
델파이 기술예측의 타당성과 신뢰성 분석에 관한 연구

권성훈* · 홍순기**

<목 차>

- I. 서론
- II. 델파이 방법
- III. 델파이의 타당성과 신뢰성
- IV. 타당성과 신뢰성 분석
- V. 결론

국문초록 : 델파이는 전문가의 의견에 기초하여 미래를 예측하는 방법이다. 델파이가 얼마나 타당하고 신뢰할 수 있는지를 검토하는 것은 중요하다. 이 논문에서는 우리나라의 IT와 BT 분야의 델파이 조사의 정확도와 정밀도를 분석하고, 그 관계에 대해 논의한다. 분석 결과, 예측결과의 정확도와 정밀도는 그 분야와 전문도에 따라 일부 차이를 보였다. 또한, 전문도가 높은 응답의 경우, 예측결과의 정확도와 정밀도 간의 유의한 상관관계가 발견되었다. 이 결과는 예측의 정밀도가 정확도를 판단하는 하나의 기준이 될 수 있음을 의미한다.

주제어 : 델파이, 타당성, 신뢰성, 정밀도, 정확도

* 성균관대학교 기술경영학과 박사과정(seonghoonk@skku.edu)

** 성균관대학교 시스템경영공학과 정교수(hongsk@yurim.skku.ac.kr)

I. 서론

기술예측의 목적은 미래 기술의 기회를 탐색하고 그 동향을 다각도로 분석하여 의사 결정자에게 유용한 정보를 제공하는 데 있다. 국가 차원의 과학기술예측조사 결과는 과학기술 분야의 제반 계획수립을 위한 근거자료로 활용될 뿐만 아니라, 미래의 연구개발 방향을 설정하는 데 중요한 역할을 한다.

델파이(Delphi)는 1960년대 이후 한국, 일본, 독일의 국가수준의 대규모 예측조사에 사용된 대표적인 예측기법이다. 델파이는 RAND 연구소에서 개발된 이후 그 유용성에 대한 일부 비판에도 불구하고 꾸준히 적용 분야를 넓혀왔다.

예측결과를 유효하게 하기 위해서는 과거 조사에 관한 평가와 분석을 통해 예측방법을 검토하는 것이 필요하다. 일본의 경우, 1회부터 4회까지의 과학기술예측조사에서 채택한 과제의 실현율(각각 69%, 68%, 73%, 66%)을 평가했다(일본 문부과학성 과학기술정책연구소·미래공학연구소, 2006). 한국의 「제1회 과학기술예측조사를 수행한 지 약 15년이 지난 현 시점에서, 그 예측결과가 얼마나 타당하고 신뢰할 수 있는지, 즉 정확도(accuracy)와 정밀도(precision)를 검토하는 것이 필요하다.

이에 본 연구에서는 1994년에 과학기술정책연구원(STEPI)이 수행한 「제1회 과학기술예측조사」의 ‘정보·전자·통신’ 분야와 ‘생명공학’ 분야를 대상으로 델파이의 타당성(validity)과 신뢰성(reliability)을 분석하고, 그 관계를 파악하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 델파이의 특성과 절차를 설명한다. 3장에서는 델파이의 타당성과 신뢰성의 개념과 관계를 고찰하고, 4장에서는 이를 이용하여 실증 분석을 실시한다. 5장에서는 연구결과의 요약과 함께 시사점을 제시한다.

II. 델파이 방법

1. 델파이의 특징

델파이는 전문가 그룹의 의견을 체계적으로 도출해서 수립하기 위해 고안된 기법으로(홍순기·오정목, 1997), 다음의 세 가지 특성을 가진다(Bright, 1978; Martino, 1993):

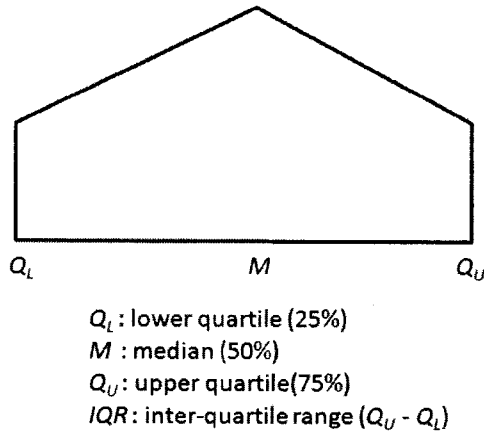
첫째, 델파이예측에서는 익명성(anonymity)이 보장되어 전문가는 자신의 의견을 자유롭게 교환할 수 있다.

둘째, 델파이에서는 그룹 전체의 예측시기를 통계적으로 집계하여 <그림 1>과 같은 중위수(median)와 사분위수범위(IQR: inter-quartile range), 즉 통계분포(statistical display)를 제시한다.

셋째, 패널이 설문서를 통해 의견을 서로 교환하고 수정응답을 하는, 즉 추론의 피드백(feedback of reasoning)에 의해 합의(consensus)가 유도된다.

중위수는 예측치를 이른 시점부터 순서대로 나열했을 때 50%에 해당하는 값이다. 상사분위수(upper quartile)와 하사분위수(lower quartile)는 각각 75%와 25%에 해당하는 응답자가 예측한 값을 나타낸다. 사분위수범위(이하 IQR)는 중위수를 중심으로 응답의 산포(dispersion)를 나타낸다.

<그림 1> 중위수와 사분위수



2. 델파이의 수행과정

전통적인 델파이는 4-라운드(round)의 설문서를 통해 수행된다. 그런데, 많은 경우 라운드 2(2nd round)에서 가장 큰 수렴이 일어난다(홍순기 외, 2007). 델파이 조사의 시간과 비용을 고려한다면, 2-라운드 미니델파이(Mini-Delphi)가 타당할 수 있다. 여기서는 과학기술정책연구원(1994)의 「제1회 과학기술예측조사」에 사용된 미니델파이를 설명한다.

라운드 1(1st round)에서 패널(panel)은 설문서를 받아 각 과제의 실현시기를 예측하

게 된다. 총괄수행자(moderator)는 설문서를 회수하여, 과제별로 예측결과를 집계하고 통계요약표를 작성한다. 이 요약표는 라운드 2의 설문서에 포함된다.

라운드 2에서 패널은 라운드 1에서 응답했던 자신의 예측결과를 수정할 기회를 갖는다. 이때 자신의 응답을 수정하지 않을 수 있으며, 자신의 응답이 IQR 밖에 위치할 경우(outliers), 그 이유를 제시해야 한다(Martino, 1993). 총괄수행자는 라운드 1과 2의 예측치 중위수·IQR, 그리고 제반 주장과 의견을 정리한다(홍순기·오정묵, 1997).

Ⅲ. 델파이의 타당성과 신뢰성

예측에서의 타당성은 예측한 사건(event)이 실현된 이후에 그 정확도를 평가하는 것이고, 신뢰성은 예측시점에서 예측결과의 정밀도(precision)를 측정하는 것이다. 예측에서 발생하는 오차는 체계적 오차(systematic error)와 비체계적 오차(random error)로 구분된다. 체계적 오차는 항상 일정한 방향으로 작용하는 편의(bias)이다. 반면 비체계적인 오차는 무작위적으로 발생하는 오차이다. 타당성은 체계적 오차와, 신뢰성은 비체계적 오차와 관련된다(채서일, 1991).

1. 타당성

델파이연구의 주요 관심사 중 하나는 그 결과가 얼마나 정확한가, 즉 타당성에 관한 것이다(Rowe and Wright, 1999). 타당성은 실제 측정결과가 대상의 실제 평균을 적절하게(adequately) 반영하는 정도이다(Babbie, 1989).

Ashton(1985)은 평균 절대 오차(mean absolute difference)로 정확도를 측정했다. Ono and Wedemeyer(1994)는 하와이의 미래에 대한 24개의 추세와 17개 사건에 대한 델파이 예측 결과의 타당성을 평가했다. 그 결과 절반 이상이 정확히 예측되었다는 것을 확인하여, 델파이는 장기예측에서 타당한 기법이라는 것을 보였다. Parenté et al.(2005)은 델파이의 반복조사와 피드백이 예측의 정확도에 미치는 영향을 검토했다. Rowe et al.(2005)은 델파이 라운드가 후반으로 갈수록 예측의 정확도가 높아진다는 것을 보였다.

델파이의 타당성은 내적 타당성과 외적 타당성으로 구분할 수 있다(Zolingen and Klaassen, 2003). 내적 타당성은 측정된 결과가 변수의 변화 때문에 일어난 것인가에 관한

문제이며(채서일, 1991), 그 방법론 자체가 바람직한가에 관한 것이다(Woudenberg, 1991).

델파이의 외적 타당성은 ‘평가기준 중심 타당성(criterion-oriented validity)’이라고 불리는데, 미래에 대한 판단과 실측치 간의 유사성에 근거한다(Zolingen & Klaassen, 2003). 즉, 외적타당성은 판단과 실제 값 간의 일치(correspondence) 정도를 나타내는 정확도를 의미한다. 본 연구의 타당성은 정확도 개념을 이용하여 분석하는, 델파이의 외적 타당성을 의미한다.

타당성 분석은 정확도를 평가하는 작업이며, 이는 실현시기(R_i)와 예측시기(M_i)의 차이에 의해 평가된다.¹⁾ 기술실현의 예측시기가 실제로 실현된 시기에 근접한다면, 그 예측조사는 정확하다고 볼 수 있다. 이 때 실현시기가 예측시기보다 이르거나 늦다는 것보다는, 절대적인 차이가 중요한 의미를 지닌다. 따라서 절대예측오차($|R_i - M_i|$)로 정확도를 측정하는 것이 합리적이다. 그런데 실현시기가 먼 미래일수록 예측의 정확도는 감소한다. 이를 보정하기 위해 절대예측오차를 예측거리²⁾($R_i - F$)로 나눈 비율을 사용한다(F 는 예측시행연도). 이 비율이 낮을수록 예측결과는 정확해지게 된다. 따라서 본 연구에서는 ‘1’에서 이를 제한 값을 정확도(A_i)로 정의한다.

$$A_i = 1 - \frac{|R_i - M_i|}{R_i - F}$$

여기서, i 는 과제번호

2. 신뢰성

예측의 신뢰성이란 동일한 개념에 대해 동일한 조건에서 측정을 반복(replication)했을 때 동일한 측정값을 얻을 가능성을 말한다(Babbie, 1989; Woudenberg, 1991; Kastein et al., 1993; Zolingen and Klaassen, 2003). 윤윤중·이종일(1998)은 예측의 신뢰성을 예측결과에 조직적인 편향(systematic bias)이 존재하지 않고, 응답자들 사이에 합의가 이루어져 있어 예측결과를 의사결정자료로 사용가능한 상태로 정의했다.

-
- 1) 본 연구에서는 기술과제에 대한 예측치의 중위수(median)를 이하 ‘예측시기’라고 하고, 그 기술과제가 실제로 실현된 시기를 이하 ‘실현시기’라고 명명한다.
 - 2) 예측거리는 실현시기에서 예측수행시점을 제한 값으로서, 장기과제 또는 단기과제로 구분하는 것처럼, 예측과제가 실현되기까지의 기간을 의미한다.

Ament(1970)는 1964년과 1969년의 상이한 패널을 대상으로 기술예측을 실시한 결과, 두 예측결과가 일관성이 있다는 것을 발견했다. Jillson(1975)은 델파이 연구의 질을 높이기 위한 지침을 만들어서 델파이의 신뢰성을 제고하고자 했다. Woudenberg(1991)는 동일한 델파이에서 참가자를 두 그룹으로 나누어 그 결과를 비교하여 신뢰성을 분석했다. Kastein et al.(1993)은 신뢰성 분석을 위해, ICC³⁾(Intraclass Correlation Coefficient)를 이용했다.

예측의 신뢰성은 정밀도의 측정을 통해 판단할 수 있으며, 정밀도는 패널의 합의 정도로 측정할 수 있다(Hill and Fowles, 1975; Ashton, 1985; Woudenberg, 1991). 그러나 델파이예측에서 합의 즉 수렴도만으로 신뢰성을 판단하기에는 부족한 점이 있기 때문에 추가적인 고려가 필요하다.

실현시기가 먼 미래일수록 예측의 수렴도가 낮아지는 것이 일반적인 현상이다(홍순기·오정목, 1997). Gordon and Helmer(1964)은 실현시기의 거리(remoteness)와 IQR의 크기 간의 높은 정(+)의 관계가 있음을 확인했다. Martino(1970)는 예측의 거리(length of the forecast)와 산포의 정도(degree of dispersion) 간에 높은 상관관계가 있음을 증명했다. 따라서 장기예측과 단기예측의 수렴도 즉 IQR만을 직접 비교하여 평가하는 것은 적절하지 않다. 이를 확인하기 위해 다음과 같이 가설을 설정한다.

H1: 델파이예측결과의 중위수와 IQR은 정(+)의 상관관계가 있다.

홍순기·오정목(1997)은 IQR을 예측수행 시점에서 예측시기의 중위수까지의 기간으로 나눈 비율로 수렴도를 정의하고 이 비율을 예측의 신뢰성에 대신했다. Landeta(2006)는 합의의 정도를 RIR(Relative Interquartile Range)로 측정했다. RIR은 IQR을 평균으로 나눈 백분율이다.⁴⁾ 델파이 결과의 대표값은 예측시기의 중위수로 주어지므로, 본 연구에서는 홍순기·오정목(1997)의 비율을 신뢰성 측정에 이용한다. 그런데 이 비율이 낮을수록 수렴도는 증가한다. 따라서 '1'에서 이 비율을 제한 값을 정밀도로 정의한다. 즉, i 번째 과제의 사분위수범위(IQR_i)를 예측거리($M_i - F$)⁵⁾로 나누고, '1'에서 이를 제한 값

3) $ICC = \frac{\text{mean square effect}}{\text{mean square effect} + \text{mean square error}}$

4) $RIR = \frac{Q_U - Q_L}{\text{Mean}} \times 100$

5) 신뢰성 측정은 타당성 측정과는 달리, 과제실현 이전에도 이루어질 수 있으므로, 예측거리의 측정에서 실현시기(R_i) 대신 예측시기(M_i)를 사용한다.

을 예측의 정밀도(P_i)로 정의한다.

$$P_i = 1 - \frac{IQR_i}{M_i - F}$$

여기서, i 는 과제번호

3. 타당성과 신뢰성의 관계

델파이예측에서는 응답자의 전문도에 따라 조사결과의 타당성과 신뢰성이 달라질 수 있다($H2$). 또한 기술 분야별로 타당성과 신뢰성이 상이할 수 있다($H3$). 이를 규명하기 위해 다음과 같이 가설을 설정한다.

$H2$: 델파이예측결과의 정확도와 정밀도는 응답자의 전문도에 따라 차이가 있다.

$H3$: 델파이예측결과의 정확도와 정밀도는 조사대상 분야에 따라 차이가 있다.

델파이연구의 대부분은 정밀도에 따라 정확도가 증가한다고 가정한다(Parenté et al., 2005). 델파이예측결과의 정밀도와 정확도 간에 상관관계가 있다면, 예측수행시점에서 정밀도를 측정함으로써 정확도를 어느 정도 파악할 수 있다는 이점이 있다. 이에 본 연구에서는 타당성과 신뢰성의 관계를 규명하고자 다음과 같이 네 번째 가설을 설정한다.

$H4$: 델파이예측결과의 정확도와 정밀도는 정(+)의 상관관계가 있다.

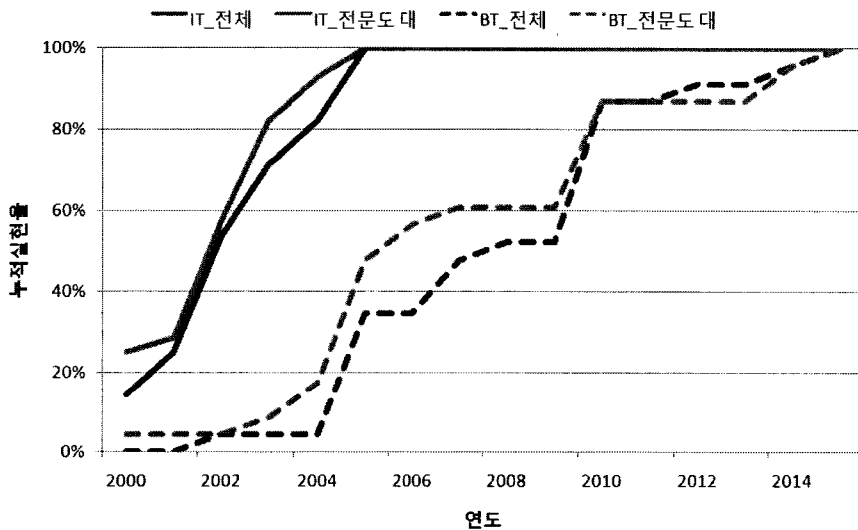
IV. 타당성과 신뢰성 분석

델파이예측결과의 신뢰성은 어떤 조사결과가 부정확한 자료에서 우연히 발견된 것이 아니라 확실한 확신을 제공한다. 반면, 타당성은 델파이예측결과가 이론적·실무적으로 의미가 있는지와 관련이 있다. 본 장에서는 델파이예측결과의 타당성과 신뢰성을 분석하고 그 관계를 살펴본다.

1. 자료 수집

본 연구에서는 15개 분야 중에서 ‘정보·전자·통신’과 ‘생명공학’을 1차로 선정했다. 그 중에서도 ‘정보·전자·통신’ 분야의 ‘통신(이하 IT)’, 그리고 ‘생명공학’ 분야의 ‘개체(이하 BT)’를 대상으로 한다. IT와 BT에 속한 과제는 각각 28개와 23개 과제이다. <그림 2>의 예측시기 분포를 보면, 전체적으로 단기과제가 상대적으로 많으며, IT 분야에서 더욱 그렇다.

<그림 2> IT와 BT의 전문도별 예측시기



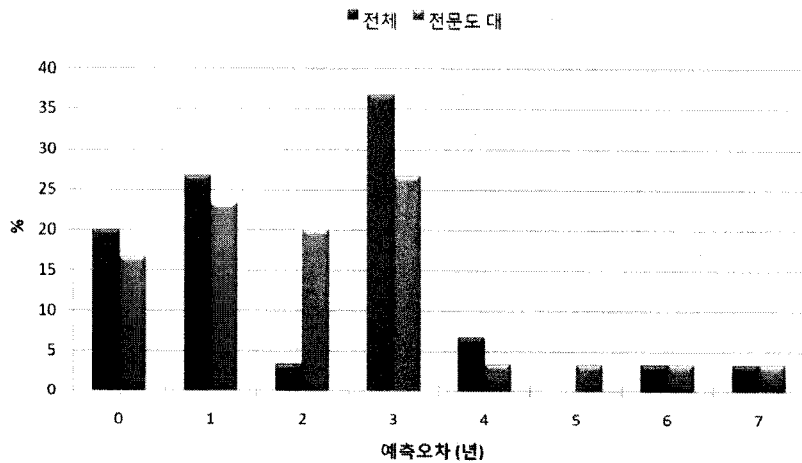
각 과제에 대해 ‘국내 예측시기’와 ‘세계 예측시기’에 대한 ‘1차 집계’와 ‘2차 집계’ 결과가 있고, ‘전체’와 ‘전문도 대(大)’인 경우의 집계결과가 있다. 본 연구에서는 ‘국내 예측시기’에 대한 ‘2차 집계’ 결과를 IT와 BT에 대해 ‘전체’와 ‘전문도 대’로 구분하여 비교분석한다.

2. 타당성과 신뢰성 분석

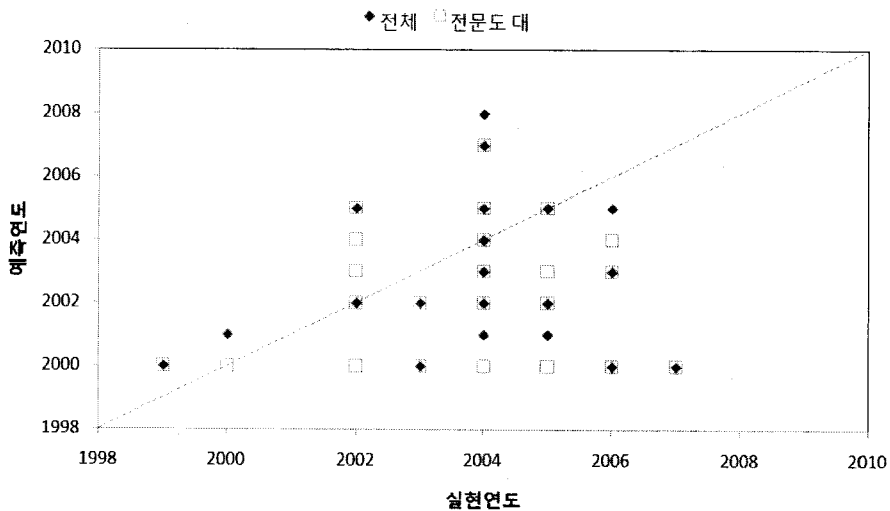
기술과제의 실현시기 판단을 위해 30명의 관련 전문가에게 자문을 구했다. 이 전문가들은 IT 관련 학과 교수 15명, 그리고 BT 관련 학과 교수 15명으로 구성된다. 실현시기

에 대한 응답을 통합하기 위한 대표치로는 최빈수(mode)를 사용했다. 그 결과, 예측오차6)는 대부분 3년 이내로 나타났으며(<그림 3>), 절반 이상의 과제가 예측보다 늦게 실현되었다(<그림 4>). IT 분야의 경우 대부분의 과제가 예측보다 늦게 실현된 반면, BT 분야의 경우 예측보다 빨리 실현되었다. IT와 BT 분야에서 아직 실현되지 않았다고 밝혀진 과제는 각각 6개와 13개였다.

<그림 3> 예측오차의 분포



<그림 4> 예측시기와 실현시기 비교



6) 예측오차 = |실현시기 - 예측시기|

타당성과 신뢰성 분석을 위해 분석대상과제에 대해 ‘전체’와 ‘전문도 대’인 응답결과 각각에 대해 정확도⁷⁾와 정밀도 값을 구했다(<부록>). 정밀도를 구할 때, 예측수행시점은 「제1회 과학기술예측조사」 보고서 발간 시점인 1994년이 아닌, 실제로 예측조사가 이루어진 1993년으로 설정했다.

타당성 분석 결과, IT와 BT 분야, 그리고 ‘전체’와 ‘전문도 대’인 응답결과의 정확도 간에는 큰 차이를 보이지 않았다(<표 1>). 이는 ‘전문도 대’라고 응답한 응답자가 그렇지 않은 응답자에 비해 전문도가 높지 않을 수 있음을 나타낸다. 다시 말해서, 델파이에서 전문도를 응답자 스스로 자가평가(self-rating)하는 방식이 응답자의 전문도를 나타내는 데 부적합할 수 있음을 시사한다.

신뢰성 분석 결과, BT가 IT 분야보다 예측의 정밀도가 높았다(<표 1>). BT 분야의 IQR 평균이 IT 분야보다 높음에도 불구하고, 이러한 결과가 나온 것은 BT 분야의 경우 IT 분야보다 장기과제가 많았기 때문인 것으로 분석된다.

IT와 BT 분야 모두 ‘전문도 대’인 응답결과의 정밀도가 ‘전체’ 응답결과의 경우보다 낮게 나타났다. 이는 전문도가 높은 응답자 간에 상당한 이견을 보임을 의미한다. 다시 말해서, 전문도가 높은 응답자는 자신의 의견에 확신을 갖는 경향이, 전문도가 낮은 응답자는 다수에 추종하는 경향이 있다는 것으로 해석할 수 있다(홍순기·오정목, 1997; 윤윤중·이종일, 1998).

<표 1> 예측결과의 타당성과 신뢰성

구분	정확도 평균(표준편차)			정밀도 평균(표준편차)		
	IT	BT	IT+BT ⁷⁾	IT	BT	IT+BT ⁷⁾
전체	0.819(0.147)	0.777(0.133)	0.805(0.144)	0.551(0.110)	0.606(0.126)	0.576(0.121)
전문도 대	0.804(0.154)	0.808(0.093)	0.806(0.137)	0.359(0.186)	0.556(0.215)	0.448(0.222)

주: IT와 BT 전체 분야

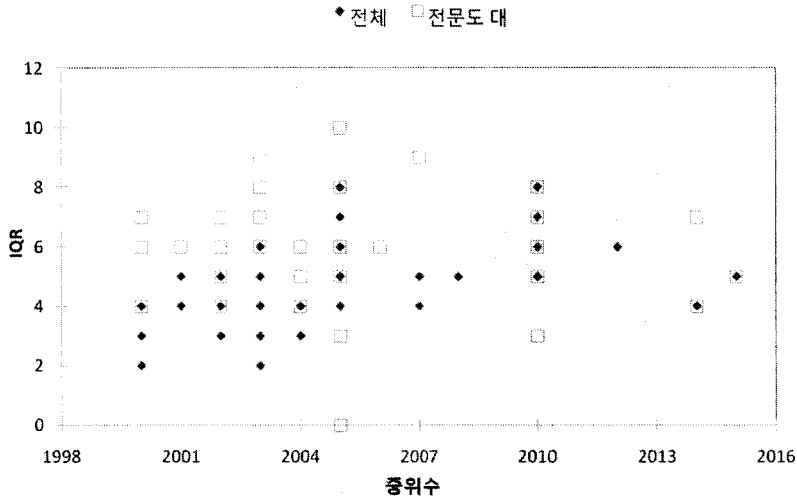
3. 가설 검증

연구가설 H_1 을 검정하기 위해 델파이예측결과의 중위수와 IQR 간의 상관계수를 구했다. ‘전체’ 응답의 경우 상관계수가 0.408로 통계적으로 유의했으나, ‘전문도 대’ 응답의

7) 타당성 분석 시 미실현 과제의 경우 정확도를 계산할 수 없으므로, 미실현 과제는 정확도 분석에서 제외한다.

경우는 유의하지 않아(<그림 5>), 전문도가 높은 응답자의 합의정도는 예측거리에 무관한 것으로 분석된다.

<그림 5> 예측의 정밀도



연구가설 H_2 과 H_3 에 대해 t 검정을 실시했다. 동일 분야 내 상이한 전문도 간의 집계 결과를 비교(H_2)할 때는 ‘대응표본 t 검정(paired sample t -test)’을, 분야 간 비교(H_3) 시엔 ‘두 표본 t 검정(two sample t -test)’을 실시했다.

정확도와 정밀도, 그리고 IT와 BT 분야 각각에 대해 ‘전체’와 ‘전문도 대’인 집계결과 간의 ‘대응표본 t 검정’을 실시했다(<표 2>). IT 분야의 정밀도는 전문도에 따라 통계적으로 유의한 차이를 보였으나, 나머지는 그렇지 않았다.

<표 2> ‘전체’와 ‘전문도 대’인 집계결과 간 ‘대응표본 t 검정’ 결과

구분		자유도	t	유의확률
정확도	IT	19	1.08	0.292
	BT	9	-0.75	0.470
정밀도	IT	27	5.85	0.000
	BT	22	1.30	0.206

‘두 표본 t 검정’은 두 모집단 분산의 동일성 여부에 따라 검정방법이 다르기 때문에, ‘두 표본 F 검정(two sample F -test)’을 네 가지 경우8)에 대해 먼저 실시했다. 그 결과

‘전체’ 집계결과에 대한 정밀도는 분야에 따라 분산이 다르다고 판단되었으나, 나머지는 그렇지 않았다.

따라서 ‘전체’의 정밀도의 경우 분야 간에 ‘이분산($\sigma_1^2 \neq \sigma_2^2$) 가정 두 표본 t 검정’을, 나머지 세 경우는 ‘등분산($\sigma_1^2 = \sigma_2^2$) 가정 두 표본 t 검정’을 실시했다(<표 3>). ‘전문도 대’인 집계결과의 정밀도는 분야에 따라 통계적으로 유의한 차이를 보였으나, 나머지는 그렇지 않았다.

<표 3> IT와 BT 분야 간 ‘두 표본 t 검정’ 결과

구분		자유도	t	유의확률
정확도	전체	28	0.732	0.470
	전문도 대	28	-0.061	0.951
정밀도	전체	44	-1.609	0.115
	전문도 대	49	-3.440	0.001

연구가설 $H4$ 을 검증하기 위해 상관분석을 실시했다(<표 4>). 그 결과 ‘전문도 대’인 응답의 경우 IT와 전 분야⁸⁾의 상관계수가 각각 0.630과 0.540으로 통계적으로 유의했다. 나머지는 대체로 정(+)의 상관관계를 보였으나, 통계적으로 유의하지는 않았다(<그림 6>). 이로 보아 전문도가 높은 응답자 집단의 경우 정밀도와 정확도 간에 상관관계가 있다고 할 수 있다.

<표 4> 타당성과 신뢰성 간 Pearson 상관계수

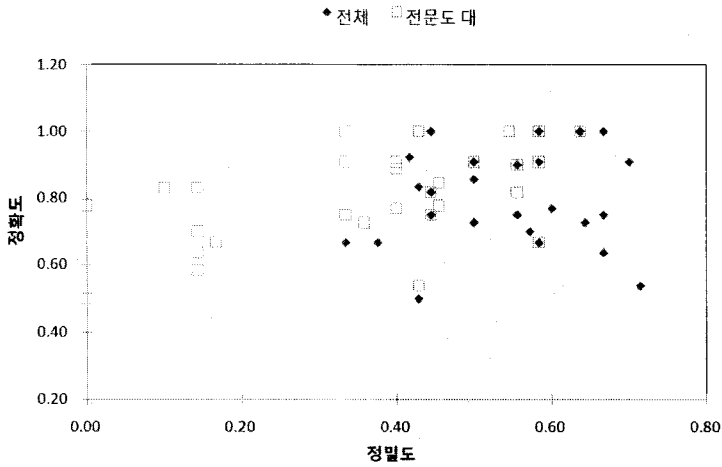
구분		상관계수	유의확률
전체	IT	0.216	0.360
	BT	-0.167	0.644
	IT+BT ⁹⁾	0.104	0.583
전문도 대	IT	0.630	0.003
	BT	0.278	0.437
	IT+BT ⁹⁾	0.540	0.002

주) IT와 BT 전체 분야

8) <표 3>의 구분 참조.

9) IT와 BT 전체 분야.

<그림 6> IT와 BT 전체 분야의 정확도와 정밀도



4. 선행연구 비교분석

지금까지 텔파이의 타당성 또는 신뢰성을 분석한 연구는 다수 있었으나(Martino, 1970; Woudenberg, 1991; Kastein et al., 1993; One and Wedemeyer, 1994; Landeta, 2006 등), 이 연구들은 타당성 또는 신뢰성만을 분석하여, 그 관계를 파악할 수 없었다. 우리나라의 경우 홍순기·오정목(1997)은 「제1회 과학기술예측조사(1994)」를, 그리고 윤윤중·이종일(1998)은 「산업기술예측(1998)」을 대상으로 텔파이의 신뢰성을 분석했다. 장기 예측에 주로 활용되는 텔파이의 특성상, 이 연구의 시점에서 예측결과의 실현여부를 평가, 즉 타당성을 분석하는 것은 불가능했다. 본 연구에서는 텔파이의 타당성과 신뢰성을 모두 검토했으며, 그 관계를 분석했다.

홍순기·오정목(1997)은 정보통신기술분야 내의 12분야, 172과제 중 한국·일본·독일·프랑스의 비교분석이 가능하고, 출현시기의 중위수가 2010년 이내의 과제를 중심으로 중요도가 높고 파급효과가 큰 10분야, 31과제를 선정하여 신뢰성을 분석했다. 윤윤중·이종일(1998)은 전자·정보통신 분야 등 4개 분야의 103개 과제에 대해 신뢰성을 분석했다.

홍순기·오정목(1997)과 윤윤중·이종일(1998)은 텔파이의 타당성은 다루지 않았으나 신뢰성은 분석했으므로, 이를 본 연구의 신뢰성 분석결과와 비교해 보는 것은 의미가 있다. 특히, 홍순기·오정목(1997)의 분석대상은 본 연구와 동일하므로 더욱 그러하다. 단, 이 연구들의 신뢰성 측정방법은 본 연구와 상이하므로, 비교분석을 위해 이 연구들의 신

뢰성, 즉 정밀도 값을 본 연구의 산출식을 이용해 변환했다.

<표 5>는 각 연구에서 분석한 정밀도를 보여준다. 한국의 델파이조사 결과가 다른 국가의 결과보다 신뢰성이 높게 나타났으며, 「산업기술예측(1998)」의 결과가 「제1회 과학기술예측조사(1994)」의 결과보다 신뢰성이 높은 것으로 나타났다. 즉 한국, 특히 「산업기술예측(1998)」에서 응답자의 합의유도(consensus building)가 성공적이었다고 할 수 있다.

<표 5> 신뢰성 분석결과 비교

연구	분석대상				정밀도			
					전체 ^{주)}		전문도 대	
	조사 국가	조사명	분야	과제수	평균	표준편차	평균	표준편차
윤윤중 · 이종일 (1998)	한국	산업기술예측(1998)	전자 · 정보통신(A)	34	0.632	0.134	0.430	0.331
			생물 · 정밀화학(B)	23	0.659	0.145	0.621	0.168
			환경 · 중전기기(C)	19	0.629	0.148	0.540	0.175
			소재 · 생산 · 자동화(D)	27	0.619	0.150	0.523	0.133
			A+B+C+D	103	0.634	0.141	0.512	0.241
본 연구		제1회 과학기술예측조사 (1994)	통신(E)	28	0.551	0.110	0.359	0.186
			개체(F)	23	0.606	0.126	0.556	0.215
			E+F	51	0.576	0.121	0.448	0.222
			정보 · 전자 · 통신	31	0.522	0.335	0.309	0.485
홍순기 · 오정목 (1997)	일본	제5회 과학기술예측조사 (1993)	정보통신	31	0.379	0.335	0.256	0.464
	독일	과학기술에 대한 독일의 델파이 보고서(1993)	정보통신	31	0.177	0.522	0.034	0.936
	프랑스	과학기술에 대한 델파이 조사 보고서(1994)	정보통신	31	0.313	0.464	0.187	0.667

주) 산업기술예측의 경우, '전체' 응답결과가 아니라, 전문도가 중(中)과 소(小)인 응답자의 응답결과임.

<표 5>의 모든 결과에서 전문도가 높은 응답자의 응답결과가 상대적으로 신뢰성이 낮은 것으로 나타났다. 이는 '전체'와 '전문도 대' 응답 중 어느 결과를 델파이 예측의 최종결과로 활용할 것인지를 선택할 때 신중을 기해야 함을 의미한다. 홍순기 · 오정목 (1997)은 전문도가 높은 전문가 응답자의 신뢰도가 예상 외로 낮게 나온 이유 중 하나로, 전문가에게 배경정보가 미리 제공되지 않은 것을 들고 있다. 즉, 기술발전과정은 정치 · 경제적 환경 등에 의해 영향을 받는데, 전문가 패널은 과학기술자로만 구성되었으므로,

이러한 환경에 대한 지식이 부족하다는 것이다.

V. 결 론

본 연구에서는 델파이의 개념과 특성을 검토하고, 한국의 「제1회 과학기술예측조사」의 IT와 BT 분야를 대상으로 델파이 기술예측의 타당성과 신뢰성을 분석했다. 또한, 정확도와 정밀도의 전문도와 분야에 따른 차이와 상관관계를 살펴보았다. 본 연구에서 타당성은 예측오차 즉 정확도를, 신뢰성은 예측결과의 수렴도 즉 정밀도를 의미한다. 이 정확도와 정밀도의 개념에는 예측거리가 반영되었다.

델파이의 3가지 특성은 패널리스트(panelist) 간의 효과적인 상호작용을 가능하게 한다. 여러 델파이실험에서 패널리스트가 다른 패널리스트의 의견에 반응하여 자기 의견을 수정함을 보였다. 그러나 델파이는 ‘다수 의견에 편승(bandwagon effect)’이라는 문제점을 여전히 안고 있다. 본 연구의 결과에서는 전문도가 높은 경우 이 문제점이 덜 발견되었다.

타당성과 신뢰성은 전문도와 분야별로 일부 유의한 차이를 보였다. BT 분야의 결과는 대체적으로 통계적으로 유의하지 않았다. BT 분야에 장기과제가 많아 샘플수가 적었던 것이 한 가지 원인으로 분석된다. IT 분야의 경우, 그 정밀도는 전문도가 상이한 응답결과 간에 유의한 차이를 보였으나 정확도는 그렇지 않았다. 또한 ‘전문도 대’인 응답결과보다 ‘전체’ 응답결과의 정밀도가 높게 나타났다. 정확도와 정밀도 간의 상관분석 결과 IT와 전 분야의 전문도가 높은 응답결과는 통계적으로 유의한 정(+)의 상관관계를 보였다. 다시 말해서, ‘전체’ 응답결과와 비교할 때 ‘전문도 대’인 경우 그 정밀도는 낮음에도 불구하고 정확도는 비슷했다. 그런데 ‘전문도 대’인 경우에만 유의한 정(+)의 상관관계를 보였다. 이는 전문도가 높은 응답자는 잘못된 다수로 추종할 가능성이 낮으며, 그들의 합의 정도가 높으면 그 결과가 타당할 확률이 높다는 것을 의미한다.

본 연구에서 델파이의 정확도와 정밀도 간의 상관관계는 일부 유의하지 않게 나타났다. 따라서 다양한 분야의 충분한 샘플에 대해 정확도와 정밀도를 분석하는 후속연구가 요구된다. 일본의 경우처럼, 과학기술예측조사 시 과거 예측과제의 실험도 함께 조사하는 것도 하나의 방법이다. 정확도와 정밀도 간의 명확한 상관관계가 확인되면, 두 변수관계의 모형화가 가능하다. 즉, 전문도가 높은 응답자의 응답치에 높은 가중치를 부여하

여, 조사결과의 정밀도를 예측결과의 정확도 예상을 위한 하나의 판단기준으로서 사용할 수 있다. 단, 이 모형의 적용을 위해서는 신뢰성의 정도를 판단할 수 있는 정밀도의 범주화가 필요하다.

본 연구는 델파이 기술예측조사에 다음과 같은 시사점을 제공한다.

첫째, 델파이예측의 정밀도를 높이기 위한 노력이 필요하다. 델파이예측조사의 타당성과 신뢰성 간의 상관관계가 높은 경우, 높은 정밀도는 정확도의 기대값을 높인다. 사전에 응답자에게 시나리오 형태로 배경정보를 제시하는 것은 신뢰성을 높이기 위한 하나의 대안이 될 수 있다.

둘째, 델파이조사 시 응답자 전문도의 객관적인 평가가 필요하다. 대부분의 델파이예측조사에서는 응답자가 전문도를 스스로 기입한다. 전문도를 객관적으로 측정할 수 있는 지표, 예를 들어 인용지수(citation index) 또는 계량서지학 지수(bibliometric index)를 도입하여 전문도의 객관성을 확보하는 것이 필요하다.

참고문헌

- 신태영 · 박재혁 · 정근하 · 김형수 (1994), 「제1회 과학기술예측조사(1995~2015년): 한국의 미래 기술」, 서울: 과학기술정책연구원.
- 윤윤중 · 이종일 (1998), “델파이방법을 이용한 기술예측의 신뢰도 분석”, 「기술혁신학회지」, 제1권 제2호, pp. 275-284.
- 일본 문부과학성 과학기술정책연구소 · 미래공학연구소 (2006), 「2035년의 과학기술」, 대전: 한국 과학기술정보연구원 역.
- 채서일 (1991), 「사회과학조사방법론」, 2판, 파주: 학현사.
- 홍순기 · 신흥식 · 박수동 (2007), 「기술예측」, 서울: 한국산업기술재단.
- 홍순기 · 오정목 (1997), “정보통신분야의 델파이 기술예측 국제비교분석 - 한국 · 일본 · 프랑스 · 독일-”, 「기술혁신연구」, 제5권, 제1호, pp. 223-248.
- Ament R. H. (1970), “Comparison of Delphi Forecasting Studies in 1964 and 1969”, *Futures*, Vol. 2, No. 1, pp. 35-44.
- Ashton, A. H. (1985), “Does Consensus Imply Accuracy in Accounting Studies of Decision making?”, *Accounting Review*, Vol. 60, No. 2, pp. 173-185.
- Babbie, E. R. (1989), *The Practice of Social Research*, 5th ed., Belmont: Wadsworth Publishing Company.
- Bright, J. R. (1978), *Practical Technology Forecasting: Concepts and Exercises*, Texas: Industrial Management Center.
- Gordon, T. J. and Helmer, O. (1964), *Report on A Long Range Forecasting Study*, California: RAND, Paper p. 2982.
- Hill, K. Q. and Fowles, J. (1975), “The Methodological Worth of the Delphi Forecasting Technique”, *Technological Forecasting and Social Change*, Vol. 7, No. 2, pp. 179-192.
- Jillson, I. A. (1975), “Developing Guidelines for the Delphi Method”, *Technological Forecasting and Social Change*, Vol. 7, No. 2, pp. 221-222.
- Kastein, M. R., M. Jacobs, R. Hell, K. Luttkik, and F. Touw-Ottenk (1993), “Delphi, the Issue of Reliability: A Qualitative Delphi Study in Primary Health Care in the Netherlands”, *Technological Forecasting and Social Change*, Vol. 44, No. 3, pp. 315-323.
- Landeta, J. (2006), “Current Validity of the Delphi Method in Social Sciences”, *Technological Forecasting and Social Change*, Vol. 73, No. 5, pp. 467-482.
- Martino, J. P. (1970), “The Precision of Delphi Estimates”, *Technological Forecasting and Social Change*, Vol. 1, No. 3, pp. 293-299.

- Martino, J. P. (1993), *Technological Forecasting for Decision Making*, 3rd ed., Columbus: McGraw-Hill.
- Ono R. and D. J. Wedemeyer (1994), "Assessing the Validity of the Delphi Technique", *Futures*, Vol. 26, No. 3, pp. 289-304.
- Parenté, R., ìl N. Hiöb, T. Silver, R. A. Jenkins, C. Poe, and R. J. Mullins (2005), "The Delphi Method, Impeachment and Terrorism: Accuracies of Short-range Forecasts for Volatile World Events", *Technological Forecasting and Social Change*, Vol. 72, No. 4, pp. 401-411.
- Rowe, G. and G. Wright (1999), "The Delphi Technique as a Forecasting Tool: Issues and Analysis", *International Journal of Forecasting*, Vol. 15, No. 4, pp. 353-375.
- Rowe, G., G. Wright and A. McColl (2005), "Judgment Change during Delphi-like Procedures: The Role of Majority Influence, Expertise, and Confidence", *Technological Forecasting and Social Change*, Vol. 72, No. 4, pp. 377-399.
- Woudenberg, F. (1991), "An Evaluation of Delphi", *Technological Forecasting and Social Change*, Vol. 40, No. 2, pp. 131-150.
- Zolingen, S. J. and C. A. Klaassen (2003), "Selection Processes in a Delphi Study about Key Qualifications in Senior Secondary Vocational Education", *Technological Forecasting and Social Change*, Vol. 70, No. 4, pp. 317-340.

□ 투고일: 2008. 10. 29 / 수정일: 2009. 03. 16 / 게재확정일: 2009. 04. 27

A Study on the Analysis of Validity and Reliability of the Delphi Forecasting in Korea

Seong-hoon Gwon · Soon-ki Hong

Abstract : The Delphi is a popular technique for forecasting based on the opinions of experts. It is important to know how valid and reliable the technique is. In this paper, we analyze accuracy and precision of the Delphi in IT and BT of Korea, and also discuss the relationship between them. As a result of the analysis, the accuracy and precision of the forecasts partly have significant differences according to their area and degree of expertise. Besides, significant correlation between the accuracy and precision of forecasts with high expertise is found. The result indicates that the precision of forecasts can be a criterion of the accuracy of them.

Key Words : Delphi, Validity, Reliability, Precision, Accuracy

<부록> 타당성과 신뢰성 측정 결과

분야	과제명	'전체' 응답				'전문도 대'인 응답				실현 연도
		중위수	IQR	정확도	정밀도	중위수	IQR	정확도	정밀도	
통신 (IT)	전화망에 음성인식 기술이 적용된 여러가지 부가서비스가 실용화된다.	2000	4	0.83	0.43	2000	6	0.83	0.14	1999
	소형·장수명의 충전용이한 휴대 전화(용적 50cc, 통화가능 시간 3시 간)가 실용화된다.	2000	2	0.54	0.71	2000	4	0.54	0.43	2006
	수백미터내의 소형셀에 의한 개 인용 무선통신망이 구축된다.	2002	5	-	0.44	2001	6	-	0.25	미실현
	DAB(digital audio broadcasting) 방송 시스템이 개발된다.	2001	5	0.67	0.38	2000	6	0.58	0.14	2005
	HDTV 용 방송장비(HD-VTR, HD- CAMERA)가 개발된다.	2000	3	0.70	0.57	2000	6	0.70	0.14	2003
	Data 압축기술을 이용하여 300채 널 이상의 프로그램 전송이 가능한 차세대 CATV가 실용화된다.	2003	4	0.77	0.60	2003	6	0.77	0.40	2006
	고해상도 텔레비전(HDTV)이 방 송, 인쇄, 통신 등의 분야에도 보급 된다.	2002	4	0.75	0.56	2002	5	0.75	0.44	2005
	소형 개인 휴대용 컴퓨터와 인공위 성간의 통신에 의한 정보의 전달이 실용화된다.	2003	5	0.91	0.50	2003	6	0.91	0.40	2004
	광신호를 전기신호로 변환하지 않 고 광 그대로 교환할 수 있는 광교 환기가 실용화된다.	2005	5	-	0.58	2005	3	-	0.75	미실현
	광대역 ISDN의 실현으로 고속 통 신 및 다중 서비스가 실용화된다.	2002	5	0.75	0.44	2002	6	0.75	0.33	2005
	방송위성을 통해 디지털 HDTV방 송이 실용화된다.	2002	5	1.00	0.44	2002	6	1.00	0.33	2002
	실시간으로 동작하는 한·영 자동 통역전화가 개발된다.	2005	5	-	0.58	2003	8	-	0.20	미실현
	지문, 필적, 음성, 표정 등의 특징을 사용하여 개인을 식별하는 보안 통 신 시스템이 실용화된다.	2002	4	0.90	0.56	2002	4	0.90	0.56	2003
	개인 ID코드를 사용해서 언제 어디 서나 자유롭게 통신할 수 있는 추 적접속교환이 실용화된다.	2004	3	-	0.73	2004	4	-	0.64	미실현
	개인대응의 전화번호가 실현되어 세계 어디에서도 희망하는 개인과 통신을 할 수 있는 퍼스널 이동통 신이 실용화된다.	2005	4	1.00	0.67	2005	5	1.00	0.58	2005
	구급차와 병원 사이에 구급의료용 화상 및 지식전송 네트워크 기술이 보급된다.	2003	2	-	0.80	2003	7	-	0.30	미실현
∴	∴	∴	∴	∴	∴	∴	∴	∴	∴	∴

분야	과제명	'전체' 응답				'전문도 대'인 응답				실현 연도
		중위수	IQR	정확도	정밀도	중위수	IQR	정확도	정밀도	
개체 (BT)	고등생물 염색체의 유전자 지도가 평균 10centimorgen 해상도로 작성된다.	2010	5	-	0.71	2014	4	-	0.81	미실현
	염색체 부분 염기서열을 바탕으로 개체의 유전적 특징이 규명된다.	2010	5	-	0.71	2010	3	-	0.82	미실현
	잔류농약 등 공해 산물에 대해 분해력이 우수한 미생물이 선발되어 활용된다.	2005	5	0.67	0.58	2005	10	0.67	0.17	2002
	식물 생육증진 효과가 우수한 vamyco-rhizal fungi가 개발된다.	2010	6	-	0.65	2006	6	-	0.54	미실현
	환경 stress에 내구성 있는 식물이 개발된다.	2008	5	0.64	0.67	2007	9	0.73	0.36	2004
	내충·내병성 식물이 개발된다.	2005	8	0.67	0.33	2004	6	0.78	0.45	2002
	고기능 생물반응기를 이용한 생산 기술이 실용화된다.	2005	7	0.92	0.42	2004	6	0.85	0.45	2006
	천연물 또는 미생물로부터 무공해, 저독성 농약이 개발된다.	2005	8	0.67	0.33	2003	6	0.89	0.40	2002
	수정란 조작 및 핵 이식에 의한 클로닝 동물이 생산된다.	2005	5	0.91	0.58	2005	8	0.91	0.33	2004
	임목 육종을 위한 유전자 재조합 기술이 개발된다.	2005	6	0.91	0.50	2005	6	0.91	0.50	2004
	웅성 불임성을 이용한 hybrid 종자의 생산기술이 개발된다.	2005	5	0.67	0.58	2005	5	0.67	0.58	2002
	우리나라 고유의 생물자원에 관한 혈통 고정기술이 개발된다.	2007	4	-	0.71	2005	3	-	0.75	미실현
	질환 모델동물들이 개발되어 생산된다.	2007	5	0.73	0.64	2005	5	0.91	0.58	2004
	로봇에 의해 운전되는 생물공정이 실용화된다.	2007	5	-	0.64	2005	0	-	1.00	미실현
	발생, 분화의 분자기구의 개요가 해명된다.	2012	6	-	0.68	2010	5	-	0.71	미실현
	수목의 크기나 형을 결정하는 기구가 해명된다.	2014	4	-	0.81	2014	7	-	0.67	미실현
:	:	:	:	:	:	:	:	:	:	