The Citizenship Education Policy Study (CEPS) represents an unprecedented attempt to design and implement a cross-cultural variation of the Delphi technique. It engaged a distinctively large team of researchers with diverse national, cultural, linguistic, conceptual, and professional resources in a collaborative effort to solicit, integrate and interpret the collective wisdom generated by a prestigious, multinational panel of experts. The expert opinion was then used to develop educational policy relevant both within the participating nations and beyond any particular policy system. In attempting this, the present study joins the tradition of its methodological predecessors in grappling with the risk, rigour and creativity required to address questions of significance to present and future social and educational systems.

ORIGIN AND EVOLUTION OF THE DELPHI METHOD

Development and description of the conventional Delphi

The Delphi was originally conceived as an intuitive, exploratory method to solicit and synthesize the forecasts of groups of experts regarding problems that did not lend themselves to precise analytical techniques but that could benefit from the application of carefully derived collective judgement. It originated in the United States as a spin-off of national defence research in the early 1950s. The US Air Force commissioned the Rand Corporation to conduct a study entitled Project Delphi, the goal of which was to 'obtain the most reliable consensus of opinion of a group of experts', by means of 'a series of intensive questionnaires interspersed with controlled opinion feedback' (Linstone and Turoff, 1975, p10).

Iteration with controlled feedback, as developed by Norman Dalkey and Olaf Helmer, was the principal methodological innovation in the original project. In 1959, with a third Rand researcher, Nicholas Rescher, they published The Epistemology of the Inexact Sciences to begin the process of establishing a theoretical justification for use of the Delphi in various forms of anticipatory research. The authors contended that the judgement of experts is permissible as scientific evidence in fields which have not yet developed to the point of establishing formal
Citizenship for the 21st Century

scientific laws. The validity and applicability of the Delphi were supported by a series of forecasting experiments conducted by Rand in 1968. Regarding estimates generated by groups, Rand researchers concluded that assessments formulated on the basis of anonymous controlled feedback tended to be more accurate than estimates emerging from face-to-face discussions (Linstone and Turoff, 1975).

These initial studies provided evidence in support of three central assumptions which have served as a philosophic basis for the development of Delphi research since its inception:

1. Human judgements represent legitimate and useful inputs in addressing research problems that are long-range, ill-defined, highly complex, and/or lack a well-developed theoretical foundation.
2. The judgement of a group is likely to be superior to that of any individual, especially if judgements are arrived at in an interactive manner involving carefully structured sharing of information.
3. Responses shared anonymously are likely to be superior (more numerous, detailed, creative and candid) to those publicly identified with their source because participants who respond anonymously are not subjected to the biasing effects of dominant individuals, group pressures towards conformity, irrelevant communication, and fear of public disapproval.

Based upon these assumptions, conventional Delphi studies are generally conducted in the following manner:

The researcher uses explicit criteria to select a panel of experts and designs a well-structured questionnaire concerning the issue(s) under consideration. Panellists are then asked to respond to the questionnaire during a series of rounds (iterations). All responses are provided individually and anonymously. Questionnaires are usually administered through the mail, and in most cases panellists are unaware of each other’s identity and interact only with the researcher or small research team (two to four members) conducting the process.

Between rounds, panellists are provided with descriptions of previous individual and group responses (controlled feedback). Group opinion is generally expressed in the form of statistical indices, a measure of central tendency (usually the median response) and a measure of dispersion (usually the interquartile range). Panellists are frequently asked to submit comments and/or justification of their personal views for review by the researchers which are often summarized and reported in subsequent iterations.

The Delphi process is concluded either when a predetermined level of agreement among panellists is achieved (group consensus), or when the responses of individual panellists have stabilized from round to round making it apparent that further repetitions will not produce the predetermined level of agreement (stabilization of disparate opinion). Between two and four rounds are usually sufficient to complete the process.

Cultural futures variations of the conventional Delphi

Although conceived as a quantitative method for acquiring expert opinion regarding technological forecasts, expanded conceptualizations acknowledging the potential
of the Delphi as a method for enhancing group communication and judgement concerning complex issues soon emerged (Linstone and Turoff, 1975). Many of the most significant problems confronting the emerging discipline of futures research were broadly cultural rather than narrowly technical in scope. For researchers compelled to respond to such concerns, the primary purpose of futures research began to shift from an emphasis on social forecasting to an emphasis on social systems design. More specifically, the purpose of cultural futures research is to engage a 'community' in developing both a shared vision of ways of life that are possible, desirable and sustainable, and feasible strategies necessary to enact that vision. The Delphi appeared particularly promising as a tool for social systems design due to its demonstrated capacity to build consensus and to enhance understanding of complex issues.

The educational impact of participation in Delphi exercises has been repeatedly demonstrated (eg, Hill and Fowles (1975), Kurth-Schai (1984, 1988, 1991), Linstone and Turoff (1975), Palkert (1986), Poolpatarachewin (1980), Scheele (1975), Weaver (1971), Weingand (1980). In addition to broadening and deepening understanding of the topic under consideration, past studies suggest that participation catalyses and supports clarification of facts and values, exploration of alternative perspectives and possibilities, and development of critical, creative and systemic thought. These effects are accomplished while emphasizing the importance of collaboration across diverse opinions and experiences in processes of social problem-solving and design. Of particular significance to the prospect of conducting cross-cultural Delphi research are the thoughts of Sam Scheele (1975) who has suggested that the educational impact of the Delphi, more specifically the opportunity to experience others' conceptions of reality, represents perhaps a greater contribution to society than the sharing of final results. He contends that focusing attention on differences in reality constructs usually yields a more refined and widely accepted definition of the appropriate construct or at least a clearer and more precise distinction between competing constructs. The redefinition of contextual realities facilitates the generation of new options and may create impetus for change whereby, 'both the dominant reality and the required change technology are invented as well as inherited, and culture is transformed as well as transmitted’ (Scheele, 1975, p43).

Adaptation of the Delphi to support cross-culturally relevant policy

Once a cultural systems design variation of the Delphi was selected, the process of adapting the technique to meet the goals, needs, constraints and challenges specific to this study could begin. A review of the literature suggested that although 'traditional professional strategies for building consensus, such as policy statements, publication and conferences, often have failed to chart a clear and uncontested sense of group agreement . . . in education the Delphi has been used effectively for applications such as curriculum planning and development and goal setting’ (Martorella, 1991, pp 83-4). The Delphi has further been described as particularly well suited to the task of generating educational policy due to its ability to accommodate multiple competing insights and interests (Cookson, 1986), its ability to accommodate multiple interpretations and to expand educators' awareness of alternative future options (Palkert, 1986), and its future orientation deemed essential due to the time lag between the imposition of policy and its educational
impact and the rapid pace of social change with which educational policy must keep pace (Weaver, 1971).

A review of past studies also revealed an important adaptation, that of using data obtained by interviewing expert panellists as the primary source of items to be incorporated in Delphi questionnaires, rather than relying solely on the judgement of researchers in completing this crucial task. Perhaps the most comprehensive efforts at integrating interviews in the Delphi are represented in the work of Chumpol Poolpatarachewin (1980) who pioneered the development of Ethnographic Delphi Futures Research (EDFR). The procedural steps of EDFR are quite similar to those of the Delphi, the primary distinguishing feature being the inclusion of a future-oriented approach to ethnographic interviewing during the first round. Drawing from past experience with EDFR, round one interviews were utilized in the CEPS study so that expert opinion could be used to guide questionnaire development, thereby assuring that central themes would be addressed from multiple, diverse and knowledgeable perspectives.

OVERVIEW OF THE RESEARCH PROCESS

In September 1993, members of the steering committee met in Hiroshima, Japan, to clarify the project's purpose and identify focal research questions. As described in Chapter 1, working definitions of key terms and concepts were negotiated, and criteria were established for the selection of research team members and policy shapers and scholars from the participating nations. The committee also began the process of developing the analytic framework and specific research strategies necessary to identify a wide range of possible future global trends, characteristics and strategies relevant to the topic of citizenship education. These results would then be used to shape policy decisions - either directly (results would be used to develop new educational policy and practice) or indirectly (results and the experience of the Delphi would serve to educate all who participate regarding the scope and complexity of the issue, and change their personal and professional attitudes and behaviour).

The project was to be completed over a period of forty-five months ending in March 1997. During the first year, twenty-six researchers (with areas of expertise that included social studies education, comparative and international education, science education and a variety of research methodologies), representing nine nations were organized into four research teams: Japan, Thailand, Europe (England, Germany, Greece, Hungary, the Netherlands) and North America (Canada, United States).

Selection of research panellists

The Delphi method is designed to collect and synthesize the opinions of panels of experts. Unlike more conventional survey methods the goal in selecting Delphi panels is not to develop a random sample representative of the general public. Instead, a purposive sampling approach is adopted. The intent is to select persons who have special knowledge or expertise in the domain(s) being examined. Predetermined selection criteria are strictly adhered to to ensure both validity and reliability.
Attention was also devoted to balancing and diversifying the membership of each panel (by gender, race/ethnicity, geographic region and area of expertise, eg, environment, economics) to ensure that a wide variety of perspectives on the topic would be carefully considered.

**Figure 3.1 Project Schedule**

During the months of August 1994–January 1995, members of the four research teams conducted interviews with carefully selected members of each expert panel. In all, 110 interviews were conducted producing a rich body of thoughtful social analysis. Three broad, open-ended questions were posed to each interviewee:

- What are the major global trends likely to have a significant impact on the lives of people during the next twenty-five years?
- What will be the characteristics required of individuals in order to cope with and/or manage these trends?
- How might these characteristics be developed, ie, what educational approaches, strategies or innovations might best implement these citizen characteristics?
The questions were sent to the interviewees for their consideration prior to the interview. Most interviews were conducted in person, and in a few cases by telephone, and tape-recorded if agreed upon. Interviews lasted 45 to 120 minutes with many respondents showing willingness to continue well beyond the time allotted and indicating enthusiasm for their involvement in what they found to be an important, thought-provoking and enjoyable process.

Upon the completion of each interview, the researcher produced a written summary identifying each concept raised by the expert. As the interview data were to be used only to generate Delphi statements, it was not necessary to translate the summary transcripts into English. Researchers then aggregated their interview data, grouping similar concepts together into statement categories such as environment, economic development, politics and government, etc, and developed draft Delphi statements.

Nearly 900 draft Delphi statements were generated across the four research teams, organized as trends, characteristics and educational strategies/approaches/innovations in relation to the three focal interview questions. In May 1995, researchers met in Minneapolis to face the task of condensing the draft statements into a workable Delphi instrument of 106 items. The researchers were divided into five multinational working groups who used the following criteria to guide their discussions and decisions in arriving at the final listing of Round Two Delphi survey instrument statements:

- State items at the appropriate level of specificity (statements should allow for discrimination in responses and should challenge respondents’ thinking).
- Synthesize related statements to avoid redundancy.
- Work towards clarity in meaning.
- Determine whether the issue provides valuable insight for developing educational policy related to citizenship.
- Ensure that the statement is relevant across cultures.
- Ensure that statements cover the range of issues raised by interviewees.

From the collection of items proposed by each working group, one shared multinational Delphi questionnaire was developed and divided into three sections, each with a specific pattern of response requested as follows:

Section I: Trends

This section of the survey consisted of sixty global trends identified by experts participating in the interview round as likely to have a significant impact upon the lives of people during the next twenty-five years. Policy experts responding to the questionnaire were asked to assess both the probability and desirability of each trend using a six-point scale ranging from Highly Likely or Highly Desirable to Not Likely or Not Desirable. For example:
Using the Delphi cross-culturally

<table>
<thead>
<tr>
<th>Desirability</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highly desirable</td>
<td>Not desirable</td>
</tr>
<tr>
<td>Not likely</td>
<td></td>
</tr>
<tr>
<td>Highly likely</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5</td>
</tr>
</tbody>
</table>

**Figure 3.2** People will continue to support economic expansion even though it may increase the stress on the environment

**Section II: Characteristics**

This section of the survey consisted of twenty characteristics of future citizens identified as necessary by experts participating in the interview round. Policy experts responding to the questionnaire were asked to select the five characteristics that will be most urgent for policy makers to consider and act upon during the next twenty-five years. For example:

Willingness to change one's lifestyle and consumption habits to protect the environment

**Section III: Strategies/Approaches/Innovations**

This section of the survey consisted of twenty-six feasible strategies, approaches or innovations identified by experts participating in the interview round as useful in preparing effective citizenry for the 21st century. Policy experts responding to the questionnaire were asked to indicate which statements they would recommend for consideration and action by policy makers during the next twenty-five years by assessing each in relation to a six-point scale ranging from Highly Recommended to Not Recommended. For example:

<table>
<thead>
<tr>
<th>Recommend</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highly recommended</td>
</tr>
<tr>
<td>6</td>
</tr>
</tbody>
</table>

**Figure 3.3** Promote schools as active centres of community life and as agents for community development

An invitation to comment on statements, to add additional statements, and/or to respond to open-ended questions describing the respondent's rationale for a particular response was also included. The original multinational questionnaire was developed in English and then translated into the native languages spoken by expert panellists (Japanese, Thai, German, Greek, Hungarian, Dutch). Back-translation was then carried out to ensure validity.
Delphi Round Two: first response to questionnaire

Questionnaires for Round Two were distributed and responded to from August through October 1995. Of the 264 policy experts who were identified and approached, 182 completed and returned questionnaires.1

As in other Delphi studies, analysis of initial responses to the research instrument focused on determining the level of consensus reached on each item and on reporting this information back to the respondents in a manner that would be useful to them in reconsidering their initial judgements during later rounds.

Criteria for analysis of consensus data

For section I, Trends, the mode, median and interquartile range scores were calculated for responses to each item. Items meeting each of the following criteria were accepted as having reached consensus:

- the mode minus the median is less than or equal to 1.0, and
- the interquartile range score is less than or equal to 1.5.2

For section II, Characteristics, items meeting each of the following criteria were accepted as having reached consensus:

- the characteristic appears in the top ten list for three of the four research teams, and
- the characteristic is selected in the top five by 25 per cent or more of all respondents.3

For section III, Strategies/Approaches/Innovations, the mode, median and interquartile range scores were calculated for responses to each item. Items meeting each of the following criteria were accepted as having reached consensus:

- the mode minus the median is less than or equal to 1.0, and
- the interquartile range score is less than or equal to 1.5.

Additional comments made by the respondents for all items were coded and recorded. Because the questionnaire was designed to solicit a multinational (rather than nation or region specific) response, it was determined that only information describing the aggregate responses of all the policy experts would be reported during the Third and final Round.

Delphi Round Three: final response to the questionnaire

Questionnaires for Round Three were distributed and responded to by policy experts from January through February 1996. Respondents received questionnaires personalized by research team members to reflect (1) the expert's Round Two response to each item, and (2) the median and interquartile range scores for trend and strategy statements and (3) the percentage of respondents selecting each characteristic as one of their five most important for the characteristics section.
Using the Delphi cross-culturally

For Sections I and III (Trends and Strategies/Approaches/Innovations), respondents were informed that consensus had not been reached for 43 of the 60 trends and 17 of the 26 strategies. These were circled on the questionnaire. Respondents were informed that although they were encouraged to review and reflect again on their judgements for every item, the researchers were most interested in responses to the circled items where there appeared to be more diversity of opinion.

For Section II (Characteristics), percentages of trait selection were reported along with indication of the expert's initial choices for feedback purposes only. No further response was requested. Experts were again invited to provide comments for all items.

In follow-up communications, respondents were encouraged to return their Round Three responses. They were also informed that should they choose not to return this final Round questionnaire by a specified date, the study would assume that their Round Two responses were to be accepted as final and would include these responses in the analysis of Round Three data. Upon conclusion of Round Three, of the 182 questionnaires distributed, 141 were returned.

Analysis and interpretation of Delphi results

The fourth and final year of the study was devoted to the analysis, interpretation and preliminary reporting of Round Three results. Early stages of these processes were completed during the Third International Meeting in Hiroshima, Japan, in June 1996. Although a range of secondary analyses focusing on comparative issues was possible, analytic activities at the Hiroshima meeting focused on interpretation of the aggregate data and development of multinational reports.

In preparation for this meeting, consensus items were organized into a prioritization scheme to assist research team members in focusing on the trends, characteristics and the educational strategies/approaches/innovations judged to be most important for shaping educational policy to prepare citizens for the 21st century. Trends reaching consensus with respect to both probability and desirability were grouped into three categories numbered in descending order of importance. The eight characteristics identified by the experts as most urgent for policy makers to consider were then listed, followed by very highly and highly recommended strategies/approaches/innovations.

Responsibilities for developing educational policy recommendations based upon key Delphi findings were then divided into categories reflecting the broad trends identified by participating experts as likely to provide the most significant challenges facing future citizens. These were: Growing Environmental and Resource Concerns, Crisis in Social Ethics, Explosion of Media and Information Technologies, and Changing Political and Demographic Patterns. Research team members were assigned to one of four multinational working groups and asked to develop draft educational policy recommendations relevant to their group's focal trend (see above) to be implemented through: Curriculum, Pedagogy/Instructional Methods/Assessment, School Organization and Administration, Teacher Education and School and Societal Interactions (i.e., relationships between local, national and international organizations/communities).

After moving through several iterations of drafting and redrafting potential policy recommendations, researchers were reorganized into writing groups, with
each group assigned to synthesize policy recommendations reflecting the organizational framework of the final report. As the Delphi process invites multiple interpretive paths, researchers as members of both multinational and national writing groups still continue to pursue the complex analytic and creative tasks of translating Delphi results into varied forms of educational policy relevant to the preparation of future citizens.

**THE PROMISE AND PERIL OF CROSS-CULTURAL DELPHI RESEARCH**

**Challenges and opportunities of methodological innovation**

Though intellectually exciting, significant challenges are posed by adopting a research technique that is not widely utilized or well understood in educational policy circles. First, the cultural futures Delphi has both variant and invariant aspects thereby inviting multiple interpretations. Second, it emerges from a distinctive philosophic framework on the basis of which methodological adaptations and innovations must be judged. And third, it provides limited opportunity for guidance or justification based on precedent. Thus, throughout the study, the demands of high levels of uncertainty, ambiguity and risk were met with a healthy sense of scepticism by research team members. In a large research team there were some for whom the Delphi was totally new. Some approached it from a psychometric stance while others were more used to dealing with qualitative data. Along this continuum the researchers worked hard to arrive at a constructive way of proceeding.

For those experienced with ethnographic techniques and skilled in eliciting rich, finely detailed and context-referenced descriptions of informants' conceptual lives, the process of 'reducing' and decontextualizing material derived from interviews into Delphi statements was highly problematic. Similar discomfort was experienced by researchers committed to forms of inferential psychological analysis for whom an expert's response to any particular Delphi statement could only be accurately understood in the context of the respondent's dispositional state and conceptual rationale for each choice.

Other methodological concerns were raised consistently by those grounded in statistically based comparative survey and scaling techniques. Pressures to employ methodological procedures ensuring construction of representative samples and comparison among essentially similar objects of study, as necessary for the appropriate use of inferential statistics, were deeply challenged by Delphic commitments to *purposive sampling*, as in the case of selecting experts on the basis of accepted criteria and then working to diversify and balance panels in relation to variables deemed important by the researchers' gender, ethnicity, area of expertise, etc. Project statisticians were also challenged by the wide range of obstacles to direct comparison inherent within the design of this study, eg, variation in composition of research teams, with Japan and Thailand as teams representing one nation and cultural heritage in sharp contrast to the North American and European teams. Similarly, the cultural/conceptual/linguistic diversity among researchers and expert panels resulted in varied interpretations of such key terms and concepts as 'citizenship' and 'expert'.

102
In response to these challenges, the researchers were faced with the need continually and collectively to work to articulate, clarify, question, reconsider, revise and newly develop both philosophic foundations and research strategies in order to achieve the study's broad-based goals. Working together, across diverse cultural and methodological experiences, the project researchers found themselves moving into largely uncharted methodological territory, catalysed by intense, engaging reflection and debate, while seeking to deepen and extend their understanding of educational research and policy design in general, and of the Delphi method in particular.

Challenges and opportunities of cross-cultural collaboration

Complex challenges are also raised by attempting to adapt a research technique centred on developing consensus, and relying traditionally on expertise emerging from shared conceptual contexts, to accommodate the varied judgements generated across multinational, cross-cultural teams of researchers and panels of experts. Throughout the course of the study, significant thought, time and energy were devoted to attempts to accommodate and constructively utilize the diversity in language, cultural heritage and patterns of social organization represented among researchers and experts. The strongly held objectives were to develop shared conceptual understandings, to reach consensus on significant policy recommendations, and to ensure full, egalitarian participation of all research participants across multiple dimensions of difference.

Developing shared conceptual understandings

Examples of the challenges posed by the struggle to arrive at shared understandings of key concepts across nine distinct social/political/cultural contexts and seven languages abound (Derricott, 1996; Ninomiya, 1996). Three are presented here, the first centred on the concept of expertise. In spite of agreements reached on criteria for expert selection, attempts at standardization in composing expert panels were challenged by definitions of ‘expert’ ranging from one who has achieved a high level of specialization in a specific field of study (e.g., a nuclear physicist or social historian), to one who holds high-level administrative or political position and power (e.g., a minister of foreign affairs or corporate CEO), to one intensely engaged in a particular experiential context (e.g., a musician as expert on trends in popular music or a child as expert on the quality of life in contemporary classrooms).

A second example concerns the nature and purpose of citizenship. Philosophical and political discussions concerning its definitive and desirable characteristics were both enlightening and highly problematic in producing Delphi statements that would be relevant within and beyond the participating nations. Contrasting images of the ideal citizen as loyal and disciplined subject willing to sacrifice personal gain to enhance the common good versus the citizen as insightful social critic and active participant in formulating and enacting social policy appeared to correspond to differences between Asian and European/North American political traditions and values. Perhaps the most difficult component of doing cross-cultural research is finding ways to cope and to collaborate across diverse and deeply held values such as these.
The third example describes the struggle for shared meaning within one research team. Euro-team meetings and written communications were conducted in English. Because four of the team members had to use their second, third, or fourth language, considerable time and effort was required to arrive at precise meanings. For example, the phrase 'Non-governmental organization' (NGO) was problematic. Early in their deliberations team members learnt that there appears to be no direct translation into German or Greek. Further consideration revealed that in the United Kingdom, NGOs are appointed and largely controlled by the government, so the concept of a 'non-governmental' organization made little sense.

Reaching consensus on policy recommendations

The struggles noted above raised interesting dilemmas upon completion of data collection, when processes of interpreting results and developing recommendations were again complicated by cross-cultural, multinational variations in the philosophy and practice of formulating and implementing social policy. Again returning to an example from the Euro-team, early in the process of developing draft policy recommendations it became clear that the five countries embraced very different practices and perspectives on educational policy. In the UK in recent years the power base for educational administration had moved from the periphery to the centre. Under the Educational Reform Act of 1988 the Secretary of State for Education assumed more than 400 new powers. The Netherlands had also developed a National Curriculum, thereby moving the balance of power in the same direction. In Germany the Federal States remained powerful, while in post-Communist Hungary the movement has been toward increasing local and professional autonomy. Variations within the Euro-team were further extended in comparison to the nations represented in the three other research teams, encompassing approaches to educational governance ranging from the highly centralized systems of Japan and Thailand to the distinctively decentralized systems of Canada and the United States.

Ensuring full participation

The task of encouraging full, egalitarian participation across an unusually large and diverse group of researchers generated its share of challenges. In addition to conventional research skills and dispositions, the need to establish and sustain trust, and to remain sensitive to the concerns of others, became essential to successful completion of all research tasks. Such qualities as patience, persistence, open-mindedness, curiosity and humour were also necessary. Because research meetings were conducted in English, careful scheduling was required to help ease the demands on non-native English speakers. Intensive two to three hours multinational working sessions were interspersed with national/regional team meetings providing opportunities to strategize, converse, relax and commiserate with colleagues sharing more similar linguistic and cultural backgrounds. Attempts were made to diversify and balance all multinational working groups taking into consideration factors including gender, English proficiency, area of interest and expertise, degree of comfort with and style of presenting personal opinions, and preferred approach to dealing with conflicting opinions. Native English speakers in multinational groups assumed responsibilities for writing and recording, and continually reminded each other to be careful not to rush or dominate discussions. Perhaps most importantly, the structure of the study allowing for face-to-face
Using the Delphi cross-culturally

interaction, over a three and one half year period, was crucial in building trusting working relationships and significant skills in cross-cultural collaboration.

Addressing concerns for methodological credibility

All research methods, but particularly those that are exploratory, must undergo careful scrutiny to assess their credibility. For all of the reasons cited throughout this chapter, and more importantly due to the social significance of the project’s goals, ensuring the reliability and validity of findings generated through this unprecedented attempt at cross-cultural Delphi research was an on-going concern.

In a well-known analysis of the conventional Delphi technique, Hill and Fowles (1975) note that reliability is usually defined in terms of the precision of measurement instruments, often demonstrated by the dependability of measurement across different replications, and usually addressed through standardization of research procedures. Although complete standardization across the complex data gathering and interpretive procedures employed here is neither necessary nor possible, the need to develop data collection and analytic techniques that could be comparable across diverse linguistic/cultural/political contexts was keenly felt. A primary rationale for actively engaging such a large, diverse and potentially unwieldy group of researchers in the design and implementation of all aspects of the research was the need to develop Delphi statements capable of soliciting comparable judgements concerning issues of shared significance across the participating nations. Intensive efforts to conscientiously translate and backtranslate questionnaires and responses between English and the native languages of respondents were also employed for the purpose of achieving this goal. Rather than relying only on the large and diverse research team to generate the Delphi questionnaire, interviews were conducted with an extensive sample of expert panellists to ensure added breadth and depth of consideration. Further, although it would have been possible in analysing Delphi results to defend the acceptance of less stringent consensus rules in light of the complexities introduced in accommodating linguistically and culturally diverse expert panels, the research team consistently chose to use the most conservative indices employed in prior studies.

Hill and Fowles (1975) further address two types of validity in their critique of the Delphi. They suggest that the first type, data validity, demonstrated by the accuracy of future projections, raises limited concern in relation to studies of this type because with respect to planning and policy-making, forecast utility is of greater importance than forecast accuracy. The second type, method validity, defined as whether the design of the method makes it possible to produce the results intended, is, however, of crucial importance here. Has participation in this Delphi process supported the generation of insights that will assist us in envisioning and enacting positive social change through education? Will the results of this study enhance our understanding, expose us to new perceptions and possibilities, expand our options, and deepen our resolve to work collaboratively and cross-culturally to prepare citizens to meet the challenges of the coming century? The chapters that follow provide evidence to answer these questions and to judge the methodological validity of this project.
IMPLICATIONS FOR THE FUTURE OF CROSS-CULTURAL DELPHI RESEARCH

In addition to providing a basis for assessing the validity and reliability of the Delphi as a tool for cross-cultural research and policy development, this study demonstrates:

- the feasibility of using a cross-cultural variation of the Delphi to enhance understanding and imagination concerning the complex topic of citizenship education
- the value of exploring specific domains of agreement and disagreement to sharpen perspectives on the status and prospects for citizenship at the dawn of the next century
- the possibility of generating significant understandings, prioritizations and projections that transcend nations, cultures and policy systems
- the opportunities and challenges posed by engaging a large, multinational team of researchers in a complex and collaborative effort
- the complexities of establishing consensus, and the generative tensions revealed in measurement and interpretation debates
- the possibility of soliciting and synthesizing the collective judgement of a prestigious panel of experts representing diverse nations and areas of specialization, and
- the possibility of affecting social and educational policy by engaging policy shapers in the Delphi process, thereby providing them with opportunities to consider a wide range of reflective and imaginative perspectives, and to develop a broader and deeper understanding of the problem under consideration and possible solution paths.

This study neither exhausts the Delphi method's exciting potentials, nor overcomes its persistent limitations. Similarly, the project represents both the promise of engaging multinational research teams and participants, and the many areas of difficulty imposed by such an endeavour. Apart from all considerations of the validity of the methodology, the quality of analyses and interpretations, or the utility of the findings and recommendations, this is the story of a complex accomplishment. A justifiable consensus is expressed by a distinguished international panel. A rigorous and innovative research design is executed with prodigious effort by an international research team. And a defensible body of challenging ideas and interpretations is generated to provide a basis for the development of educational policy.

ENDNOTES

1. Based upon his research, Thomas Macmillan (1971) noted that a panel size of at least 17 is necessary to ensure the lowest rate of error in Delphi studies, thus requiring at least 68 experts distributed across the four panels used in this study, a number well exceeded by the total number of 182 expert respondents. Further, it is the question of the quality of the panel experts that is most important in the Delphi, i.e., that each of them met the selection criteria.
2. Following Dalkey and Helmer's (1963) and then Helmer's (1966) lead, Delphi studies have traditionally used two indices to determine consensus and to report
findings back to Delphi panellists: the median and interquartile range. The
median is used instead of the mean as a measure of central tendency because
expert panels are not comprised of a random sample. Interquartile range is
used to provide a sense of each statement’s degree of clarity, ie, statements for
which the interquartile range falls beyond a prescribed range are considered
ambiguous and therefore not indicative of consensus. The smaller the accept-
able interquartile range, the more likely it is that the results reflect authentic
consensus of expert judgement.

3. The concept of consensus is a complex one. Although an item may have reached
consensus in reference to the aggregate data determined in the manner described
here, it may not have reached consensus among experts responding from a
particular nation or region. Because experts were selected based in part on
their ability to think globally from within their particular cultural context (to
balance culturally specific and global perspectives), it was determined that
formulation of multinational policy recommendations would focus on items
reflecting aggregate consensus, whether or not consensus is also achieved on
a team by team basis.

REFERENCES

Cogan, J J (1994) Notes from February CEPS Steering Committee Meeting, Bangkok,
Thailand.
3–13.
ive on Methodological Issues in Conducting Cross-Cultural Research’, Paper
presented at the Annual Western Regional Conference of the Comparative and
International Education Society, University of Hawaii, January 3–5.
Helmer, O (1966) The Delphi Method for Systematizing Judgments about the Future,
Institute of Government and Public Affairs, UCLA.
thesis, University of Minnesota, Minneapolis.
Linstone, H and Simmons, W H C (eds) (1977) Futures Research: New directions,
Addison-Wesley, Reading, MA.
Linstone, H and Turoff, M (eds) (1975) The Delphi Method: Techniques and applica-
tions, Addison-Wesley, Reading, MA.
annual meeting of the California Junior Colleges Association Commission on
Research and Development (3 May 1971), Monterey, California.
Citizenship for the 21st Century


Scheele, S (1975) 'Reality Construction as a Product of Delphi Interaction', in H A Linstone and M Turoff (eds), The Delphi Method: Techniques and applications, Addison-Wesley, Reading, MA.
