

생물자원 전통지식 추출을 위한 델파이조사의 신뢰성 연구[†]

이기훈¹, 송미장², 김 현³

¹전주대학교 경영학과 · ²전주대학교 대체의학과 · ³전주대학교 생명자원융합과학과

접수 0000년 0월 1일, 수정 0000년 0월 0일, 게재확정 0000년 0월 0일

요 약

본 연구는 문헌 및 현장에서 얻은 생물자원지식들의 가치를 평가하기 위해 전문가들을 대상으로 실시한 델파이 조사 (Delphi method)의 신뢰성을 평가하였다. 델파이조사는 전문가들이 다른 사람의 의견과 관계없이 독립적으로 전문성 있는 평가를 내린 다음에 다음 단계에서 다른 전문가들의 의견을 참고하며 자신의 의견을 수정하는 절차를 갖는다. 본 연구에서는 문헌에서 얻은 전통지식 100건, 현장에서 취득한 지식 100건 등, 모두 200건을 우선 선정한 후 전문가 6인을 선정하여 각 지식의 가치에 대하여 두 차례 평가하도록 하였다. 그 결과 두 차례의 평가점수는 연관성이 매우 높으면서 2차에서는 다른 전문가들의 의견을 수용해 어느 정도 자체수정이 발생하여 각 문항에 대한 평가점수의 표준편차가 줄어들었다. 본 조사의 신뢰성 (reliability)을 파악하기 위해 일반적인 신뢰도 계수인 크론바하 알파와 함께 일반화가능도 (generalizability) 계수를 구하였다. 이 두 신뢰도 분석을 통해 2차 평가 후 평가의 신뢰도가 상승하여 전문가에 의한 델파이 조사의 신뢰도가 매우 높다는 사실을 지지하였으나 일반화 가능도 분석 결과를 해석하는 과정에서 다른 결과를 유추할 수 있었다. 신뢰도계수가 증가하였음에도 불구하고 평가자간의 편차는 증가하여 신뢰도가 높아진 것은 평가가 상향되고 평균에 회귀하는 경향으로 잔차변동이 줄어서이지 평가자간의 의견수렴이 이루어진 결과로 볼 수는 없다는 사실이었다. 이러한 결과를 토대로 신뢰도 계수와 함께 평가자 간의 분산을 파악하여 델파이조사의 추가적인 단계 (round)가 필요함을 제시하였다.

주요용어: 델파이조사, 생물자원전통지식, 신뢰성, 일반화가능도.

1. 서론

지식정보화 시대에서 지식재산권은 그 경제적 가치가 크고 새로운 성장동력으로 인식되기 때문에 이에 대한 발굴과 보호가 매우 중요하다. 그러한 지식재산 중에서 최근에는 국가가 보유한 전통지식에 대한 관심도 높아지고 이를 활용한 기술에 특허권을 부여하여 타국으로부터 보호하려는 움직임이 생겨나고 있다 (Kim과 Song, 2013; Bubela와 Gold, 2012). 국제적 흐름으로는 생물다양성협약 (Convention on Biological Diversity), 당사국회의 (Conference of the Parties), 나고야 의정서 등에 의해 생물자원에 대한 지식이 인류의 공동자원이 아닌 주권적 권리로 인정받게 되었고 이에 따른 권리 인정 및 이익의 공유 등이 규정되고 있다. 우리나라도 전통지식에 대한 발굴과 보호에 관심을 가지면서 지식재산권 제도의 보완과 전통지식의 DB 구축이 필요하게 되었다 (Ministry of Environment, 2010; Kim과 Song, 2011; Choi 등, 2014).

[†] 본 논문은 농촌진흥청 연구사업(세부과제번호: PJ009566032015)의 지원에 의해 이루어진 것임.

¹ (560-759) 전북 전주시 효자동 3가, 전주대학교 경영학과, 교수. E-mail: khlee@jj.ac.kr

² (560-759) 전북 전주시 효자동 3가, 전주대학교 생명자원융합과학과, 조교수.

³ (560-759) 전북 전주시 효자동 3가, 전주대학교 대체의학과, 교수.

전통지식을 발굴하고 이를 DB로 구축하기 위해서는 특정 전통지식의 가치를 평가하는 작업이 선행되어야 하므로 평가기준을 구축하는 작업도 필요하다. 우리나라의 전통지식을 평가하기 위해 한국지식재산연구원의 Ryu 등 (2012)은 산업적 활용성을 중심으로 전통지식에 대한 가치평가기준을 제안하였는데 이는 기술의 과학성, 유용성, 경제성, 활용성, 시장수준 등이다. 이러한 평가기준은 객관적 자료를 통해 계산할 수 있는 평가기준이지만 전문가가 정성평가할 시에 참조할 수 있는 정보로 활용하기를 권장하기 때문에 본 연구에서는 전통지식에 대한 전문가의 평가를 위해 전통지식의 가치를 실용성 (practicality), 경제적 가치(economic value), 산업화 가능성 (industrialization), 이용 가능성 (availability), 잠재적 가치 (potential value) 등의 특성으로 세분화였다.

본 연구에서는 고문헌에서 얻은 생물자원지식 100건, 현장에서 취득한 지식 100건 등, 모두 200건을 우선 선정한 후 전문가 평가에 의해서 유용한 생물전통지식을 선정하는 방법을 제안하였다. 이런 연구를 위한 사전 단계로 고문헌과 현장에서 추출한 전통지식을 전문가 패널에게 우수성, 실용성 등을 중심으로 평가하도록 하여 전문가 패널 5명 중 3명이 복수 추천한 생물전통지식을 기초로 하여 문헌에서 얻은 지식 100건, 현장에서 취득한 지식 100건 등, 모두 200건을 선정하였다. 이렇게 취합된 200건의 생물자원지식의 가치를 평가하기 위해 전문가 6인을 선정하여 각 지식에 대하여 델파이조사 (Delphi method)를 실시하였다. 1차 (first round)로 온라인을 통하여 전문위원에게 전통지식의 가치에 대한 특성별 평가를 의뢰하였고 그 결과를 취합하여 전체 전문가의 평균을 각 평가위원에게 알려주고 수정하는 2차 (second round) 평가를 받았다.

두 차례의 델파이조사가 적절하게 잘 이루어졌고 조사의 횟수가 충분한지를 파악하기 위해서 전문가 평가에 대한 신뢰도 분석을 행하고 통계적 검정을 실시하였다. 즉, 차수 (round)를 거듭할수록 델파이조사의 신뢰도는 증가하였는지를 파악하기 위하여 차수별 신뢰도 계수 (Cronbach alpha)와 일반화가능도 (generalizability)를 계산하였고, 또한 델파이조사의 성격과 목적에 부합하도록 전문가들 평가의 차수간 변화와 공감대 (consensus) 형성이 잘 이루어졌는지를 파악하고자 하였다.

본 연구의 2장에서는 델파이 조사와 신뢰성 측도에 대하여 소개하는데, 특히 일반화가능도가 익숙한 개념이 아니기 때문에 비교적 상세하게 설명하였다. 3장에서는 델파이 조사 결과에 대한 비교와 신뢰성 분석을 실시하고, 4장에서 결론과 제안사항을 서술하였다.

2. 델파이 조사와 신뢰성

2.1. 델파이 조사의 개요

델파이 조사는 구조화된 조사기법으로 원래는 전문가 패널들에게 체계적이고 인터랙티브하게 예측문제를 조사하는 방법으로 개발되었다. 1950년 대 초에 미국 RAND (research and development) 연구소에서 시작되어 가장 믿을 만한 전문가 그룹의 합의를 얻어내기 위하여 여러 차례 질문과 피드백을 반복하여 합치된 의견을 유도하고자 하였다 (Linstone과 Turoff, 1975). 전문가들은 같은 질문에 한번 답하는 것이 아니라 두 회 (round) 이상, 보통은 3-4 회 답하도록 설계되어 있으며 각 회에서 전 회에 다른 전문가들의 의견을 공유함으로써 매 단계에서 그들의 판단을 수정할 수 있는 기회를 준다. 이러한 다단계 응답 기회를 통해 전문가들이 각자의 의견을 수정하고 의견을 좁혀가며 가장 적절한 답에 수렴해줄 것으로 기대하고 있다 (Gordon, 1994).

델파이조사에 가장 적합한 경우는 정보탐색의 주된 근원이 데이터 실증이 아닌 전문가의 판단에 의할 때이다 (Ziglio, 1996) 어떤 사안에 대해 아직 공식적으로 추인된 의견이 없을 때 전문가들의 합의와 논쟁을 통해 좋은 결론을 도출해낼 수 있다는 것이 델파이조사의 아이디어이다 (Gordon, 1994). 합의를 위해 전문가들이 같은 자리에 모여 의견을 교환하는 것도 좋은 방법일 수는 있지만 면대면 회의인 경우 누군가 회의를 주도할 수 있고 권위자에게 자신의 의견을 양보하는 등 소수자에 의하여 의견이 왜곡될

수 있는 여지가 있기 때문에 분리된 공간에서 개별적으로 의견을 피력하고 이를 취합하는 것을 원칙으로 한다. 취합된 전체 의견은 다음 단계에서 전문가에게 다시 개별적으로 보내지고 각 전문가들은 다른 전문가들의 의견을 참고하여 자신의 의견을 수정하는 절차를 갖는다 (Linstone과 Turoff, 1975). 여러 번 반복해서 전문가의 의견을 구하는 방법은 비용이 많이 들기 때문에 현재 많은 델파이조사는 온라인을 통해 이루어지고 있는데 본 조사에서도 이메일로 질문을 보내고 답변을 받고 다시 수정하는 절차를 수행하였다. Gordon과 Pease (2006)은 실시간으로 다른 의견을 참조해서 자신의 의견을 수정하는 웹기반의 실시간(real-time) 델파이 조사를 개발하기도 하였고, Gnatzy 등 (2011)은 이러한 방법이 전통적인 델파이 조사에 비교해 타당성 (validity)과 신뢰성 (reliability)을 개선하고 있음을 실증하였다. 본 조사는 온라인상에서 실시간으로 수정하는 최신의 델파이조사 형태는 아니지만 단기간 내에 답변을 얻고 풍부한 피드백을 제공함으로써 온라인 델파이조사의 장점을 취득하고자 하였다. 또한 어떤 권위자에 의한 평가 객관성의 침해를 방지하기 위하여 일반적인 델파이조사와 마찬가지로 전문가들 간에 익명성을 보장하였다.

2.2. 델파이 조사의 신뢰도

Sackman (1974)은 델파이 연구에서 타당성과 신뢰성 평가에 관해 소홀히 하는 연구풍토를 비판하였는데 이는 전문가들이 지식과 경험을 공유하며 불완전한 결론에 대한 수정을 가하는 델파이연구의 특성상 불가피한 것으로 여겨지기도 하였다 (Fink 등, 1991). Pill (1971)은 델파이조사가 부차적으로 전체 그룹의 평가를 참고하기 때문에 얼마나 평가자가 확고한 평가를 반복해서 내리는가를 측정하는 신뢰성 평가는 의미가 없다고 하였다. 또한 Murphy 등 (1998)은 델파이연구는 새로운 지식을 창출하는 과학적 방법이 아니고 유용한 지식을 최적으로 사용하도록 하는 절차로 간주해야 한다고 주장하여 타당성 검증에도 다른 의견을 보였다.

그래서 델파이 조사에서는 일반적인 연구에서 제안하는 신뢰도나 타당성 측도보다는 새로운 사실의 가치를 평가하는 다양한 과학적 기준을 제시하게 되었고 이들 중 Heshusius (1990)의 적합도 기준 (goodness criteria)이 하나의 예이다. Fink 등 (1991)은 델파이 연구에서 밝혀진 사실을 신뢰하기 위해서는 여러 측도들을 같이 점검해야 한다고 하였다. Murphy 등 (1998)은 델파이 연구결과를 평가할 수 있는 방안을 제시하였는데, 즉 황금기준 (gold standard)과의 비교, 다른 원천을 통해 얻은 결과와 비교하는 기준타당도 (criterion validity), 일관성 등 내부 논리 (logic)의 평가, 안면타당성 (face validity) 평가 등이다.

일반적인 조사와 마찬가지로 델파이조사에서도 크론바하 알파 (Cronbach's α)를 이용해 신뢰도를 보고하는 경우도 있다 (예. Tomasik, 2010; Graham 등, 2003). Graham 등 (2003)은 크론바하 알파가 패널들간의 동의 정도를 측정하는데 매우 유용한 통계량이라 주장하였다. 크론바하 알파는 다음과 같이 정의된다.

$$\alpha = \frac{k}{(k-1)} \left(1 - \frac{\sum \sigma_j^2}{\sigma_x^2} \right) = \frac{k}{(k-1)} \left(\frac{\sum \sum_{i \neq j} \sigma_{ij}}{\sigma_x^2} \right),$$

여기서, k 는 평가자 패널의 수이고 σ_j^2 은 j 번째 평가자의 분산(variance), σ_{ij} 은 i, j 번째 평가자의 공분산 (covariance)이고, σ_x^2 은 각 평가자의 분산과 공분산의 합, 즉 공분산행렬의 전체 원소의 합이 된다.

신뢰성을 평가하는 방법으로 사전평가 (pre-test)와 사후평가 (post-test)의 일치성 (consistency)을 분석하는 방법이 있으나 델파이 조사의 특성상 다른 평가자의 의견을 참고하여 자신의 의견을 수정하는 과정을 갖기 때문에 매회 일관적인 평가를 내린 사실로 신뢰성을 판단할 수는 없다. 그래서 현재 신뢰도 측정은 매회 분리하여 신뢰도계수를 구하는 방식이 유일하다고 할 수 있다. 반복된 조사에서 일관성보다는 의견의 수렴 (consensus)을 이루었느냐가 델파이조사의 가치를 판단하는 기준이 될 수 있다. 회를 거듭할수록 응답자들 간의 표준편차 (SD: standard deviation)가 줄어드는 것은 의견 수렴이 발생하고

있다고 해석할 수 있기 때문에 (Rowe와 Wright, 1999; Schmidt, 1997) 평가점수의 분산을 분석하는 방법으로 신뢰도를 평가하는 방법을 추천하였다.

2.3. 일반화가능도

Cronbach 등 (1963, 1972)은 새로운 신뢰성 측도로서 일반화가능도 (generalizability)를 제안하였다. 신뢰성 측도로 가장 많이 사용되고 있는 크론바하 알파를 제안했던 저자들은 알파 계수가 문항 또는 평가자의 내적일관성에 관한 측도이고 안정성에 관한 측도가 되지 못함을 파악하였다. 평가자 간에 차이를 보는 안정성의 평가는 상관계수나 검사-재검사 방식으로 할 수 있으나 이는 평가가 내포하는 오차변동의 크기를 측정하지 못하기 때문에 평가 설계의 개선에 사용하기가 어렵다는 점을 개선하기 위하여 측정오차 (measurement error)의 다양한 원천 (source)을 평가하는 일반화가능도 이론을 제안하였다 (Brennan, 2001).

일반화가능도 이론에서는 i 번째 지식항목에 대하여 r 번째 평가자 (rater)에게 받은 평가점수의 모형을 다음과 같이 정의할 수 있다.

$$\begin{aligned} X_{ir} &= \mu + (\mu_i - \mu) + (\mu_r - \mu) + (X_{ir} - \mu_i - \mu_r + \mu) \\ &= \mu + \nu_i + \nu_r + \nu_{ir}; i = 1, \dots, n_i, r = 1, \dots, n_r, \end{aligned}$$

여기서 μ 는 모집단 (population)과 전집 (universe)에서 나온 전체 평균 ($\mu = E_i E_r X_{ir}$)이고, ν 는 각 변인들의 효과 (주효과와 상호작용)라 할 수 있다. 그리고 모집단 평균 (population mean)은 $\mu_i = E_r X_{ir}$, $\mu_r = E_i X_{ir}$ 등이다.

일반화가능도 이론에서 개별적인 평가점수는 단지 무한한 모집단과 가능한 측정의 여러 전집에서 나온 표본 (sample)으로 간주한다. 예를 들어, i 번째 항목을 n_r 명의 평가자가 평가점수를 부여하였다면 무한한 항목의 모집단으로 부터 항목표본을 얻었고 평가자는 수많은 가능한 평가자에서 선택된 평가자이기 때문에 가능한 전집 (universe of admissible observation)에서 뽑은 표본으로 표현되는 것이다. 일반화가능도 이론에서는 문항, 평가자, 평가시기 등 (test item, rater, test occasion)을 국면 (facet)이라 하는데 통계학의 요인 (factor)과 유사한 개념이지만 무한히 수준을 확장할 수 있다는 점이 다르다. 측정의 개체는 주로 피험자이며 이들의 개인적 차이는 분명한 사실이고 실험자는 이런 차이에 관심을 갖는 것이기 때문에 개체간의 차이는 오차의 요인으로 간주하지 않는다. 본 연구에서는 지식항목이 피험자와 같은 역할을 한다. 위 모형은 오차에 영향을 주는 국면이 평가자 뿐이기 때문에 1국면 설계 (single-facet design) 모형이라 정의한다.

신뢰도가 높은 설계는 각 국면에서 차이가 적을 것이며 이러한 차이는 오차의 개념으로 파악한다. 각 피험자에게 수차례 국면의 수준에서 측정한 값의 평균은 개인의 진점수 (또는 전집점수)의 추정값인데 모든 오차는 국면에 의해 생기는 것이므로 분산성분 (variance components)으로 그들의 영향력을 측정할 수 있다. 앞에서의 모형의 지식항목 평가점수의 분산성분은 다음과 같으며 이들 값은 ANOVA 분석을 통해서 추정할 수 있는데 추정된 분산 성분을 통해 신뢰도를 추정할 수 있다.

$$\sigma^2(X_{ir}) = \sigma_i^2 + \sigma_r^2 + \sigma_{ir}^2.$$

각 항목에 대한 평가는 n_r 명의 평가자에 의해 행해지는데 평가자는 수많은 전집에서 뽑힌 사람들이므로 항목의 평가점수의 기대값이 항목의 전집점수 (universe score)가 된다. 이때 관측된 지식항목의 평가자에 대한 평균점수 (mean score)의 모형과 분산성분을 다음과 같이 표현할 수 있다.

$$\begin{aligned} X_{iR} &= \Sigma X_{ir} / n_r = \mu + (\mu_i - \mu) + (\mu_R - \mu) + (X_{iR} - \mu_i - \mu_R + \mu), \\ \sigma^2(X_{iR}) &= \sigma_i^2 + \sigma_R^2 + \sigma_{iR}^2 = \sigma_i^2 + \sigma_r^2 / n_r + \sigma_{ir}^2 / n_r. \end{aligned} \quad (2.1)$$

상대오차 (relative error)는 지식항목의 관측된 편차 (observed deviation)와 그것의 전집점수편차 (universe score deviation)의 차이 (difference)로 정의된다.

$$\delta_i = (X_{iR} - E_i X_{iR}) - (\mu_i - E_i \mu_i).$$

이때 상대오차분산 (relative error variance)은 다음과 같이 정의한다.

$$\sigma^2(\delta_i) = \sigma^2[(X_{iR} - \mu_R) - (\mu_i - \mu)] = \sigma_{ir}^2/n_r.$$

Cronbach 등 (1972)는 일반화가능도 계수 (generalizability coefficient)를 다음과 같이 전집점수의 분산과 전집점수분산 및 상대오차분산의 합의 비 (ratio)로 정의한다.

$$E\rho^2 = (\sigma_r^2)/(\sigma_r^2 + \sigma^2(\delta)), \quad (2.1) \quad \text{에서} \quad \sigma^2(\delta) = \sigma_{ir}^2.$$

분산분석에서 평가자간에 분산이 크다면 평가자의 주관이 평가에 큰 영향을 주는 것을 의미하고, 일반화가능도 계수값이 크면 평가가 우연성보다는 평가자의 주관을 잘 반영하고 있으므로 신뢰도가 높다는 의미가 된다.

모형 (2.1)의 일반화가능도 계수의 불편 (unbiased) 추정값은 평균제곱합의 기대값 성질을 이용하여 분산분석표에서 구할 수 있다.

$$E\hat{\rho}^2 = \frac{\hat{\sigma}_r^2}{\hat{\sigma}_r^2 + \hat{\sigma}_{ir}^2/n_r} = \frac{MS_r - MS_{ir}}{MS_r},$$

여기서, MS_r 은 평가자 간의 평균 제곱합 (Mean Squares due to raters), MS_{ir} 은 평가자와 항목의 교호작용에 의한 평균 제곱합 (Mean Squares due to interaction)이다. 1국면 설계에서 추정된 일반화가능도 계수는 크론바하 알파와 일치한다 (Lee, 2014).

그런데 본 연구에서는 평가자들에게 항목에 대한 여러 속성에 대하여 평가하도록 하였기 때문에 i 번째 지식항목의 t 특성에 대하여 r 번째 평가자에게 받은 평가점수의 모형을 다음과 같이 정의할 수 있다.

$$X_{itr} = \mu + \nu_i + \nu_t + \nu_r + \nu_{it} + \nu_{tr} + \nu_{ir} + \nu_{itr}; \quad i = 1, \dots, n_i, t = 1, \dots, n_t, r = 1, \dots, n_r,$$

여기서 μ 는 모집단과 전집에서 나온 전체 평균이고, ν 는 각 변인들의 효과 (주효과와 상호작용)라 할 수 있다. 일반적인 일반화가능도 교재에서는 이러한 모형의 경우 2국면 설계 (two-facet design)로 분석하지만 측정의 개체는 항목과 특성이기 때문에 이들간의 차이는 분명한 사실이고 실험자는 이런 차이에 관심을 갖는 것이기 때문에 항목과 특성 간의 차이는 오차의 요인으로 간주하지 않고 오차에 영향을 주는 국면이 평가자 뿐이기 때문에 1국면 설계 (single-facet design) 모형으로 분석하도록 한다.

분산을 각 요소 (components)로 분해하고 각 요소별 분산의 추정값을 얻기 위해서 다음 Table 2.1과 같은 분산분석 (ANOVA)표를 얻는다.

이 때 전집은 평가자 뿐이기 때문에 평균점수의 모형을 X_{itR} 로 적고 이때 상대오차분산을 다음과 같이 정의한다.

$$\sigma^2(\delta_{it}) = \sigma^2[(X_{itR} - \mu_R) - (\mu_{it} - \mu)] = \sigma_{iR}^2 + \sigma_{tR}^2 + \sigma_{itR}^2 = \sigma_{ir}^2/n_r + \sigma_{tr}^2/n_r + \sigma_{itr}^2/n_r.$$

일반화가능도 계수 추정값은 평균제곱합의 기대값 성질을 이용하여 분산분석표에서 구할 수 있다.

$$E\hat{\rho}^2 = \frac{\hat{\sigma}_r^2}{\hat{\sigma}_r^2 + \hat{\sigma}(\delta)},$$

여기서, $\hat{\sigma}_r^2 = [(MS_r - MS_{tr} - MS_{ir} + MS_{itr})]/(n_i n_t)$, $\hat{\sigma}^2(\delta) = \hat{\sigma}_{it}^2/n_r + \hat{\sigma}_{tr}^2/n_r + \hat{\sigma}_{itr}^2/n_r$, $\hat{\sigma}_{ir}^2 = [(MS_{ir} - MS_{itr})]/n_t$, $\hat{\sigma}_{tr}^2 = [(MS_{tr} - MS_{itr})]/n_i$, $\hat{\sigma}_{itr}^2 = MS_{itr}$.

Table 2.1 Mean Suares and Expectations when number of components is 3

Source of variation	df	Mean Square	Expected Mean Square
Item(<i>i</i>)	$n_i - 1$	MS_i	$\sigma_{itr}^2 + n_r\sigma_{it}^2 + n_t\sigma_{ir}^2 + n_in_r\sigma_i^2$
Property(<i>t</i>)	$n_t - 1$	MS_t	$\sigma_{itr}^2 + n_i\sigma_{tr}^2 + n_r\sigma_{it}^2 + n_in_r\sigma_t^2$
Rater(<i>r</i>)	$n_r - 1$	MS_r	$\sigma_{itr}^2 + n_t\sigma_{ir}^2 + n_i\sigma_{tr}^2 + n_in_t\sigma_r^2$
<i>it</i>	$(n_i - 1)(n_t - 1)$	MS_{it}	$\sigma_{itr}^2 + n_r\sigma_{it}^2$
<i>tr</i>	$(n_t - 1)(n_r - 1)$	MS_{tr}	$\sigma_{itr}^2 + n_i\sigma_{tr}^2$
<i>ir</i>	$(n_i - 1)(n_r - 1)$	MS_{ir}	$\sigma_{itr}^2 + n_t\sigma_{ir}^2$
<i>itr</i>	$(n_i - 1)(n_t - 1)(n_r - 1)$	MS_{itr}	σ_{itr}^2

3. 델파이 조사결과

3.1. 기초통계량

6명의 평가위원에게 200개의 전통지식에 대하여 5개 특성의 가치에 대하여 5점 척도로 평가하는 1차 델파이 조사후 각 특성별 평가의 전체 평균을 구하였다. 그리고 전체 평균자료 및 순위 등을 공개하고 동일한 6명의 평가위원에게 전통지식의 가치를 재평가하는 2차 델파이 조사를 실시하였다. 주관적이고 독립적인 판단과 평가를 보장하고 (1차) 타인의 정보를 참고하고 수정할 수 있는 기회를 부여하여 (2차) 객관적인 평가를 내릴 수 있도록 하는 델파이 조사의 장점을 가질 수 있도록 조사를 구성하였고 통계적인 분석은 SPSS를 이용하였다. 다음 Table 3.1은 문헌과 현장에서 얻는 전통지식 각 100건에 대한 두 차례 평가에 의한 평가점수의 평균과 표준편차 등 기초통계량 값이다.

Table 3.1 Simple statistics of the survey data

source	round	statistic	practicality	economic	industriali-	availability	potential	average
				value	zation		value	
literature review	first	mean	3.202	3.042	2.932	2.988	3.443	3.121
		SD	0.816	0.812	0.879	0.831	0.823	0.832
	second	mean	3.207	3.077	2.985	3.030	3.493	3.158
		SD	0.691	0.613	0.628	0.649	0.698	0.656
field interview	first	mean	2.955	2.756	2.541	2.713	3.040	2.801
		SD	0.806	0.786	0.891	0.847	0.829	0.832
	second	mean	3.045	2.773	2.658	2.780	3.190	2.889
		SD	0.803	0.670	0.636	0.724	0.767	0.720

두 차례의 평가점수는 연관성이 매우 높았다 (문헌자료의 피어슨 상관계수: $r = 0.9816$, $p < 0.001$, 현장자료의 피어슨 상관계수: $r = 0.9715$, $p < 0.001$). 어느 정도 평균 값의 변경이 발생하였음을 알 수 있는데 전반적으로 2차 평가시 평균이 상승하였다. 또한 2차 델파이 조사 후 각 문항에 대한 평가의 표준편차가 줄어들었음을 알 수 있었다. 실용성에 대한 평가의 표준편차는 전 100개 항목에 대하여 1차에서는 0.816이었으나 2차수정후에는 0.691로 의견 수렴이 발생하였음을 알 수 있다. 전반적으로 2차 평가시 평균이 상승하고 편차가 줄어들었으므로 의견수렴이 이루어졌음을 추측할 수 있지만 좀 더 세밀한 분석이 필요하다.

3.2. 델파이 조사의 신뢰성 분석

평가자가 어떤 항목에 대하여 특성별로 일관적인 평가를 내리는지 신뢰성 분석을 통해 검증해볼 필요가 있다. 한 항목에 대하여 6명의 평가자가 평가를 한 형태이므로 크론바하 알파계수로 신뢰도를 평가하였다. 다음 Table 3.2에서 신뢰도 계수가 1차에서는 0.54에서 0.75 정도로 나타나고 있지만 2차에서

는 0.78에서 0.87 정도로 이보다 상승하는 패턴을 보이고 있다 (예. 문헌의 실용성 (practicality) 평가 신뢰도가 1차: 0.5428 $\Rightarrow$ 2차: 0.7825). 이는 2차 평가 후 평가의 신뢰도가 상승하였음을 보여주며 전문가에 의한 델파이조사의 신뢰도가 높음을 의미한다.

그러나 평가자의 분산 추정값이 각 특성별로 2차에서 증가 또는 감소하는 경향을 보여 신뢰도가 증가하고 있지만 이것이 공감대 (consensus)를 형성하는 증거가 되지는 못하고 있음을 시사하였다.

Table 3.2 Reliability of the survey

source	round	statistic	practicality	economic value	industrialization	availability	potential value
literature review	first	rater variance, $\hat{\sigma}_r^2$	0.1372	0.1414	0.2426	0.1358	0.1153
		item variance, $\hat{\sigma}_i^2$	0.1143	0.2250	0.3004	0.1861	0.2502
		$i \times r$ variance, $\hat{\sigma}_{ir}^2$	0.5778	0.5669	0.5984	0.6218	0.6074
		Cronbach's α	0.5428	0.7043	0.7507	0.6423	0.7119
	second	rater variance, $\hat{\sigma}_r^2$	0.2074	0.1221	0.1179	0.1351	0.1912
		item variance, $\hat{\sigma}_i^2$	0.1862	0.2730	0.303	0.2318	0.3930
		$i \times r$ variance, $\hat{\sigma}_{ir}^2$	0.3106	0.3025	0.3138	0.3309	0.3388
		Cronbach's α	0.7825	0.8441	0.8528	0.8078	0.8744
field interview	first	rater variance, $\hat{\sigma}_r^2$	0.1922	0.1795	0.2722	0.2118	0.1761
		item variance, $\hat{\sigma}_i^2$	0.1059	0.0972	0.1444	0.2001	0.1260
		$i \times r$ variance, $\hat{\sigma}_{ir}^2$	0.4988	0.4992	0.5828	0.5675	0.5612
		Cronbach's α	0.5602	0.5388	0.5978	0.6790	0.5739
	second	rater variance, $\hat{\sigma}_r^2$	0.3922	0.1885	0.1512	0.2319	0.3128
		item variance, $\hat{\sigma}_i^2$	0.1715	0.1999	0.1976	0.2668	0.2814
		$i \times r$ variance, $\hat{\sigma}_{ir}^2$	0.2835	0.2975	0.2905	0.3281	0.3152
		Cronbach's α	0.7840	0.8013	0.8032	0.8299	0.8427

3.3. 델파이 조사의 일반화가능도 분석

신뢰도는 자료의 전체분산에서 진점수 (true score)의 분산 비율로 표현되는데 앞의 식에서는 분산을 다음과 같은 요소 (components)로 분해할 수 있다.

Table 3.3에 의하면 평가자 분산, 항목의 분산, 특성 간 분산이 모두 2차 조사에서 증가하였고, 크론바하 신뢰도 계수나 일반화가능도 계수 모두 1차 조사에 비해 2차조사에서는 높은 값을 나타내었다. 이는 두번째 델파이 조사가 첫번째 조사에 비해 평가자의 신뢰도가 개선된 것을 의미한다. 또한 항목간의 분산이 커지는 것은 다른 평가자의 의견을 반영해 항목간 차이의 구별이 뚜렷해지고 있음을 말해주며 특성 간의 분산이 커지는 것은 각 특성에 따른 평가가 좀더 차별적으로 이루어지고 있음을 의미한다. 이는 각 특성별로 평가하는 조사의 구성타당도 (construct validity)가 개선되었음을 의미한다.

그런데 극단적인 평가들의 감소로 인하여 각 평가의 오차항 분산이 줄어들었지만 평가자 간의 분산이 증가한 것은 주목할 만한 사실이다. 응답자들 간의 표준편차 (SD)가 줄어드는 것은 의견 수렴이 발생하고 있다고 해석할 수 있고 (Rowe와 Wright, 1999; Schmidt, 1997), 델파이조사는 전문가들에게 반복적으로 의견을 받아 그들이 의견이 수렴하기를 기대하고 있는데, 2차 델파이조사에서 평가자들간의 편차가 오히려 커지는 현상이 발생하였다. 즉, 전체적인 변동이 줄고 상대오차도 줄어 신뢰도 계수, 일반화가능도 계수 모두 증가하였지만 이것이 평가자들간의 의견이 좁혀져서가 아니고 평균 회귀 등에 따른 현상 때문일 수도 있다. 그러므로 신뢰도가 개선되었다고 델파이조사가 그 목적인 바를 달성하였다고 말하기 어렵다. 이는 일반화가능도나 크론바하 신뢰도 계수로는 파악할 수 없는 현상으로 각 요소별 분산을 분해하는 일반화가능도 유도과정에서 각 국면별 편차의 변화를 파악하여 추가적인 차수의 조사가 필요한지를 결정해야 함을 의미한다.

Table 3.3 Estimated variances and generalizability coefficients

source	source	first		second	
		MS	Estimate of Variance	MS	Estimate of Variance
literature review	Item(<i>i</i>)	8.388	0.2026	8.598	0.2520
	Rater(<i>r</i>)	55.684	0.0950	67.236	0.1272
	Property(<i>t</i>)	25.522	0.0322	25.162	0.0368
	(<i>tr</i>)	6.131	0.0595	2.933	0.0275
	(<i>ir</i>)	2.236	0.4103	0.885	0.1414
	(<i>it</i>)	0.260	0.0126	0.330	0.0254
	(<i>itr</i>)	0.184	0.1842	0.178	0.1779
Generalizability		0.4657		0.6875	
field interview	Item(<i>i</i>)	5.775	0.1282	7.135	0.2070
	Rater(<i>r</i>)	75.331	0.1320	116.540	0.2255
	Property(<i>t</i>)	23.812	0.0269	29.010	0.0429
	(<i>tr</i>)	7.662	0.0746	3.156	0.0298
	(<i>ir</i>)	1.891	0.3370	0.828	0.1312
	(<i>it</i>)	0.244	0.0064	0.271	0.0165
	(<i>itr</i>)	0.206	0.2056	0.172	0.1718
Generalizability		0.5620		0.8025	

4. 결론

본 연구에서는 문헌 및 현장에서 얻은 생물자원지식들의 가치를 평가하기 위해 전문가들을 대상으로 델파이 조사의 신뢰성과 적절성을 평가하였다. 델파이 조사는 전문가들이 아직 결론이 나지 않은 미확정된 가치에 대한 평가를 개별적으로 평가를 내리고 다음 단계에서 다른 전문가들의 의견을 참고하며 자신의 의견을 수정하는 절차를 갖는다. 본 연구에서 1차는 온라인을 통하여 전문위원에게 평가를 의뢰하여 전문위원들 간 상호 영향을 배제하였으며 2차에서는 1차의 평가 결과를 취합하여 전체 평균을 각 평가위원에게 알려주고 수정된 평가를 받았다. 두 차례의 평가점수는 연관성이 매우 높으면서 2차에서는 다른 전문가들의 의견을 수용해 각 문항에 대한 평가의 표준편차가 줄어들었다. 또한 전반적으로 2차 평가시 평가점수의 평균이 상승하는 경향을 볼 수 있는데 이는 전문가들이 잘 모르던 지식에 대하여 다른 전문가의 평가를 참고해 재고하는 경향을 보였을 것임을 추측하게 한다. 본 조사의 신뢰성을 파악하기 위해 일반적인 신뢰도 계수인 크론바하 알파와 평가점수의 구성요소별 분산을 분석하는 일반화가능도 분석을 실시하였다. 이 두 분석을 통해 2차 평가 후 평가의 신뢰도가 상승하여 전문가에 의한 델파이 조사의 신뢰도가 높다는 사실을 지지하였다.

그러나 델파이조사의 목적은 평가자의 의견 수렴인데 일반화가능도 유도과정에서 신뢰도가 높아짐에도 평가자 간의 분산은 증가하는 현상을 발견하였다. 이는 델파이조사의 우수성을 검증할 때 단순한 신뢰도 또는 일반화가능도 만으로는 부족함을 의미한다. 일반화가능도가 증가하더라도 평가자 간의 분산이 증가하면 아직 의견의 수렴이 이루어진 것이 아니므로 추가적인 델파이 조사가 필요하다 할 수 있다.

델파이 조사는 평가자의 의견이 수렴되어야 하므로 신뢰도가 증가하였다고 조사를 마칠 것이 아니라 평가점수의 요소별 분산을 분석하여 평가자 간의 분산이 작아질 때까지 의견수렴을 위하여 반복해야 한다. 그런데 본 연구에서 전체적인 분산이 줄어들더라도 평가자 간의 분산이 증가할 수 있음을 실증하였고 이는 일반적인 분산의 계산, 신뢰도 계수의 유도로 파악할 수 없음을 보였다. 즉, 조사의 신뢰성과 의견의 수렴성을 파악하기 위해 일반화가능도 계수를 유도하는 과정을 포함하여 각 요소별 분산의 추정값을 분석할 것을 제안한다.

참고문헌

- Brennan, R. L. (2001). *Generalizability Theory*, Springer-Verlag, New York.
- Bubela, T. and Gold, E. R. (2012). *Genetic resources and traditional knowledge*, Edward Elgar Publishing Inc., Northampton, MA.
- Choi, K., Kim, H. and Song, M. J. (2014). Selecting order of priority using Delphi and statistical method. *Journal of the Korean Data & Information Science Society*, **25**, 1161-1170.
- Cronbach, L. J., Gleser, G. C., Nanda, H. and Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. John Wiley, New York.
- Cronbach, L. J., Nageswari, R., and Gleser, G. C. (1963). Theory of generalizability: A liberation of reliability theory. *The British Journal of Statistical Psychology*, **16**, 137-163.
- Eom, J. S. and Kim, S. B. (2013). Study on the defence R&D project risk analysis using AHP. *Journal of the Korean Data & Information Science Society*, **24**, 557-569.
- Fink, A., Kosecoff, J., Chassin, M. and Brook, R. (1991). *Consensus Methods: Characteristics and Guidelines for Use RAND*, Santa Monica, California.
- Gnatzy, T., Warth, J., von der Gracht, H. and Darkow, I. L. (2011). Validating an innovative real-time Delphi approach - A methodological comparison between real-time and conventional Delphi studies. *Technological Forecasting & Social Change*, **78**, 1681-1694.
- Gordon, T. J. (1994). *The Delphi method*, American Council for the United Nations University, Washington, DC.
- Gordon, T. and Pease, A. (2006). RT Delphi: An efficient, "Round-less" almost real time Delphi method. *Technological Forecasting and Social Change*, **73**, 321-333.
- Graham, B., Regehr, G. and Wright, J. G. (2003). Delphi as a method to establish consensus for diagnostic criteria. *Journal of Clinical Epidemiology*, **56**, 1150-1156.
- Heshusius, L. (1990). *Goodness Criteria*. In *The Paradigm Dialog* (Guba E.G. ed.). Sage, Newbury Park, CA, pp. 198-201.
- Kim, H. and Song, M. J. (2011). Analysis and recordings of orally transmitted knowledge about medicinal plants in the southern mountainous region of Korea. *Journal of Ethnopharmacology*, **134**, 676-696.
- Kim, H. and Song, M. J. (2013). Analysis of traditional knowledge about medicinal plants utilized in communities of Jirisan National Park (Korea). *Journal of Ethnopharmacology*, **153**, 85-89.
- Lee, K. H. (2014). *Statistical Data Analysis using SPSS*, Freedom Academy, Seoul.
- Linstone, H. A. and Turoff, M. (Eds.). (1975). *The Delphi method: Techniques and applications*, Addison-Wesley, Boston, MA.
- Ministry of Environment (2010). *ABS guidebook*, Ministry of Environment, Seoul.
- Murphy, M. K., Black, N., Lamping, D. L., McKee, C. M., Sanderson, C. F. B., Askham, J. et al. (1998). Consensus development methods and their use in clinical guideline development. *Health Technology Assessment*, **2**, 1-88.
- Pill, J. (1971). The Delphi method: substance, context, a critique and an annotated bibliography. *Socio-Economic Planning and Science*, **5**, 57-71.
- Rowe, G. and Wright, G. (1999). The Delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting*, **15**, 353-375.
- Ryu, T. K. et al. (2012). *A Study on Intellectual Property Valuation : the Case of Traditional Knowledge*, Korea institute of intellectual property, Seoul.
- Sackman, H. (1974). *Delphi assessment: Expert opinion, forecasting and group process*, Rand Corporation, Santa Monica, CA.
- Schmidt, R.C. (1997). Managing Delphi Surveys Using Nonparametric Statistical Techniques. *Decision Sciences*, **28**, 763-774.
- Tomasik, T. (2010). Reliability and validity of the Delphi method in guideline development for family physicians. *Quality in Primary Care*, **18**, 317-26.
- Ziglio, E. (1996). *The Delphi method and its contribution to decision-making*. In M. Adler & E. Ziglio (Eds.), *Gazing into the oracle: The Delphi and its application to social policy and public health*, Jessica Kingsley, London, England.

Reliability of Delphi survey for traditional knowledge on agricultural resources[†]

Ki Hoon Lee¹, Mi-Jang Song², Hyun Kim³

¹Department of Business Administration, Jeonju University

²Department of Integrated Bio-Resource Science, Jeonju University

³Department of Alternative Medicine, Jeonju University

Received 1 0000, revised 0 0000, accepted 0 0000

Abstract

In the knowledge and information age, to discover and protect Intellectual Properties would be very important for their economic value as a major growth engine. This study evaluated the reliability of a Delphi survey conducted by experts to assess the value of agricultural resources knowledge obtained from literature reviews and field interviews. Delphi method is collecting the opinions of experts for several rounds repeatedly, in the next round the experts have chance to modify their opinion. Scores between two rounds are highly correlated and standard deviations are declined for second round to imply that some correction of their evaluations are made. To check reliability of Delphi survey of two rounds Cronbach's reliability coefficient and Generalizability coefficient are derived. The Cronbach alpha's supported the reliability of the method, but the Generalizability analysis revealed some unexpected results while checking the variance components of sources of measurement errors. Despite the increased reliability coefficients, the deviations between the raters are increased which means that additional rounds are required to get consensus, the goal of Delphi research.

Keywords: Agricultural resources knowledge, Delphi method, generalizability, reliability.

[†] This work was carried out with the support of "Cooperative Research Program for Agriculture Science & Technology Development (Project No. PJ009566032015)" Rural Development Administration, Republic of Korea.

¹ Professor, Department of Business Administration, JeonBuk 560-759, Korea.
E-mail: khlee@jj.ac.kr

² Assistant Professor, Department of Integrated Bio-Resource Science, JeonBuk 560-759, Korea.

³ Professor, Department of Alternative Medicine, JeonBuk 560-759, Korea.