

## **LA TÉCNICA DELPHI COMO ESTRATEGIA DE CONSULTA A LOS IMPLICADOS EN LA EVALUACIÓN DE PROGRAMAS**

*Esther Martínez Piñeiro*

Facultad de Ciencias de la Educación  
Universidad de Santiago

### **RESUMEN**

*En este artículo presentamos la técnica Delphi como una interesante estrategia de consulta a los implicados en la evaluación de programas sociales.*

*Tras exponer sus principales características y algunas consideraciones sobre la configuración del panel de expertos y el proceso de aplicación, revisamos los dos criterios empleados en la selección de los items y la finalización del proceso de consulta, el consenso y la estabilidad. Analizamos así los principales estadísticos utilizados a la hora de concretar dichos criterios, tanto los más habituales, basados en la mediana y la media o en pruebas de bondad de ajuste, como otros menos extendidos en este tipo de investigaciones.*

*Finalizamos con algunas consideraciones metodológicas sobre el uso de esta estrategia a la luz de los problemas surgidos en el estudio Delphi realizado en torno a la evaluación de los cursos de formación ocupacional de la comunidad gallega.*

***Palabras clave:** Técnica Delphi, panel de expertos, medidas de consenso, medidas de estabilidad, evaluación de programas, participación de audiencias.*

### **ABSTRACT**

*This paper introduces the Delphi Technique as an interesting strategy for taking up with those people who involved in social programs evaluation.*

*To begin with, the technique main features and some considerations on the experts panel configuration and on the application process are presented. Next, consensus and stability, the two criteria for items selection and for consultation process finalisation, are revised. Thus, the*

*key statistics used to fix the aforementioned criteria are analysed; both the more usual, based upon the mean and the median or upon goodness of fit tests, and some others less commonly applied to this kind of researches.*

*Finally some methodological considerations about the utilisation of this strategy are introduced in the light of some problems arisen from the Delphi Study carried out in the realm of occupational training courses evaluation in Galicia.*

**Key Words:** *Delphi Technique, experts panel, consensus measurements, stability measurements, program evaluation, stakeholder groups participation.*

## **I. UTILIDAD DE LA EVALUACIÓN: EL PAPEL DE LAS AUDIENCIAS**

Durante las últimas décadas hemos asistido a la paulatina democratización de las instituciones en todos los ámbitos sociales, y así también en la evaluación de las políticas públicas. En esta línea, se han ido articulando sistemas de evaluación con los que las administraciones rinden cuentas a la sociedad sobre el buen uso de los recursos públicos. A pesar de este avance, la presencia de los colectivos cercanos al servicio o programa analizado se limita habitualmente a colaborar como principales informantes en el estudio y a ser conocedores de los resultados. Compartimos, sin embargo, la idea de que el control de la ciudadanía sobre tales políticas no será real si ésta no participa activamente en estos procesos de evaluación.

La colaboración de los grupos o colectivos afectados en la evaluación de los programas públicos no es sólo un derecho social, donde el debate abierto entre los distintos grupos propicia la reflexión y el aprendizaje colectivo, sino que también posibilita, desde el punto de vista metodológico, mejorar la utilidad de la evaluación, criterio fundamental de calidad de los estudios evaluativos. Dicha colaboración mejora el proceso evaluativo en tres direcciones principales:

En primer lugar, centra el estudio evaluativo en aquellas cuestiones de interés no sólo para los responsables del programa, principales clientes, sino también para otros colectivos afectados. De esta manera la evaluación dará respuesta a sus necesidades priorizando y seleccionando la información relevante. En segundo lugar, genera una mayor credibilidad del equipo evaluador, acompañada de una mayor transparencia del proceso y de una mayor confianza en los resultados. En tercer lugar, posibilita una mejor comprensión del proceso que es así construido por los participantes, o cuando menos seguido a menor distancia. En definitiva, dicha participación puede conllevar un mayor impacto social de la evaluación, que se concreta, principalmente, en la puesta en marcha del proceso posterior de toma de decisiones fundamentado con consecuencias reales en el programa y su contexto.

En la actualidad la consideración de criterios democráticos en la evaluación de programas es una tendencia que cobra fuerza por encima de otros enfoques (De Miguel, 2000). En cualquier caso, la participación de los colectivos implicados en la evaluación, tanto en la proposición de criterios como de posibles contenidos, requiere establecer mecanismos que permitan «escucharlos». En este sentido parece deseable en muchos programas socioeducativos poder llegar a establecer un cierto nivel de consenso entre

las distintas audiencias, las que mantienen con frecuencia intereses y puntos de vista muy distintos. El método Delphi se presenta como una interesante estrategia de recogida y análisis de información en esta línea.

## **2. LA TÉCNICA DELPHI**

La técnica Delphi constituye una estrategia de análisis de opinión que puede ser entendida como un tipo específico de entrevista en grupo aunque con características particulares (Del Rincón et al., 1995; Ruiz Olabuénaga y Ispuzua, 1989). Esta puede describirse como un proceso de comunicación entre varios sujetos, mediante la contestación de sucesivos cuestionarios a través de los que se llega a un consenso en torno una cuestión sin que exista la interacción personal entre ellos (Uhl, 1991). El objetivo central de esta técnica es, por lo tanto, consensuar la opinión de un grupo de personas en torno a una determinada problemática, cuyas soluciones o posible situación futura es difícil de predecir.

Delphi se sustenta en base a dos razones: en primer lugar, la mayor adecuación de las técnicas de recogida de opinión grupales frente a las individuales en estas situaciones, ya que varias personas proporcionan mejor información que una sola. En este sentido es importante conocer la opinión de varios expertos sobre una posible problemática, entendiéndose que el consenso entre ellos dará lugar a una visión que ofrezca mayor confianza. En segundo lugar, esta técnica supera ciertas limitaciones de aquellas donde existe interacción cara a cara entre los participantes, como la influencia de sujetos dominantes, el ruido semántico y la presión del grupo hacia la conformidad.

### **2.1. Características**

Son tres las características más importantes que se le otorgan (Dalkey et al., 1972): el anonimato, la retroalimentación controlada y el consenso estadístico del grupo. El anonimato viene dado porque los miembros del panel no conocen la identidad de los otros participantes. Esto les permite expresar su opinión y cambiarla sin miedo a ser amonestado por otro miembro; además la influencia personal está anulada. La retroalimentación controlada es posible porque la secuencia de cuestionarios en los que se comunica a los participantes los resultados de la vuelta anterior evitan el ruido semántico, ya que la discusión puede ser reorientada y pueden evitarse sesgos que distorsionen la marcha de la investigación. Por último, el consenso estadístico se consigue porque cada miembro recibe información del grado de acuerdo en que se encuentra con respecto al grupo, pero gracias al anonimato no sabe con respecto a qué sujeto concreto está en distensión, esto ayuda a proteger los sesgos producidos por el pensamiento del grupo, donde la opinión de una persona o minoría prestigiosa puede dictar las decisiones de la mayoría.

### **2.2. Proceso de aplicación**

Esta técnica consiste en la aplicación de sucesivos cuestionarios con el objetivo de lograr el consenso entre el conjunto de participantes. Comienza el proceso con la deli-

mitación del problema y la selección de los sujetos que serán consultados, denominado panel de expertos, los que reciben un primer cuestionario generalmente de preguntas abiertas. El investigador elabora a partir de las respuestas recibidas el segundo cuestionario, muchas veces ya de respuestas cerradas que posibilitan un tratamiento estadístico de las respuestas del grupo. Analizadas éstas se remite a cada participante el tercer cuestionario en el que se ofrece información sobre la puntuación promedio dada por el grupo y su puntuación anterior, con el objeto de que, si el participante lo considera, modifique su respuesta; normalmente se acompaña de un espacio para justificar los motivos que le llevan a mantener su posición si así lo hace. La remisión de cuestionarios finaliza cuando se obtiene un alto consenso entre los participantes en todas las cuestiones, o bien cuando se considera que el consenso no aumentará más en posteriores consultas (Fernández-Ballesteros, 1995). No existe, por lo tanto, un número de consultas fijo, ya que este depende del grado de consenso que se alcance en las sucesivas vueltas y también del punto de partida del proceso. Si bien distintos autores señalan que el primer cuestionario está compuesto por preguntas abiertas (Fernández-Ballesteros, 1995 y Ruiz Olabuénaga y Ispizua, 1989) otros (Pérez-Campanero, 1991) indican que en el caso de estar previamente definido el problema en términos precisos se suprimen los primeros cuestionarios y comienza el proceso con los cuestionarios diseñados con preguntas cerradas, lo que acostumbra a seguir muchas de las investigaciones que emplean esta técnica con el objetivo de abreviar y simplificar al proceso de recogida de datos.

### 2.3. Instrumentos de recogida de datos

Es habitual comenzar el proceso utilizando un cuestionario de preguntas abiertas que permitan recabar información amplia y rica de todo el conjunto del panel evitando la posible omisión de aspectos de interés. En este caso deberán ser preguntas concretas y claramente formuladas que conduzcan a la transformación de dichas respuestas en nuevas preguntas cerradas. La técnica Delphi permite utilizar en esta segunda fase distintos tipos de preguntas, los más frecuentes son (Landeta, 1999):

- Jerarquizaciones: deben clasificarse un conjunto de elementos del primero al último atendiendo a los criterios especificados en la pregunta. Generalmente recibe el valor menor, el uno, el elemento prioritario, obteniendo el valor mayor aquel situado al final de la clasificación.
- Valoraciones: preguntas generalmente con respuesta en escala graduada, tanto verbal como numérica, en la que el participante asigna o marca la puntuación al elemento considerado. Este es el tipo de pregunta más utilizada, ya que resulta más rápida la contestación del cuestionario y se facilita el análisis de los resultados. También es frecuente en estudios que tienen como objetivo final priorizar o seleccionar elementos del conjunto presentado a los panelistas.
- Comparaciones: dados los elementos de análisis de dos en dos, el miembro del panel debe indicar cual de los dos es preferente atendiendo a los criterios especificados. Cada elemento es así contrastado con los demás, asignándole un uno

si es el marcado de los dos, y un cero en el caso contrario. La suma de los puntos de cada elemento servirá para otorgarle una puntuación por cada sujeto del panel.

- Elecciones: el sujeto participante debe escoger de entre un conjunto de elementos el que considera más adecuado según los criterios fijados.
- Estimaciones cuantitativas: el panelista ofrece un valor numérico estimado puntual o, en ocasiones, un intervalo.

Es el equipo de investigación el que debe considerar las ventajas e inconvenientes de cada uno teniendo en cuenta los objetivos del estudio, la composición de panel de expertos, la posible longitud del instrumento, etc. En ocasiones, atendiendo al tema de estudio se recurre al empleo de distintos tipos de preguntas en el mismo instrumento, si bien esto conlleva un mayor esfuerzo de análisis.

#### **2.4. El panel de expertos: elección de los sujetos y tamaño**

La técnica Delphi pretende consensuar la opinión en torno a un tema de un grupo de sujetos considerados «expertos» en él. Inicialmente, se entendía como tales a aquellas personas con un amplio conocimiento y experiencia en la materia, cuyas predicciones podían ser así bastante acertadas. En la actualidad, y debido en parte a la utilización de esta técnica en contextos muy diversos en los que se precisa conocer el punto de vista de otros colectivos, como es el caso de la evaluación de programas, este término amplió sus límites, y el panel de participantes en el estudio puede acoger a tres tipos de experto (Landeta, 1999):

- Los especialistas: o aquellos con conocimientos, experiencia y capacidad predictiva, que se identifican con la idea clásica de experto en el contexto de esta técnica.
- Los afectados: es decir aquellos que sin tener una formación o experiencia relacionada con el tema en cuestión están implicados en él.
- Los facilitadores: sujetos con capacidad de estimular al grupo, de proponer nuevos puntos de vista, etc.

Los objetivos y el contexto del estudio llevarán a la selección de un panel con mayor o menor representación de cada uno de estos tipos de experto. En todo caso, la inclusión de más de un tipo conlleva una configuración del panel en grupos y subgrupos que deberán tenerse en cuenta en las decisiones que se tomen sobre los criterios y pautas de análisis de datos.

Es frecuente en los estudios que utilizan esta técnica la participación de dos paneles de configuración distinta. El primero de ellos está formado exclusivamente por especialistas en la materia, con el objetivo de consensuar entre ellos las líneas y contenidos relativos la problema que se está analizando; sus respuestas servirán así para la construcción del instrumento que será contestado por el segundo panel. Este lo integran sujetos relacionados con el tema del estudio, pero no son necesariamente especialistas.

Debe considerarse también en la selección de los participantes criterios tales como la motivación real del sujeto a participar —ya que en caso contrario existe un elevado riesgo al abandono antes de finalizado el proceso—, el conocimiento sobre el tema, el coste del proceso, las dificultades organizativas, etc.

Además debe decidirse el número de personas que formarán dicho panel. Ruiz Olabuénaga e Ispizua (1989) señalan entre los diez y los treinta componentes. Otros autores (Landeta, 1999) apuntan que no es recomendable, por cuestiones de operatividad, que supere los cincuenta. Entendemos que el número de miembros del panel depende de los objetivos y las condiciones en que se desenvuelve cada estudio, debiendo ser lo suficientemente amplio para permitir la diversidad de opiniones (Fernández-Ballesteros, 1995) pero haciendo factible el propio proceso, por lo que no existe un número idóneo de participantes; este ha de adecuarse a cada situación específica de investigación. En todo caso, los sujetos seleccionados para participar en este proceso no constituyen una muestra de una población determinada, sino que son personas que representan distintas posiciones sobre el problema formulado.

## 2.5. Análisis de los resultados

La aplicación de esta técnica requiere, necesariamente, que deban ser analizados los resultados parciales a medida que se obtienen —es decir los procedentes de cada una de las consultas— por cuatro razones básicas:

- El contenido del cuestionario empleado en cada ronda depende de los resultados de la fase anterior. Esto ocurre tanto si nos encontramos tras un cuestionario de preguntas abiertas, propio de la primera consulta, donde el análisis de las respuestas —su clasificación y categorización— permitirá diseñar las preguntas del siguiente cuestionario, como si estamos ante cuestionarios de preguntas cerradas; en este caso es posible la supresión de elementos del cuestionario cuando se hayan alcanzado los niveles esperados de acuerdo o estabilidad en determinadas preguntas, las cuales no serán incluidas en rondas sucesivas.
- Tras el primer cuestionario de preguntas cerradas los siguientes ofrecen al panelista información estadística sobre los resultados del grupo y su posición respecto a él, con el objetivo de que este valore de nuevo su respuesta y la modifique si lo considera oportuno.
- En el marco de determinados estudios que utilizan esta técnica se contempla la inclusión de nuevas preguntas, fruto de las aportaciones de los miembros del panel, que deberán ser revisadas, formuladas e incluidas en el instrumento siguiente.
- Los resultados de cada ronda condicionan la finalización del proceso. El equipo de investigación deberá considerar su conclusión a la luz de dichos resultados, sin olvidar otros condicionantes, principalmente la mortalidad de la muestra en cada ronda, pero también la propia disponibilidad de recursos y de tiempo y la rentabilidad real de realizar una nueva consulta después de haber conseguido un determinado nivel de consenso entre los participantes.

El análisis de los datos debe ir, por lo tanto, precedido de la determinación de los criterios que rigen la selección de los elementos del cuestionario sobre los que se seguirá o no realizando la consulta al grupo, así como, si procede, de los que conducen a la finalización del proceso sin haber alcanzado la selección de todos los elementos. Son dos los criterios que se utilizan, el consenso entre los miembros del panel y la estabilidad de sus respuestas.

- El consenso: responde a la propia filosofía de la técnica, ya que su principal objetivo es, precisamente, la convergencia entre las opiniones de los participantes. Para conocerla es posible utilizar los siguientes estadísticos:
  - Rangos basados en los cuartiles: el rango semiintercuartílico, es decir la distancia media entre el primer y el tercer cuartil; el rango intercuartílico, cuando se toman la diferencia entre ambos, o el rango intercuartílico reducido (RIR), esto es, la diferencia entre el tercer y el primer cuartil dividido entre la mediana. Estos tres rangos poseen la resistencia de la mediana (Palmer Pol, 1999), ya que se basan en la distancia existente entre percentiles; en estos casos, los valores no están condicionados por la existencia de puntuaciones extremas.
  - El coeficiente de variación: es decir la desviación típica de las respuestas dadas a la pregunta dividida entre la media. Este coeficiente, a pesar de ser utilizado en este tipo de estudios, presenta algunos inconvenientes. En primer lugar, no es aconsejable con distribuciones asimétricas, en la que existan valores extremos que condicionen el valor de la media. En segundo lugar, distribuciones con el mismo consenso tendrán coeficientes distintos dependiendo del valor de la media en una de ellas; así pues, el coeficiente variará aun cuando el grado de acuerdo, es decir la dispersión de las respuestas, sea el mismo.
  - Intervalo de confianza de la media: media más menos un determinado valor de desviaciones típicas. Identifica así los sujetos extremos muy por encima o debajo de la media, que se sitúan fuera de un determinado porcentaje de casos. Cuando la puntuación dada por un sujeto a una pregunta está fuera de este rango se solicita que vuelva a considerar su respuesta. Este criterio implica el análisis de las respuestas de cada sujeto y la configuración del cuestionario de la ronda siguiente adaptado a cada uno de ellos, ya que estos reciben las preguntas en las que su respuesta está fuera de estos límites, con la referencia a la puntuación media del grupo. Puede utilizarse cuando las preguntas poseen formatos diferentes ya que toma como referencia la dispersión relativa de cada distribución. El uso de un intervalo de confianza respecto de la media presenta inconvenientes en distribuciones asimétricas, donde el valor de la media se sitúa cerca de los extremos de la escala de respuesta utilizada, ya que el porcentaje del área tomado inicialmente como referencia de la distribución normal sufre modificaciones.
  - Porcentaje en alguna o algunas de las categorías de respuesta. Este criterio es usado en el caso de las preguntas de tipo escalar, donde se pueden considerar dos categorías contiguas. En estos casos debe establecerse previamente valor del estadístico que es considerado aceptable, excluyendo la pregunta de la siguiente

consulta por alcanzar el nivel de consenso esperado. Este es un referente muy utilizado debido a la claridad que presenta para los propios participantes, tomando los porcentajes alcanzados en los últimos valores de la escala cuando el objetivo que se persigue es seleccionar los elementos que consensuadamente son considerados más importantes.

- Cociente entre la desviación típica y la desviación típica uniforme: si la desviación típica de la variable es menor que un determinado porcentaje de la desviación típica uniforme de la distribución, el ítem es seleccionado. La desviación típica uniforme viene dada por el valor que tomaría la desviación típica cuando cada categoría recibe igual número de elecciones (Renom Pinsasch, 1992), es decir, en caso de encontrarnos con una distribución al azar. Este permite el uso de un valor teórico tomado como referente para todos los ítems que presenten el mismo formato. De esta forma su utilización evita los problemas derivados de distribuciones asimétricas o de aquellas con valores de la media diferentes.
- Pruebas de bondad de ajuste: permiten comparar las frecuencias de respuesta observadas en cada categoría con las frecuencias esperadas, que deben ser previamente decididas. Entre ellas destacamos:
  - Chi cuadrado: Compara las frecuencias observadas de cada categoría con las teóricas. En el caso de preguntas de respuesta dicotómica cada frecuencia esperada deberá ser por lo menos de cinco. En preguntas con más categorías no debe ser utilizada cuando en más del 20% de las categorías de respuestas tengan frecuencias esperadas menores que cinco (Siegel, 1975).
  - Prueba de Kolmogorov-Smirnov: en el caso de muestras pequeñas en las que no se contemple la condición de las frecuencias en cada categoría antes indicada o cuando sean preguntas de tipo escalar, es más poderosa esta prueba de bondad de ajuste (Siegel, 1975).

El problema de las pruebas de bondad de ajuste es definir previamente la distribución teórica con la que se establecerá la comparación, ya que esto, en definitiva, supone definir el consenso en términos de respuesta.

El criterio de consenso presenta el inconveniente de que la consulta continúa hasta haber alcanzado el nivel esperado por lo que puede conllevar a la repetición excesiva de aplicaciones, sobre todo cuando los miembros del panel hayan considerado sus respuestas en consultas anteriores y optado por no modificarlas. Este es uno de los inconvenientes que se señalan de esta técnica, la tendencia a forzar a los participantes a modificar su respuesta para alcanzar la convergencia de sus opiniones (Fazio, 1987).

- La estabilidad: este criterio se utiliza para decidir la finalización del proceso tras la comparación de las respuestas dadas en una ronda y la anterior. La existencia de estabilidad no implica el consenso del grupo, sino la escasa variación en las respuestas dadas por este. En este caso es posible la utilización de diversos estadísticos, en su mayor parte complementarios:
  - Variación en los coeficientes de dispersión: Dado el coeficiente de una ronda y la anterior, si la diferencia entre ambos es menor que un nivel establecido previa-

mente se considerara aceptable la estabilidad de las respuestas. En este caso podrán utilizarse los tres coeficientes anteriormente indicados, es decir podrá tomarse como referencia la variación del recorrido semiintercuartílico, intercuartílico, intercuartílico relativo, o en el coeficiente de variación, si bien siguen presentando los mismos inconvenientes que el correspondiente coeficiente.

- T de student para muestras relacionadas: contraste de las medias en dos vueltas sucesivas. Cuando la probabilidad asociada al valor de t sea superior al nivel de significación determinado previamente podrá finalizarse el proceso, ya que no se observan diferencias entre la media de respuesta en una y otra consulta.
- Test no paramétricos. En los casos en que no sea recomendable la utilización de la prueba de t, porque los datos no cumplen las condiciones de normalidad, podrán emplearse pruebas no paramétricas para muestras relacionadas. El test de McNemar y el de Wilcoxon son los más recomendables en estos estudios. El primero de ellos es particularmente adecuado en diseños «antes-después», situación precisamente que se da en el análisis de la estabilidad. Esta prueba atiende a la significación de los cambios, pero no a su dirección y magnitud, por lo que en preguntas de tipo escalar supone pérdida de información. En estos casos resulta más poderosa la prueba de Wilcoxon, especialmente con muestras grandes (Siegel, 1975).
- Proporción de expertos que modifican su valoración. Debe decidirse con antelación qué porcentaje de sujetos que no cambiaron su respuesta es considerado aceptable; cuando este se alcance, dicha pregunta no será incluida en la vuelta siguiente.
- Índice de concordancia: Para el análisis de la estabilidad es posible utilizar índices de concordancia entre dos vueltas sucesivas como el coeficiente kappa de Cohen. Este tiene en cuenta el porcentaje de acuerdo entre las respuestas dadas en ambas consultas sucesivas introduciendo la corrección de la posible concordancia por azar.

Como en el caso de los otros índices debemos establecer previamente el nivel aceptable de concordancia entre dos vueltas. Es frecuente en estudios observacionales considerar aceptables aquellos por encima de 0,7 (Bakeman y Gottman, 1989). Hay que tener en cuenta que este coeficiente entiende por desacuerdo la existencia de discrepancia entre la puntuación de una y otra vuelta sin atender a la magnitud de dicha diferencia, cuando esto puede tener interés en los estudios Delphi.

El uso de criterios de estabilidad presenta una seria dificultad, la mortalidad de la muestra, muy frecuente en este tipo de estudios. Cuando esto sucede se debe tomar para el cálculo de los coeficientes sólo aquella parte de la muestra que participó en las dos vueltas consideradas.

Dados los inconvenientes asociados tanto al consenso como a la estabilidad, puede ser aconsejable la utilización de ambos criterios conjuntamente. La especificación de un estándar atendiendo al consenso permite seleccionar en la primera consulta las cuestiones en las que existe el nivel esperado. El criterio de estabilidad requiere, necesaria-

mente, que contemos al menos con dos rondas para poder tomar decisiones, por lo que podrá ser utilizado tras la segunda o tercera consulta. A la hora de seleccionar el criterio y el estadístico correspondiente, debe considerarse el tamaño de la muestra, la normalidad de la distribución y el tipo o tipos de pregunta utilizadas en el instrumento de recogida de datos.

Tomadas estas decisiones, los resultados obtenidos en la primera ronda de preguntas cerradas determinarán el formato y contenido del cuestionario posterior, que no incluye ya las preguntas en las que atendiendo a los criterios fijados se alcanzó el nivel deseado, pero que ofrece información sobre la respuesta promedio dada por el grupo en cada una de las preguntas que lo forman. El equipo de investigación deberá decidir que información estadística se ofrece a los panelistas, para lo que considerará los criterios de consenso o estabilidad fijados, el perfil de los sujetos que lo forman, así como la existencia de subgrupos en el marco del panel que pueda conllevar, en ocasiones, a indicar más de una medida de tendencia central, la del conjunto del panel, y la del grupo al que pertenece el sujeto. Lo más frecuente es ofrecer la media o la mediana, si bien es más adecuada la media recortada, aunque también se puede indicar el porcentaje de respuesta y en ocasiones una medida de dispersión, si bien esto requiere contar con un grupo de participantes con ciertos conocimientos para comprender a información ofrecida.

### **3. LA TÉCNICA DELPHI EN LA SELECCIÓN DE CONTENIDOS EVALUATIVOS EN LOS PROGRAMAS DE FORMACIÓN OCUPACIONAL**

La técnica Delphi, inicialmente utilizada en estudios de carácter prospectivo es hoy también una estrategia empleada en otros ámbitos de investigación, entre ellos, como ya indicamos, en estudios evaluativos. En nuestro caso, esta herramienta de recogida y análisis de información fue utilizada en el marco de los cursos de Formación Ocupacional de la Comunidad Autónoma Gallega. Este estudio pretendía priorizar los contenidos evaluativos que en el contexto de los cursos de formación ocupacional eran considerados, de forma consensuada, especialmente importantes por los distintos grupos implicados en estas acciones. Con ello buscábamos así realizar una primera consulta entre las audiencias más cercanas a estos programas que sirviera de referencia en la selección de los contenidos evaluativos, hasta ahora basada tan sólo en las decisiones del propio evaluador.

Para la configuración del panel establecimos cinco tipos de «experto» atendiendo a la relación con estos cursos: profesores de las universidades gallegas, personal de la administración laboral, técnicos pertenecientes a los centros formativos, docentes y alumnado. A su vez se pretendía que en el marco de cada uno de estos grupos hubiera personas con vinculación a instituciones distintas —privadas, sindicales, gestionadas directamente por la administración autonómica, etc.— y de las cuatro provincias gallegas.

El panel quedó constituido, tras la renuncia de algunos de los invitados y la falta de respuesta de otros, por 58 expertos en la primera ronda y 45 en la segunda. En la tabla siguiente presentamos información más detallada al respecto.

TABLA 1  
COMPOSICIÓN DEL PANEL DE EXPERTOS

Grupo	Invitados 1ª vuelta	Aceptantes	Válidos	Invitados 2ªvuelta	Aceptantes	Válidos
Universidad	6	6	5	5	5	5
Administración	9	9	4	4	4	4
Técnicos	27	25	18	18	14	13
Docentes	24	24	20	19	16	15
Alumnado	25	25	11	10	8	8
Total	91	89	58	56	47	45

El instrumento empleado recogía un conjunto de aspectos, indicadores en muchos casos, de los cursos que podrían ser evaluados. Se optó por el formato de pregunta con respuesta de tipo escalar, de 1 a 5 puntos. Cada participante debía otorgar a cada uno de estos aspectos dos puntuaciones, una relativa a la relevancia y otra a la prioridad. En la escala utilizada el uno indicaba así poca relevancia o prioridad, el tres moderada y el cinco muchísima. El cuestionario contemplaba, además, la posibilidad de que los participantes propusieran nuevos elementos de valoración en cada uno de los bloques en que se presentaba de forma estructurada la información.

Previamente a la recogida de datos se decidieron los criterios de selección de los ítems en los que se consideraba había suficiente acuerdo, lo que conllevaba seleccionar también los que requerían nueva consulta. Debido a los objetivos del estudio se decidió también el criterio que llevaba a la eliminación de los ítems considerados de poca relevancia y prioridad por el grupo de participantes.

Se desestimaron aquellos ítems con una puntuación promedio baja, tanto si existía un amplio consenso en su valoración como si no. En concreto, se tomaron dos estadísticos de referencia, el valor de la media del total de la muestra, que debía ser mayor de 3,5 en ambas dimensiones, relevancia y prioridad, y la mediana en cada uno de los cinco grupos que forman el panel, que debía ser superior a tres también en cada una de las dimensiones.

De los ítems inicialmente no desestimados por superar el valor indicado de la media y de las medianas se seleccionaron aquellos en los que existía consenso, para lo que se fijaron dos referentes complementarios. En primer lugar, el porcentaje de valoraciones altas emitidas, —es decir correspondientes a puntuaciones de cuatro y cinco— debía ser mayor o igual al ochenta por ciento en ambas dimensiones; en segundo lugar, cuando esto no ocurría, se analizó la dispersión de las respuestas emitidas, de forma que si la desviación típica era igual o inferior a 0,90, el ítem era también seleccionado, lo que se corresponde con el 63,64% de la desviación típica uniforme.

## ALGUNAS CONSIDERACIONES METODOLÓGICAS

A lo largo del diseño y puesta en práctica de este estudio nos encontramos con diversas dificultades derivadas de la propia técnica y de los objetivos propuestos en la investigación, dificultades que atañen, fundamentalmente, a decisiones de tipo metodológico relacionadas con la configuración del panel, la mortalidad de la muestra, los criterios de selección de items y los consecuentes puntos de corte, y la finalización del proceso.

### Configuración y mortalidad de la muestra

En primer lugar entendíamos que el panel de expertos debía estar compuesto por sujetos de los distintos grupos relacionados con este tipo de cursos en el contexto gallego; esta decisión nos llevó a definir qué colectivos debían integrar el panel y a utilizar una estrategia de selección de sus miembros que garantizara la presencia de personas de los grupos propuestos. Así optamos por un muestreo no probabilístico por cuotas no proporcional que nos permitiera otorgar suficiente representación a los distintos sectores poblacionales y a la posible pluralidad de opiniones.

Si bien no es recomendable paneles tan numerosos, por la dificultad que presenta el desarrollo del proceso, la previsión de una moderada participación en el estudio nos llevó a invitar inicialmente a una amplia muestra de expertos. Como esperábamos, a pesar de la aceptación, parte de ellos no cubrieron el cuestionario, lo cubrieron fuera de plazo o lo entregaron en su mayor parte en blanco.

Nos encontramos así con el abandono propio de estudios en los que se requiere consultar en más de una ocasión a un mismo sujeto. Para evitarlo en la mayor medida posible los cuestionarios fueron entregados y recogidos en mano, y en casos excepcionales remitidos, por correo una vez cubiertos, dándole al participante el sobre preparado para ser enviado. A pesar de ello, como se observa en la tabla presentada anteriormente, existió el abandono trece panelistas de una ronda a la otra.

Esta mortalidad de la muestra no es inusual en estos estudios, pero debe tenerse en cuenta a la hora de valorar los resultados y de decidir la finalización del proceso. En nuestro caso, como puede observarse en la tabla, ninguno de los que abandonaron tras la primera vuelta pertenecía a los sectores de la universidad o de la administración, dos grupos muestrales con los que manteníamos, por motivos profesionales, mayor relación, lo que nos lleva a pensar, que en este como en otros casos, la existencia de vínculos con los propios investigadores, contribuyó a la participación. Por otra parte, hemos de señalar que de los trece que abandonaron, cinco de ellos pertenecen a la localidad en la que se optó, por limitaciones temporales y de recursos, por el envío de los cuestionarios por correo y no su recogida en mano.

La opinión que manifestarían en rondas sucesivas aquellos que abandonan es forzosamente desconocida pero no por ello carece de importancia; habría que preguntarse como afectarían sus valoraciones a los resultados. Ya que esto resulta imposible, sí cabe analizar los datos de la vuelta en la que participaron si no contáramos con ellos. En nuestro caso, bajo esta condición, es decir analizando los datos de la primera vuelta

solamente contando con los sujetos que participaron también en la segunda, observamos que la gran mayoría de los ítems presentan medias más altas, y desviaciones típicas más bajas; es decir la valoración de los distintos aspectos es mayor y la dispersión en la respuesta es menor, por lo que cabría decir que existe mayor consenso entre los panelistas. De este modo ítems que no fueron seleccionados en la primera consulta si lo hubieran sido si el panel estuviera constituido por este grupo restringido de cuarenta y cinco sujetos. Cabe pensar, por lo tanto, que en nuestro caso los que abandonaron son globalmente sujetos con valoraciones más bajas de los ítems que el resto del panel, situación que se confirma en otros estudios en los que se observó una relación significativa entre el grado de disconformidad y el nivel de abandono (Landeta, 1999).

### **Selección de criterios**

La determinación de los criterios de selección de los ítems conlleva la operativización del consenso y la necesidad de establecer límites entre lo que es un nivel aceptable y lo que no. La falta de investigaciones anteriores sobre el tema es lo más usual, por lo que no existen referentes previos sobre la tendencia de respuesta. En todo caso la configuración concreta de cada panel es un importante condicionante, aún más si está formado por grupos distintos que pueden tener valoraciones muy dispares sobre los elementos de la consulta, las que deben tenerse en cuenta además de las valoraciones globales del panel.

El contenido concreto de cada ítem, el formato de pregunta utilizado e incluso la disposición de estos en el cuestionario son algunas cuestiones que pueden influir en las valoraciones de los participantes, pudiendo existir niveles de consenso óptimos muy distintos entre ítems. Es, por lo tanto, una cuestión compleja que se debe resolver en el marco de cada estudio.

En nuestro caso la propia finalidad de la investigación obligaba a establecer el punto de corte que determinaba la selección o no de los aspectos valorados. Dada la constitución del panel se esperaba que los ítems seleccionados no sólo fueran importantes para el conjunto de participantes sino también para cada uno de los cinco grupos que lo integraban, por ello se conjugaron dos criterios, la media del panel y la mediana de los grupos, teniendo en cuenta el menor tamaño de estos. Por otra parte, en la escala utilizada el tres indica moderada relevancia y prioridad por lo que con las dos medidas establecidas, —media de tres y mediana de 3,5— se garantizaba que todos los elementos seleccionados fueran considerados cuando menos importantes para todos los grupos y para el conjunto de la muestra.

En el caso del consenso se optó por utilizar dos criterios complementarios. El porcentaje de acuerdo, que garantiza la selección de los ítems con mayor importancia y acuerdo del panel, y la comparación de la desviación típica de cada ítem con la desviación típica uniforme correspondiente. El porcentaje establecido de referencia, 80% es el utilizado en este tipo de estudios. Así, se considera un consenso alto cuando la suma de los porcentajes de dos puntuaciones contiguas oscila en torno al setenta y cinco u ochenta por ciento (Pérez-Campanero, 1991). La selección de la proporción de la desviación típica uniforme atendió a las virtudes que presenta este criterio frente a otros ya expuestos, este supera las limitaciones de aquellos que se ven alterados cuando las dis-

tribuciones son asimétricas, como era nuestro caso, ya que toma como referencia un valor teórico dependiente de la escala utilizada, no de la distribución observada del ítem, al tiempo que es el mismo valor de referencia para todos los ítems.

### **Finalización del proceso**

Analizados los resultados de la segunda vuelta, consideramos la posibilidad de aplicar un tercer cuestionario que recogiera aquellas cuestiones en las que no se había alcanzado el nivel esperado de consenso. Son dos las razones principales que nos llevaron a desestimar esta nueva consulta. Por un lado el reducido número de ítems sobre los que aún no se habían alcanzado los niveles de convergencia establecidos, lo que junto con las limitaciones de recursos y tiempo, nos llevó a valorar la rentabilidad de esta tercera vuelta. Por otro, la posibilidad de que de nuevo se produjera el abandono de una parte de los panelistas, con lo que se vería afectada la validez externa del estudio.

Estos tres elementos —los resultados alcanzados, la mortalidad de la muestra, y la rentabilidad de una nueva aplicación— deben considerarse en este tipo de investigaciones a la hora de decidir si realizar una nueva consulta. Así también no todas las cuestiones planteadas a los participantes han de tener la misma relevancia, por lo que el equipo de investigación debe valorar si es oportuna esta nueva consulta cuando los aspectos más importantes ya fueron seleccionados. Finalmente, a estas consideraciones habrá que añadir las posibles limitaciones temporales y presupuestarias.

### **Consideraciones finales**

La técnica Delphi, todavía escasamente utilizada en el ámbito de la investigación evaluativa, presenta interesantes potencialidades como estrategia para integrar los intereses de los distintos grupos vinculados al programa que se pretende valorar. Por un lado, en ocasiones, las diferencias entre dichos grupos hacen aconsejable el empleo de una estrategia de recogida de información que evite el encuentro entre ellos y las posibles influencias y presiones de grupo. Por otro, el excesivo coste y tiempo que supone el uso de entrevistas grupales o grupos de discusión, con las dificultades añadidas en muchos casos de la dispersión geográfica, aconsejan en ocasiones esta técnica.

Delphi resulta idónea cuando se pretende integrar propuestas de distintos colectivos, ya que se emplea para determinar de forma consensuada la asignación de pesos a los diferentes elementos que definen un problema, de cara a establecer prioridades. Por ello puede facilitar la convergencia de puntos de vista inicialmente diferentes en torno a diversas cuestiones que afectan al diseño y puesta en marcha de la evaluación. En concreto, la priorización de criterios o de contenidos puede ser consensuada entre las audiencias. Los resultados de dicha consulta podrán ser considerados por el equipo evaluador que dispondrá de información que le permita ofrecer una propuesta evaluativa que vaya más allá de los propios intereses del cliente o de las decisiones técnicas que tome como experto.

En el marco de programas de intervención estables, como los cursos de formación ocupacional, el uso de esta técnica permite recoger información relevante que posibilite

la creación de una batería de indicadores útiles en el diseño sistemas de evaluación que deben ser puestos en práctica de forma periódica. Los resultados pueden así convertirse en una referencia imprescindible cuando se pretende tener presente el punto de vista de las audiencias, sin tener que consultarlas en cada aplicación.

Finalmente, la técnica Delphi se presenta como una vía para la democratización de los procesos de evaluación de programas y servicios públicos en un momento en que la participación de los ciudadanos no es ya sólo un derecho social, sino que se convierte en una pauta metodológica deseable.

## BIBLIOGRAFÍA

- Bakeman, R. y Gottman, J.M. (1989). *Observación de la interacción: introducción al análisis secuencial*. Madrid: Morata.
- Dalkey, N.J., Rourke, D.L., Lewis, R., e Snyder, D. (1972). The Delphi method: An experimental study of group opinion. En N.J. Dalkey (Dir). *Studies in quality of life*. (pp. 13-55). Massachusetts: Lexington Books.
- Del Rincón, D., Arnal, J., Latorre, A. e Sans, A. (1995). *Técnicas de investigación en Ciencias Sociales*. Madrid: Dykinson.
- De Miguel, M. (2000). La evaluación de programas sociales: fundamentos y enfoques teóricos. *Revista de Investigación Educativa*, vol 18, nº 2, 289-317.
- Fazio, L.S. (1986). The Delphi: education and assessment in Institutional Goal Setting. *Evaluation Studies: Review Annual*, 11, 287-296.
- Fernández-Ballesteros, R. (1995). *Evaluación de programas: una guía práctica en ámbitos sociales, educativos y de salud*. Madrid: Síntesis.
- Hakim, S. e Weinblatt, J. (1993). The Delphi process as tool for decision making. *Evaluation and program planning*, 16, 25-38.
- Landeta, J. (1999). *El método Delphi. Una técnica de previsión para la incertidumbre*. Barcelona: Ariel.
- Palmer Pol, A.L. (1999). *Análisis de datos. Etapa exploratoria*. Madrid: Pirámide.
- Pérez-Campanero, M.P. (1991). *Cómo detectar las necesidades de intervención socioeducativa*. Madrid: Narcea.
- Renom Pinsach, J. (1992). *Diseño de test*. Barcelona: Engine.
- Ruiz Olabuénaga, J.I. e Ispuzua, M.A. (1989). *La descodificación de la vida cotidiana*. Bilbao: Universidad de Deusto.
- Siegel, S. (1975). *Estadística no paramétrica aplicada a las ciencias de la conducta*. Méjico: Trillas.
- Uhl, N.P. (1991). Delphi Technique. En Lewy, A. (Ed.), *The International Encyclopedia of curriculum* (pp. 453-454). Beverly Hills, Ca.: Sage.

Fecha de recepción: 18 de enero de 2002.

Fecha de aceptación: 30 de octubre de 2003.