1970). And while a manned landing on Mars was not considered feasible in 1970 (called a "miss" by Amant), NASA has a very different viewpoint today.

The Delphi design falls under the general category of "consensus development techniques," which in turn are under the general grouping of action research approaches (Vernon, 2009). Consensus techniques are typically applicable when there is limited evidence or when the existing evidence is contradictory in the specific topic of interest. Delphi itself is uniquely applicable in areas where there is little prior research or where advantage could be realized in the collective subjective judgment of experts (Hejblum et al., 2008). It has also been applied in large, complex problems plagued with uncertainty and in situations where causation can't be established (Yang, Zeng, & Zhang, 2012).

Delphi is predominantly qualitative in nature, but it can have a quantitative component depending on the specific application. As such, its primary characteristics match those of interpretivism. Examples of qualitative-only and partially quantitative strategies are presented later in this paper.

The Delphi Panel Process

Design Overview

The basic design involves assembling groups of experts without concern for geography, and who then reply to a number of "rounds" involving response to a specific question or questions through e-mail (Linstone & Turoff, 2002). After each round, participants receive feedback of the group response which typically takes the form of points of agreement listed in order of most- to least-often mentioned.

Historically speaking, the Delphi method falls into one of three versions which differ by their purpose. A "Policy" Delphi is used when there is a need to devise a strategy to address a specific problem; a "Classical" Delphi is used to forecast the future; and, a "Decision-Making" Delphi is used to achieve better decision making. While these design versions may differ in purpose, the execution of the design can take many different forms irrespective of the purpose. Different versions may execute the exact same or very different designs, depending on the specific study objective, as will also be discussed later in this paper.

The rounds process repeats itself with the goal of reducing the range of responses until "consensus" is achieved (Linstone & Turoff, 2002). With each repetition, specific responses would receive increasing or decreasing mention, eventually being pared down to an outcome acceptable to

all. It is worthwhile noting at this point that consensus does not mean 100% agreement, as it might be extremely difficult to get groups of individuals representing different constituencies with varying viewpoints and priorities to reach unanimity. Delphi consensus typically ranges from 55 to 100% agreement, with 70% considered the standard (Vernon, 2009). Dalkey and Helmer found that early responses exhibited wide ranges of alternatives, but were quickly distilled after very few iterations (Fischer, 1978).

Role of the Researcher

The role of the researcher is twofold: the first is that of “planner,” and later that of “facilitator” as opposed to “instrument” in the case of more traditional qualitative designs. In carefully designed and executed panels, the risk of researcher bias are minimal, if not nil, as the researcher’s primary task is that of planner/coordinator/recorder, and the back-and-forth communication between researcher and panel members provides for internal process auditing. One should note that “contributor” appears nowhere in the responsibilities of a Delphi researcher.

In planning a Delphi study, the principal tasks include identifying the discipline, number, and content of groups, and establishing the method and procedures of communication. Given the evolution of communication technology in recent years, it is perhaps hard to imagine a Delphi study being conducted through the mail using written or typed documents, but there may be occasions when this might be necessary. Most, if not all, Delphi dissertation studies will be conducted using e-mail, and in certain, very rare circumstances, if feasible, panel members might be brought together at the end to finalize the study outcomes. Melynk, Lummus, Vokurka, Bursm, and Sandor (2009) convened a Delphi panel consisting of supply chain managers and leading figures in supply chain research, with the objective to identify and prioritize the key issues and challenges facing the discipline as it transitioned from a tactical business practice with focus on cost and delivery, to one more strategic in nature where it became a point of differentiation from competitors. The panel established the priority list and later were invited to come together to discuss and extend the findings. Practically speaking, however, it’s highly doubtful that a doctoral candidate will have the personal resources to bring together panel members for a face-to-face session. Such events are typically associated with professional and corporate organizations.

The first area of consideration in planning a Delphi panel includes identifying the disciplines to be invited to participate in the panel. The concept of stakeholder in the study’s purpose is particularly relevant here. Researchers should ask themselves “which groups have a professional interest in achieving the study purpose?” The answer provides indication of those groups who should make up the panel. Studies involving student classroom behavior, for example, might include groups of teachers, parents, and behavioral specialists as participants. Studies involving airport designs would include community planners, engineers, architects, airport operations experts, and aviation personnel, among others. And finally, a current study under supervision by the author has convened a panel of educators, managers, and behavioral specialists to identify the “soft skills” necessary for Millennials (those born after 1980) to succeed in the 21st Century workplace, and to determine how best to incorporate those skills in college courses and curricula.

Who qualifies as an “expert” invited to participate is of critical importance. In the Melynk et al. study (2009), publication in scholarly journals provided a minimum qualification threshold for researcher participation on the panel. Generally speaking, participant invitation criteria should include those measurable characteristics that each participant group would acknowledge as those defining expertise, while still attempting to recruit a broad range of individual perspectives within those criteria. If the participant group were to consist of college faculty, for example, one might establish an academic rank (e.g., Associate Professor or above), having achieved tenure, or a minimum time on faculty criterion, and select faculty meeting that threshold while representing different academic disciplines. Researchers must ignore the appeal of becoming an arbiter of who

participates. Studies where the researcher claims to be the primary judge of relevant participant experience invite skewed outcomes.

It's also important to avoid the temptation to select members of a group who are "representative" of the discipline involved. Choosing a representative sample is something typically sought out in quantitative studies so the results can accurately portray the total population. Representation is not a quality actively pursued in Delphi studies; expertise is. There is no intent to extrapolate panel outcomes to any larger group, or to predict what another panel might conclude. The objective is to include individuals who can speak knowledgeably from the position of the group to which they belong. Generally speaking, the more diverse the perspectives of potential panel members, the broader the number and type of alternatives the panel will produce and consider. Finally, a researcher planning a Delphi study using international participants should require fluency in the language chosen for the panel.

In the researcher's role as facilitator, the task is essentially one of controlling the debate. The very nature of panel selection wherein a broad range of perspectives is purposely sought, when taken together with the anonymity provided by the process, leads to and perhaps even encourages individual panel members to present opinions that might be considered extreme. As long as the facilitator avoids being opinionated and functions in a non-judgmental manner, these "outliers" can receive a level of attention equivalent to the more popular opinions, although they might never achieve consensus. The benefit of the Delphi design lies in finding those areas where the panel finds consensus, and even the most extreme position expressed might trigger other panel members to alter their positions to some degree, or give rise to alternatives not previously considered.

The Basic Design

As noted above, the basic design involves assembling groups of experts who are geographically remote, and who then reply to a number of "rounds" involving response to a specific question or questions through e-mail (Linstone & Turoff, 2002). There is no standard or typical number of groups that would constitute the panel, although two or three groups are those most often seen in the literature. The number of groups necessary for participation should be based on those stakeholder groups most directly affected by the topic of the study. For example, Gjoligaj (2014) employed a Delphi panel to establish a sports club ("team" in some countries) management program at an Albanian university for her Management Education program dissertation. She convened a panel of educators, sports club/team managers, and government officials and asked them to list the competencies that should be developed in the management program, and then to rank them from the most to the least important. After three rounds, the panel had agreed on 11 competencies to be integrated into the university program, ranging from leadership (the most frequently listed) to facility management (the least often listed). The educator, club/team manager, and government official groups had a direct and significant stake in the quality of the study outcome, and in the eventual outcome of the university program once developed. One might think that athletes presumably had an interest in only the "best" educated team managers, but they were not direct stakeholders in the education process, nor were they concerned with every aspect of sports club/team management.

The size of the overall panel is another consideration. There is no standard when it comes to panel size; neither has it ever been established what constitutes a large or small panel. Akins, Tolson and Cole (2005) noted that panels have been conducted with just about any size. They also noted that panels of less than 10 are rare, as are panels over 1,000. Typical panels seem to fall in the 10 to 100-member range and consist of either two or three expert groups, again depending on stakeholder interest. Researchers should be attentive to balancing membership across expert groups as much as possible. Even with geographically diverse and anonymous participation, one particular

expert group could still dominate the process to some degree if it were significantly larger than the others.

As with any dissertation research effort, the design process begins with a problem statement, and in the flow typically associated with a dissertation, this leads to the purpose statement, which in turn generates the research question(s). It is from the research question that the researcher is not only able to establish the appropriateness of the Delphi panel design over other alternatives, but also to design the panel structure and establish the composition and criteria for membership. The research question itself can also provide the starting point for panel deliberations.

Key Panel Design Characteristics

There are two design characteristics that are critical to all Delphi panels, irrespective of topic or approach, and are inviolate and irreducible, and they are anonymity and feedback. Without both, any presumed Delphi design is flawed. The designers of the method sought to encourage debate that was independent of individual personality and influence of professional reputation. The designers also wanted to ensure all contributions received equal weight, at least at the start, and only through panel debate might any individual contribution be modified or eliminated.

Anonymity

The first characteristic critical to the execution of the design is participant anonymity (Yousuf, 2007). Recalling that averaging opinions of individuals collected separately was often more accurate than opinions reached through face-to-face discussion, and noting that dominant participants and *groupthink* limited the effectiveness of the face-to-face group, keeping panel members isolated from each other allows each individual freedom of expression without outside pressure or influence. Use of the web and conducting panel business by e-mail through the researcher/facilitator is particularly effective (and is essentially required) to maintain this privacy and confidentiality. All panel members should communicate individually with the researcher. The researcher needs to exert extreme care in ensuring that communication with each panel member is maintained on that individual basis. It is even possible that two individuals sharing an office could be members of a Delphi Panel but be unaware of the other's participation.

Feedback

The second design characteristic critical to executing all Delphi panels is feedback. Panel deliberations begin with one or more questions for individual members to consider. The results of the initial question(s) are collected and consolidated by the researcher/facilitator and then returned to panel members in a series of iterations (called "rounds") until consensus is reached. In each iteration, panel members are asked to review the outcome of the previous round and either agree with that outcome or recommend changes along with their rationale in making those changes. Without the ongoing feedback embedded in the round procedures, the process is more akin to generic inquiry than the consolidated opinion of experts. The various ways in which iterations might be implemented are discussed later in this paper.

Panel Membership

Selecting individuals who meet expertise qualifications for panel membership is critical and cannot be overstressed. It is well worth noting that it is the respective disciplines of the panel members that determine what those qualifications are, and not the researcher. It is probably not necessary to recruit a recent Nobel laureate in medicine for a study on identifying the ideal treatment sequence for a specific type of cancer, for example; but it would be appropriate to seek out oncologists who regularly treat that type, as well as medical researchers seeking a cure for the condition as prospective members of the panel. Years in specific practice and holding specific certifica-

tions or credentials are examples a researcher might use in choosing panel members. Panel membership criteria should be measurable and identifiable, but not subject to researcher judgment.

Selection of expert participants can also bring with it the potential for bias in that the individuals selected for the panel can bring with them positions known to the researcher. Hasson, Keeney, and McKenna (2000) make a case for seeking impartiality in recruiting panel members, but this might not always be possible as experts bring with them their own preconceptions. It can be avoided in many ways, however, one of which was employed in the supply chain panel reported earlier in this piece (Melynk et al., 2009). In that case, the researchers were identified through their scholarly publications, and the practitioners through contact with professional societies. Presumably no panel member who ultimately participated was known personally to the researchers before the study commenced. As in interview research, the ideal would be to establish qualification criteria for participant selection and then apply those criteria in recruiting panel members. Seidman's (2006) Phase 1 suggests a process by which interview candidates in any qualitative study are qualified for study participation and is conceivably applicable to Delphi panel recruitment. To preclude risk of bias, any known relationship, including casual acquaintance, between researcher and potential panel members should be an exclusionary criterion (Murphy et al., 1998).

Each group participating in the panel could also have different criteria that determined if an individual was an expert in their respective field. It's perhaps intuitive that Delphi panels that include different disciplines would look to each of those disciplines for the expert threshold. That said, no matter what the topic or problem any given panel could be asked to address, there are certain criteria that apply to membership on all Delphi Panels (Akins et al., 2005):

Interest

Potential participants should express interest in the topic and a willingness to participate through to project completion. Interest can be determined in any of a number of ways: through an academic publication record, via membership in a professional society/working group dedicated to the topic, through networking websites like LinkedIn, or simply by contacting potential panel members and inquiring. Interest in the topic can of itself generate willingness, but it alone is not sufficient. It is worthy of note that, despite interest and willingness, panels have reported member participation varied from round to round (Rupprecht, Birner, Gruber, & Mulder, 2011)

Time

Potential participants should have time available to dedicate to panel activities. Depending on the problem/topic the panel is asked to address, activities can be very time consuming (Williams & Webb, 1994). As a general rule, the larger the panel overall, the higher number of groups comprising the panel, or the more complex the topic the panel is asked to address will individually or collectively demand greater amounts of time on the part of panel members. While it is doubtful that a researcher could reasonably approximate the time commitment required of panel members, ensuring that prospective panel members are made aware of participation expectations as part of the recruitment process can reduce attrition later.

Written communication

Ability of potential panel members to communicate in writing is critical. In a world where communication is increasingly conducted through "sound bite" technology (e.g., Instant Message, Facebook, Twitter) or in formal presentations using PowerPoint as opposed to issuing formal reports, it is important that panel members be able to articulate their written positions clearly and succinctly. In a typical second or subsequent round, when panel members receive the previous panel-generated list of alternatives and are asked to respond with any changes along with their rationale, the ability of a participant to articulate his or her reasoning becomes critical. Eliciting a

“that makes sense” response from other panel members is highly dependent on the persuasiveness of the logic expressed an individual’s written argument.

In the case of an international panel, it’s also important that members be fluent (at least reasonably so) in the language identified for panel activities. That language can be the one native language most prevalent among panel members or one based on a specific requirement for a language in which panel members shared fluency. In the Gjologaj (2015) study, Albanian was the most common native language, but English was the panel’s “official” language due to intent to present/publish in English. That panel members possessed English fluency was a participation criterion.

Typical Delphi Designs

Throughout the literature, researchers will often see designs listed as either “Delphi” or “Modified Delphi.” Delphi (called “Conventional Delphi” herein) is defined the process wherein panel experts initiate the alternatives in response to the researcher’s question(s). Modified Delphi indicates the process whereby the initial alternatives in response to the researcher’s questions are carefully selected before being provided to the panel (Custer, Scarcella, & Stewart, 1999).

The Delphi method can be applied in contexts that exhibit varying mixtures of quantitative and qualitative techniques. The format presented to the panel together with the technique used to determine its outcomes are what determine a particular panel’s design. Following are some examples.

Conventional Delphi Designs

Conventional Delphi designs employ a group communications process targeted at achieving consensus through a series of questionnaires presented to an expert panel in multiple iterations (Hsu & Sandford, 2007). The researcher/facilitator asks the questions and records, consolidates, and transmits panel responses for each iteration until consensus is achieved. Looking at application of Delphi in a hypothetical example, followed by summaries of two actual studies can provide perspective on the range of possibilities in using this design.

Hypothetically speaking, suppose a researcher sought to convene a panel consisting of high school guidance counselors, college admissions officers, and college faculty to address the research question: “In order of priority, which are the five most important factors college admissions officers should consider in the admissions decision?” The process would begin with the researcher asking the question in order to solicit as many opinions as possible based on panelists’ own experience and expertise. Panel members would be given a reasonable amount of time to consider the question and respond. After collecting all responses, the researcher/facilitator would then tabulate the results and generate a list of factors based both on how often a factor appeared on the submissions and where it appeared on each list, in order to provide the panel an overall indication of their collective judgment. This could be done by assigning inverted point values to each factor on each list. The most important factor listed would be awarded 5 points and the 5th factor would receive 1 point. Point totals for all factors listed would be combined in descending order of point totals.

How the subsequent rounds would be executed can exhibit some variations, usually depending on the complexity of the question(s) asked. In the admissions example, the initial list (without the point totals) would be circulated for round 2 to the panel, and members asked to review, comment, and revise the list if warranted or to approve as is (this is rare). In those cases where a panel member revised the list, that member would be asked to provide a rationale for the change. The researcher would consolidate the responses (including areas where changes were recommended and explained) and circulate this round 2 list to the entire panel for round 3, asking the same

questions (review, comment, revise, or approve). Any subsequent rounds would continue to refine the list of factors until consensus was achieved, although in most cases, this third round typically achieves the 70% agreement necessary for consensus (Vernon, 2009).

In another example, the researcher might ask panel members to classify the factors according to the impact they might have on admission. An individual factor might be listed as “critical,” meaning failure to meet a minimum standard would be grounds for denial of admission. Another factor might be listed as “desirable,” meaning that failure to meet a minimum standard would not of itself cause a denial of admission, but failure to meet standards in more than one category might result in denial. Subsequent rounds might seek agreement on the critical factors, and only if the resulting list failed to identify the initially asked “five most important factors” would the desirable factors be considered.

Rivera (2013) convened a panel of 31 allied health professionals (physical therapists, occupational therapists, and certified child life specialists) from various clinical practice settings across the United States. Round 1 began with three questions that sought to define community reintegration, identify the barriers to reintegration, and isolate the most effective treatment strategies for community reintegration in adolescents and young adults with spinal cord injuries. Round 1 produced a total of 161 responses to the three questions, but after eliminating duplications, 44 themes resulted. Themes were phrased into statements to initiate Round 2 (e.g., “Community integration is defined as . . .”), and respondents asked to indicate relative agreement with each statement using a 7-point Likert scale. Only 10 (of the original 31) panel members responded to Round 2, but all three disciplines were still represented. Rivera used measures of central tendency (means and medians) and standard deviations to determine agreement. A similar process was implemented for Round 3 and yielded a 92.5% agreement across the panel. It should be noted that in analyzing panel contributions using the Likert scale, the author’s criterion for agreement with the statement was based on whether the scale mean and median (rating of “4” of the “7” options) was exceeded. Because the study questions did not add any qualifiers, such as the “single most critical barrier” or the “three most critical barriers” to reintegration, all statements that exceeded a mean and median of “4” were deemed to be in agreement.

In the case of more complex studies that generate extensive lists of alternatives for panel consideration, studies like the Wynaden et al. (2014) effort that narrowed mental health nursing research priorities might not use either of the above two hypothetical processes. While first two rounds in their study involved invitations sent to all nurses in Western Australia, the final round involved more stringent screening criteria in order to achieve consensus within a reasonable number of iterations. Panel members in round one were asked to identify the five most important issues for research. A total of 97 nurses responded with 390 individual suggestions that the researchers condensed into five categories using thematic analysis software. This allowed the researchers to narrow the 390 individual responses into 56 broad research questions. The second round involved some 127 nurses, who were tasked with rating the relative importance of the 56 questions using a 5-point Likert scale. This yielded a range of research questions from the most important—“Are there alternative primary care models that can be adopted to reduce the pressure on acute inpatient mental health beds?” (p.20)—to the least important—“Do uniforms promote identify and professionalism in mental health nurses?” (p. 21). The 10 most important research priorities from round 2 were submitted to a panel of senior mental health nurses to obtain their consensus in ranking these 10 questions in order of importance.

Modified Delphi Designs

As noted previously, Modified Delphi designs typically do not consult the expert panel to generate answers to the round 1 question(s). Rather the researcher collects the initial answers to the

question(s) through some other means and presents them to the panel to begin the consensus-seeking process. Some examples of this approach include the following:

- Researcher collects an initial list of responses based on a review of the relevant literature and disseminates it to the expert panel. The panel would then be asked to rank the list according to a specific criterion provided by the researcher. Panel members would be encouraged to add to this initial list based on their own experiences in order preclude researcher from controlling or limiting alternatives.
- Researcher conducts a series of interviews with individuals either within or outside the study panel, summarizes the results, and presents them to the panel. Using thematic analysis software (as a guard against researcher interpretation), an initial list of alternatives would be generated and disseminated to the panel. If the interviews were conducted with individuals external to the panel, panel members would also be encouraged to add to this initial list based on their own experiences.
- Researcher provides the results of a survey administered to a group external to the panel. While this approach eliminates the risk of interpretation or intervention by the researcher, panel members would still be encouraged to add to the list of alternatives based on their own experiences.

McBride (2015) used the interview approach to generate the Round 1 alternatives. The purpose of the study was to determine the leadership theory or style most effective in leading employees with diagnosed Post-Traumatic Stress Disorder (PTSD). She assembled a panel of 10 individuals in leadership positions within the Federal government, both military and civilian, with service tenure ranging from 17 to 37 years, who currently supervise (or have supervised in the recent past) federal employees who were diagnosed with PTSD according to accepted diagnostic standards. The interview data was subjected to analysis to identify common themes and patterns using thematic analysis software, which were then subjected to two rounds of panel activities to determine consensus. While both transformational and situational leadership were isolated by the panel as being effective in dealing with PTSD-diagnosed employees, situational leadership was ultimately identified as the most effective.

Joyner and Smith (2015) convened a modified Delphi panel to ensure whether a proposed faculty-developed framework for an undergraduate curriculum in dairy manufacturing was in alignment with current industry needs. The 21 panel members represented the dairy industry, regulatory agencies, and academic institutions. Panel members were presented the proposed curriculum and courses, and two rounds were used to gather the feedback. The first round was used to determine how the curriculum/course content met the needs of the current dairy industry and to determine if there were any gaps or excessive overlaps. Because the overall objective of the study was to ensure alignment between the curriculum and courses with the needs of the industry, the focus was on ensuring that the program providing graduates with the necessary entry-level skills as opposed to higher level positions. Panel members were asked to provide feedback in the form of a 5-point Likert scale rating for each course and how appropriate they were to the curriculum and to the needs of the industry. The results of Round 1 were used to make changes to the curriculum and courses and these were presented to the panel for round 2. The overall Likert scale ratings for the round 2 curriculum and courses were overwhelming rated as “highly appropriate” (Likert level 5) or “appropriate” (Likert level 4), with the preponderance of ratings achieving the highly appropriate rating.

Advantages and Disadvantages

Like any research design, Delphi provides benefit and value when it is determined to be the most suitable approach to address the research problem and answer the research question. Yang et al.(2012) noted its suitability for studies that exhibited the following properties:

- Subjective expertise and judgmental inputs;
- Complex, large, multidisciplinary problems with considerable uncertainties;
- Possibility of unexpected breakthroughs;
- Causal models cannot be built or validated;
- Particularly long time frames; and
- Opinions required from a large group, Anonymity is deemed beneficial. (p. 78)

Following are some specific examples of advantages and disadvantages of using Delphi:

Advantages

The main advantage of the Delphi Method comes in achieving consensus in areas of uncertainty or in situations lacking in causation (Powell, 2003). This is particularly true in the case of studies that focus on topics where multiple stakeholder groups are potentially involved. The Gjologaj (2015) and Rivera (2013) dissertations provide typical examples, as does the Joyner and Smith (2015) study. In no particular order of importance, some other advantages include the following.

Flexibility and simplicity

Delphi studies can be relatively straightforward to design and flexible in how those designs are put together. The Gjologaj panel (2015) addressed a single issue (competencies) and asked the panel one question. Responses were recycled until consensus achieved. The Kim and Aktan panel (2014) asked three questions and combined data gathering (qualitative) and assessment (quantitative) components until the purpose was satisfied. In pursuing a Delphi design, the researcher could start the process by presenting initial responses to the panel question gleaned from the literature (as was done in the Joyner and Smith modified design panel) or simply ask the question of panel members (as Gjologaj did). Researchers might even employ a validated quantitative instrument to begin the process. It's critical, however, that a researcher not compromise rigor for simplicity.

Knowledge sharing

One area not often mentioned in regard to Delphi designs centers on the ability of different disciplines to share knowledge and stimulate new ideas that apply to the purpose of the panel (Pill, 1971) and which can broaden the knowledge base of other panel members. In the hypothetical admissions criteria discussion, having faculty involved in the panel can provide insight (and potentially correct misconceptions among other participant groups) which could broaden the perspective of the other two panel groups.

Cost effectiveness

Unless the researcher decides to integrate an instrument for which there is a "fee for use" involved, pursuit of a Delphi design is highly cost-effective (Williams & Webb, 1994). It is the researcher's time, as well as the time of participants (all of whom are volunteers), that constitute the principal cost. Beyond time, there is little expense unique to the study. It makes use of tools and supplies already available.

Freedom of expression

Freedom of expression is a direct result of the anonymity required in all Delphi panels. It gives panel members considerable latitude in presenting their opinions and potentially offers different perspectives to others without fear of criticism (de Villiers, de Villiers, & Kent, 2005). Anonymity also eliminates participants feeling threatened in reporting their opinions and positions. Not knowing who said what eliminates overt judging of individual contributions by other panel members.

Ease of communications

In its infancy, the Delphi Method employed physical artifacts (typed and mailed/faxed documents) that severely limited the efficiency of panel operations and essentially eliminated the opportunity for international participation. The availability and ease of electronic communications have gone a long way to ameliorate those issues, but in turn have created issues of their own, especially for researchers. This is particularly true in maintaining participant anonymity. With the exception of disseminating individual round questions and subsequent summaries of round results, all communications between the researcher and each panel member should be a one-to-one basis.

Membership variations

As experienced in the Rivera (2013) dissertation and the Wynaden et al. (2014) study, the number of panel members participating in each round does not have to remain constant throughout the entire process. It can vary in a number of ways: members can drop out or skip a round and return later, or members can skip the initial round but join a subsequent one, to name just two. Researchers need to be attentive to maintaining the number of groups and the balance in membership cross the groups and to ensuring criteria for membership are maintained.

Lack of geographical limitation

Electronic communications have virtually eliminated geographical boundaries that affected the design of physical artifact Delphi studies. The Gjologaj panel (2015) was entirely located in Albania, but the researcher was in Brooklyn, New York. Kim and Aktan (2014) were located in Korea and Turkey, respectively. Researchers are no longer bounded by geography in recruiting participants or in addressing issues of international concern.

Disadvantages

Like most research designs, Delphi is not without flaw, but also like most research designs, those flaws arise with the researcher and not the design. Flaws can appear from shortcomings of the researcher or from panel members. Following are some typical examples.

Researcher bias

Given the extraordinary authority and influence of the researcher in the process, bias may creep into the process, even unintentionally (Linstone & Turoff, 2002). How the question(s) are formulated and who is invited to participate can become tools for the researcher's positions to prevail. Just as interview scripts in dissertation proposals are field tested by experts to ensure they don't "steer" responses, it would be in the best interest of a Delphi researcher to have an outside expert review the formulation of the question(s). That outside expert could be the dissertation committee Chair or another faculty member familiar with the Delphi design.

The temptation for a researcher to select panel members with known positions on the problem is another way a researcher can exert bias. While the policy of "no known relationship" (even casual

acquaintance) espoused by Murphy et al. (1998) would seem to preclude selection of individuals who share the researcher's positions, this may not always be possible in areas where the field of experts is limited. Researchers can also exert bias in panel member selection by appointing themselves as arbiters of participant qualifications. As noted earlier, it should be the discipline involved that determines expertise, and not the researcher.

Researcher shortcomings

A potential disadvantage lies in the researcher imposing his or her preconceptions on respondents, especially in the case of a modified design that uses a researcher generated literature review for Round 1. Another can be exhibited by poor summarizing of panel contributions or incomplete presentation of the group response for the next round (de Villiers et al., 2005). The expertise of individual panel members in their subject areas and their willingness to be open and direct in their contributions should provide an offset to researcher shortcomings. Previous discussion of modified Delphi designs noted that no matter how initiating material was generated, panel members should never be constrained from adding to the alternatives generated. It's important for researchers to recognize that their role is not one of contributor, but rather than of facilitator.

Panel member anonymity and petulance

Vernon (2009) reported an occasion where participants expressed minority opinions but dropped out when asked to explain them (remaining members had to deal with them anyway). McKenna (1994) noted that anonymity could provide temptation for participants to be less than fully motivated and be less rigorous or less serious in their contributions. Also problematic is the possibility that a panel member can drop out at any point and for any reason and not just the ones already mentioned.

It's the topic, not the method

Grant and Kinney (1992) suggested that panel deliberations could get bogged down in discussions/debates over the method as opposed to the topic. Vernon (2009) reported that panel members with expertise in research methods who espouse a positivist perspective may challenge the method as opposed to contributing topic-related positions.

Sackman (1975) indicated that the consensus process does not lead to the "best" option, but rather to a "watered down" version. This was later echoed by Rennie (1981) who asserted that the process yielded only non-controversial statements that would produce the least common denominator. And while these might be valid concerns, one might reasonably ask, however, what defines "best"?

If an education dissertation proposal asked the research question and convened a Delphi panel to pursue "how best to integrate social/emotional learning into elementary school curricula," what defines best could conceivably vary depending on who was asked. Teachers might identify one process as "best," but the school psychologist might well say another was "best." And if either required parental involvement, that group might disagree with either or both. In the end, what defines "best" is the one that meets the objective and instills the learning into the students, as agreed by all. That's the fundamental benefit of a Delphi panel approach.

Other Factors to Consider in Delphi Design

To this point, many of the major criteria for designing a Delphi study have been addressed:

- Panel size (number and membership of participating groups);
- Panel member recruiting (definition of “expert” and participant time, interest, and language facility);
- Choice of design approach (conventional or modified); and
- Process to maintain participant confidentiality and provide feedback to the panel.

In addition to these, three additional factors are put forth for consideration.

Field Tests or Pilot Studies

Field tests and pilot studies are not techniques normally associated with Delphi designs, but there are cases where employing either (or both) would be to the researcher’s advantage. One example where a field test might be appropriate occurs in Round 1 if the researcher plans an interview of panel members or when a researcher-designed survey is planned. Ensuring that the interview script or survey is comprehensive and thorough in addressing the study topic might best be achieved by submitting the instrument to a field test before implementation.

In the case of complex studies involving large panels that include multiple professional groups each with a widely diverse membership, it may be to the researcher’s advantage to “test” the process in the form of a pilot study. A pilot test might also be advantageous in the case where the author plans a self-designed survey instrument. Proving statistical validity and reliability is not necessarily the objective in these pilot studies, but rather how well the design would function when executed.

Use of Outside Help

When looking to include a Round 1 interview script or survey, researchers should not be reluctant to seek outside expertise in reviewing any planned instruments in a modified study design. Unlike a traditional field test, which focuses on the credibility and transferability of an instrument and employs individuals with expertise in the specific research method involved, field tests of Delphi instruments seek to determine if the instrument is comprehensive and thorough in the topical area and instead might employ one or more subject matter experts.

Data Saturation

When pursuing an interview-based research design, such as narrative inquiry or phenomenology, attention to achieving data saturation is a key component in the process of answering the research question and achieving the study purpose. In those situations where a researcher plans to conduct interviews of panel members or outsiders to initiate Delphi panel deliberations, achieving saturation is not a concern. The purpose of the interviews is to provide a starting point for panel deliberations, and it is those panel deliberations that will answer the research question and achieve the study purpose. Also contributing to overlooking saturation is the opportunity for panel members to contribute beyond any list of alternatives provided to them.

Delphi Design in Recent Dissertations

Some interesting statistics can be extracted from the dissertation abstracts that appeared in the 78,104 dissertations published by UMI Dissertation Services in the two years between January 1, 2014, and December 31, 2015. A search for keywords in the abstracts reflecting the various re-

search designs revealed the following: 9,486 classified the design as case study, 1,266 as grounded theory, 288 as ethnography, 400 as phenomenology, but only 200 described the study's design as "Delphi."

Based on the above statistics, one might legitimately ask why the design hasn't achieved greater dissertation research exposure. No doubt there could be a number of reasons why it hasn't enjoyed greater popularity. Doctoral candidate comfort level and committee Chair expertise in one design area over others is perhaps one. Lack of exposure to the design in Research Methods courses is certainly another. But still another reason may lie in the availability of design resources used by both new and experienced researchers. Discussion of the Delphi design is typically found only in databases that provide links to articles that reported studies using the method, or in the doctoral dissertations already completed (both using "Delphi" as one of the search terms). Unless one can assess Delphi as an alternative to generic inquiry or single/multiple case study in investigating and answering a research question, for example, it can get overlooked as a viable research alternative. Well-known texts in qualitative research methods and designs such as Creswell (2014), Christensen, Johnson, and Turner (2013), or Leedy and Omrod (2013) don't mention it at all. Another possible reason centers on a perception the Delphi is a technique only for forecasting future events or for problem solving that addresses a practical concern or shortcoming. But isn't a practical concern or shortcoming the starting point for a significant number of contemporary dissertations?

Perhaps the cause of the previously listed issues (not covered in Research Methods courses or methodology texts) has evolved from what had been the harshest criticism of Delphi that has appeared over the years and might still permeate some segments of the research methods community. That criticism came in 1975 when Sackman highlighted it as having limited scientific value. Linstone and Turoff (2002) took a contrary position, pointing out that Sackman needed to look beyond methodological tradition. More recently, the work of Yang, Zeng, and Zhang (2012) specifically noted the applicability of Delphi for complex studies that called for subjective input where no cause and effect could be established. This category of studies would call for a process outside the realm of traditional science. It would appear that Delphi shares the same scientific value as generic qualitative research in the form of narrative inquiry, which regularly collects subjective input when no cause and effect has been established. And since dissertation committees and peer-review panels regularly accept dissertations and articles using Delphi, its place within the realm of research methods should be acknowledged, and more importantly, utilized.

Summary

Despite being the least known and utilized among dissertation research designs, the Delphi Panel has been experiencing increasing acceptance recently. It is uniquely applicable to a number of research applications that involve a postpositive/constructivist perspective, and it offers potential advantages not available in many of the more traditional qualitative designs such as case study, ethnography, or grounded theory. It comes with disadvantages as well, but most of these can be overcome with careful attention to the design of the process. Delphi designs are flexible in that they can be organized in different configurations, and they are particularly applicable in pursuing complex topics where there is no clear causal relationship and the subjective judgments of experts could generate breakthroughs.

Delphi panels are based on two principles: the opinion of experts is a mixture of knowledge and speculation, and the averaging of separately collected opinions provides a more accurate picture than a collective opinion resulting from a face-to-face group discussion. Additional facets that provide the foundation of the method are concentrated on the researcher, who performs the functions of planner and facilitator, but not instrument as found in other research designs. Key proper-

ties of the design are participant anonymity (to encourage freedom of expression) and feedback (panel discussion are summarized and fed back to members).

In designing the panel, the researcher must decide which groups of experts will best arrive at satisfying the study purpose, how many individuals should be included in each group, and what criteria will be used for membership. In the case of membership criteria, it is expertise as determined by the discipline of the group and not the opinion of the researcher that's the primary professional consideration. This, taken together with the selection of individuals who are unknown to the researcher, will provide the best approaches to staffing the panel. Willingness of potential participants to devote the time and interest in the specific topic are also considerations, as is the ability to communicate, particularly in the case of international panels.

Once the panel is formed, members are asked specific questions and provide responses which the researcher consolidates and feeds back to the panel in a series of "rounds" until consensus is achieved. It's important to note that unanimity is not the objective; achieving 70% concurrence among panel members is considered the standard (Vernon, 2009).

Any research design has its advantages and disadvantages, both in selection and use. Delphi designs offer distinct advantages in dealing with topics where problem solving is a desired outcome or when causation cannot be established. Problems in business and education typically affect different constituencies and having those constituencies work collectively to find what works best for all together is an ideal dissertation research design, as is trying to determine "why" something has occurred when there is no clear cut cause precipitating it.

A key consideration to the applicability of a Delphi design centers on the topic affecting multiple constituencies. While Delphi designs have been employed in single constituency studies (see earlier discussion of the 2014 Wynaden et al. study), in most cases research questions that focus on a single constituency provide little beyond a voting process on the various responses. The relative advantage of employing a Delphi design over a more direct integration of the multiple perspectives that individual members of a single constituency would provide in response to a series of questions delivered in a narrative inquiry format is something that would flow directly from the study purpose.

Reference List

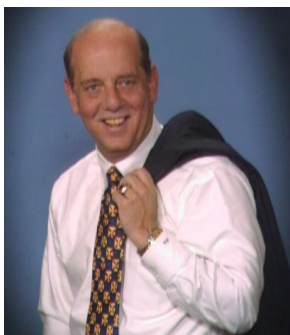
- Akins, R., Tolson, H., & Cole, B. (2005). Stability of response characteristics of a Delphi panel: Application of bootstrap data expansion. *BMC Medical Research Methodology*, 5(37). doi:10.1186/1471-2288-5-37
- Amant, R., (1970). Comparison of Delphi forecasting studies in 1964 and 1969. *Futures*, 2(1), 35-44. Retrieved from <http://www.sciencedirect.com/science/article/pii/0016328770900066>
- Christensen, L., Johnson, R., & Turner, L. (2013). *Research methods, design, and analysis* (12th ed.). Upper Saddle River, NJ: Pearson Education.
- Creswell, J. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage Publications.
- Custer, R., Scarella, J., & Stewart, B. (1999). The modified Delphi technique - A rotational modification. *Journal of Career and Technical Education*, 15(2). Retrieved from: <https://ejournals.lib.vt.edu/JCTE/article/view/702/1013>
- Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, 9(3), 458-457. Retrieved from <http://pubsonline.informs.org/doi/abs/10.1287/mnsc.9.3.458>

Delphi Panels

- deBoer, J & Hale, J. (2002, November/December). The oracle at Delphi: Was she really stoned? *Archeology Odyssey*, 6/2002. Retrieved from <http://www.biblicalarchaeology.org/daily/ancient-cultures/daily-life-and-practice/the-oracle-of-delphi%E2%80%94was-she-really-stoned/>
- de Villiers, M., de Villiers, P., & Kent, A. (2005). The Delphi technique in health sciences education research. *Medical Teacher*, 27(7), 639-643. doi: 10.1080/13611260500069947
- Fischer, R. (1978). The Delphi method: A description, review, and criticism. *Journal of Academic Librarianship*, 4(2), 64-70. Retrieved from <http://www.journals.elsevier.com/the-journal-of-academic-librarianship/>
- Gjologaj, V. (2014). *Developing a sports club management competency model for Albania: A Delphi study*. (Doctoral Dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3621740).
- Grant, J., & Kinney, M. (1992). Using the Delphi technique to examine the content validity of nursing diagnoses. *International Journal of Nursing Terminologies and Classifications*, 3(1), 12–22. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1744-618X.1992.tb00193.x/abstract>
- Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*, 32(4), 1008-1015. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1046/j.1365-2648.2000.t01-1-01567.x/full>
- Hejblum, G., Ioos, V., Vibert, J., Bøelle, P., Chalumeau-Lemoine, C., Chouaid, C., Valleron, A., & Guidet, B. (2008). A web-based Delphi study on the indications of chest radiographs for patients in ICUs. *Chest Journal*, 133(5), 1107–1112. doi:10.1378/chest.06-3014
- Hsu, C., & Sandford, B. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment Research & Evaluation*, 12(10), 1-8. Retrieved from <http://pareonline.net/pdf/v12n10.pdf>
- Janis, I. (1982). *Groupthink: Psychological studies of policy decisions and fiascos* (2nd ed). Boston: Houghton-Mifflin.
- Joyner, H., & Smith, D. (2015). Using Delphi surveying techniques to gather input from non-academics for development of a modern dairy manufacturing curriculum. *Journal of Food Science Education*, 14, 88-117. doi:10.1111/1541-4329.12062
- Kim, M., & Aktan, T. (2014). How to enlarge the scope of the curriculum integration of mathematics and science (CIMAS): A Delphi study. *Eurasia Journal of Mathematics, Science, & Technology Education*, 10(5), 455-469. doi:10.12973/eurasia.2014.1115a
- Leedy, P., & Omrod, J. (2013). *Practical research: Planning and design* (10th ed.). Upper Saddle River, NJ: Pearson Education.
- Linstone, H. A., & Turoff, M. (Eds.). (2002). *The Delphi method: Techniques and applications*. Retrieved from <http://is.njit.edu/pubs/delphibook/delphibook.pdf>
- McBride, M. (2015). *A modified Delphi approach to posttraumatic stress disorder and leadership theory*. (Doctoral Dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3692072).
- McKenna, H. (1994). The Delphi technique: A worthwhile research approach for nursing? *Journal of Advanced Nursing*, 19(6), 1221-1225. doi:10.1111/j.1365-2648.1994.tb01207.x
- Melynk, S., Lummus, R., Vokurka, R., Burns, L., & Sandor, J. (2009). Mapping the future of supply chain management: A Delphi study. *International Journal of Production Research*, 47(16), 4629-4653. doi:10.1080/00207540802014700.
- Murphy, M., Black, N., Lamping, D., McKee, C., Sanderson, C., Askham, J., & Marteau, T. (1998). Consensus development methods, and their use in clinical guideline development. *Health Technology Assessment*, 2(3), 1-88. doi:10.3310/hta2030

- Pill, J. (1971). The Delphi method: Substance, context, a critique and an annotated bibliography. *Socio-Economic Planning and Science*, 5(1), 57–71. Retrieved from <http://www.sciencedirect.com/science/article/pii/0038012171900413>
- Powell, C. (2003). The Delphi technique: Myths and realities. *Journal of Advanced Nursing*, 41(4), 376–382. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1046/j.1365-2648.2003.02537.x/full>
- Rennie, D. (1981). Consensus statements. *New England Journal of Medicine*, 304(11), 665–666. Retrieved from <http://www.nejm.org/doi/pdf/10.1056/NEJM198103123041110>
- Rivera, R., Jr. (2013). *Assessing community reintegration in adolescents and young adults with spinal cord injury: A Delphi study*. (Doctoral Dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3574181).
- Rupprecht, M., Birner, K., Gruber, H., & Mulder, R. (2011). Dealing with diversity in consulting teams: Results of two Delphi studies. *Human Resource Development International*, 14(5), 561-581. doi: dx.doi.org/10.1080/13678868.2011.618348
- Sackman, H. (1975). *Delphi critique*. Lanham, MD. Lexington Books.
- Seidman, I. (2006). *Interviewing as qualitative research: A guide for researchers in education and the social sciences* (3rd ed.). New York, NY: Teachers College Press.
- Vernon, W. (2009). A Delphi technique: A review. *International Journal of Therapy and Rehabilitation*, 16(2), 69-76. Retrieved from doi: <dx.doi.org/10.12968/ijtr.2009.16.2.38892>
- Von der Gracht, H. (2008). The Delphi technique for futures research. In H. Von der Gracht, *The Future of Logistics: Scenarios for 2025*, pp. 21-68. doi: 10.1007/978-3-8349-9764-7_3
- Williams, P., & Webb, C. (1994). The Delphi technique: A methodological discussion. *Journal of Advanced Nursing* 19, 180–186. doi: 10.1111/1365-2648.ep8540360
- Wynaden, D., Heslop, K., Omari, O., Nelson, D., Osmond, B., Taylor, M., & Gee, T. (2014). Identifying mental health nursing research priorities: A Delphi study. *Contemporary Nurse*, 47(1-2), 16-26. ISSN: 1037-6178 PMID: 25267123
- Yang, X., Zeng, L., & Zhang, R. (February 2012). Cloud Delphi method. *International Journal of Uncertainty, Fuzziness & Knowledge-Based Systems*, 20/1, 77-97. doi:10.1142/S0218488512500055
- Yousef, M. I. (2007). Using experts' opinions through Delphi technique. *Practical Assessment Research & Evaluation*, 12(4). Retrieved from <http://pareonline.net/getvn.asp?v=12&n=4>

Biography



Dr. Jay Avella holds a BS in Chemistry from Rensselaer Polytechnic Institute in Troy, NY; a MA in International Affairs from Catholic University in Washington, DC; a MBA from Capella University; and the Ph.D., also from Catholic University in International Affairs. He recently completed chairing his 80th dissertation and has been appointed as a dissertation methods quality reviewer at a prominent US education institution in addition to teaching research methods courses.