

集團構成員數를 考慮한 確率的 意見 收斂方法 (Group Format Selection Considering the Effect of Group Size in Aggregating Probabilistic Opinions)

朴 奭 根*
趙 成 九*

ABSTRACT

In this study three types of aggregation methods such as the Estimate-Talk-Consensus(ETC) process, the Estimate-Talk-Estimate(ETE) process, and as a new approach the Estimate-Talk-Leader's Estimate (ETLE) process are compared to find which one of the three group processes considered is more effective than others. We, also, investigate the effect of group size on the performance of the group processes.

Some experiments were conducted. It was shown that both the ETC and the ETLE processes performed better than the ETE process in approaching correct estimates in this judgmental task. As the size of group increased, only the ETC and the ETE processes were shown to result in positive effect.

1. 서론

기업, 정부, 연구소, 학교, 병원 등과 같은 오늘날의 조직들은 한 개인의 직관과 경험만으로는 그 조직이 직면하고 있는 문제 자체의 성격도 파악하기 힘들 경우가 많고, 관련된 이익집단들에게 자신이 내린 판단을 정당화시킬 필요가 생길 수도 있다. 더구나, 이와 같은 상황에 처한 의사결정자가 관심대상인 모수들에 대한 지식이 충분하지 않다면, 그는 다른 사람들, 가능하다면, 그 분야의 전문가들에게 자문을 구할 필요가 있을 것이다. 특히, 고려하고 있는 문제의 성격상 확률적 자료가 필요하다면, 그 의사결정자는 자문위원들의 계

량화된 판단들을 확률분포의 형태로 얻을 필요가 발생하게 된다. 이러한 경우, 전문가들은 그들 자신이 느끼는 불확실성을 확률형태로 표현할 수 있어야 할 뿐만 아니라, 하나의 '집단(group)' 확률분포를 형성하기 위한 방안을 강구하여야 할 것이다.

본 연구에서는 Winkler[19]의 연구에서 집단구성원들간의 의견차를 좁혀 주는 효과를 가진 것으로 나타난 집단적 재평가법, Gustafson et al.[6]의 연구에서 가장 우수한 성과를 가져 온 것으로 밝혀진 명목집단법, 그리고 본 연구에서 제시하는 대표자 재평가법 등을 비교하고자 한다. 또한, 이

* 동국대학교 산업공학과

방법들에 대하여 2인 집단, 4인 집단 및 7인 집단을 할당하여

- (1) 각 의견수렴방법에 있어서, 피실험자들이 개별적으로 평가과정을 수행하는 단계(개별평가 단계)와 집단을 구성하고 구성원들간의 집단토의를 거친 후 평가하는 단계(집단토의후 단계)로부터 얻은 성과를 비교해 봄으로써 집단적 의견수렴방법의 필요성을 조사하고,
- (2) 집단구성원수(group size)가 각 의견수렴방법의 성과에 미치는 효과를 조사함으로써 집단구성원수와 의견수렴방법 사이의 상관관계를 알아 보고,
- (3) 이를 통하여, 집단구성원수에 따른 효과적인 의견수렴방법들을 제시하고자 한다.

2. 복수인확률유도방법에 관한 기존연구

복수인확률유도(multiperson probability elicitation)를 위한 방법은 수리적(또는 통계적)인 것과 행위적인 것으로 나누어 살펴 볼 수 있다. 먼저, 수리적(또는 통계적) 방법에 관한 연구결과들을 살펴보면, Stone[17]을 비롯한 많은 학자들은 한 집단을 이루는 구성원들의 총체적 의견들을 선형 결합 형태를 갖는 하나의 확률분포(Stone은 이 분포를 'opinion pool'이라 칭하였음)로 표현하는 방법을 제안했는데, 이와 같은 'opinion pool', 또는, '가중평균'모형에서 쟁점이 되는 것은, 각 집단구성원(또는 전문가)의 분포에 대한 가중치를 결정하는 일이다[5, 17, 19, 21].

그런데, 집단에 의한 성과의 질이, 평균적으로 볼 때, 개인으로 부터 도출된 그것을 능가한다는 실증적 결과들이 얻어짐에 따라 집단구성원들간의 상호교류(interaction) 효과를 연구하는 행위적 접근방법이 활발하게 연구되어 왔다[9, 18, 19, 21].

Winkler[19] 및 Winkler & Cummings[22]는 합의분포(consensus distribution)에 도달하는 네 가지 방법들을 실증적으로 비교분석하였다. 그 결

과, 피이드백 및 재평가법(Feedback & Reassessment)과 집단적 재평가법(Group Reassessment, 또는 Estimate-Talk-Consensus)으로부터 구성원들간의 의견차를 좁혀 주는 효과를 얻을 수 있었지만, 피이드백 및 재평가법을 수행하는 데에는 절차상의 번거로움이 있음을 지적했다.

또한, 집단구성원들간의 상호교류 정도가 상이한 세 종류의 집단적 추정방법들을 개인적 추정방법과 비교분석한 Gustafson et al.[7]의 연구결과에 따르면, 명목집단법(Estimate-Talk-Estimate)으로부터 가장 우수한 성과를 얻었다고 한다[7, p. 289]. 그런데, Winkler[19]의 집단적 재평가법은 결과적으로 동일한 집단내의 구성원들이 집단토의를 통해 하나의 합의적 성격을 띤 분포를 이끌어 내야 하기 때문에, 구성원들에게 심리적 부담을 주기 쉽고, 그로 인하여, 집단적 평가방법으로부터 기대할 수 있는 성과증가분을 상쇄시킬 수도 있다는 단점이 있다.

반면에, Gustafson et al.[7]의 명목집단법은 행위적이 아닌 수리적 의견수렴방법이라는 점으로 인하여 적용상의 제약을 받을 수도 있다는 문제점을 안고 있다. 본 연구에서는 실제로 의견수렴과정을 필요로 하는 많은 상황들과 흡사한 성격을 지닌 대표자 재평가법(Estimate-Talk-Leader's Estimate)을 기존의 다른 연구들에서 제시되었던 방법들과 비교해 보고자 한다. 이 방법은 집단토의과정을 통해 다른 구성원들의 의견을 들어 볼 수 있는 집단적 재평가법의 장점을 살리면서, 합의분포를 유도해야 하는 심리적 부담으로 인하여 집단토의 도중에 일어나기 쉬운 구성원들간의 의견충돌을 줄일 수 있다는 장점이 있다. 또한, 실제적으로, 기업이나 기타 공공기관의 사장 또는 위원장이 이와 같은 대표자의 역할을 수행하고 있는 경우가 많이 있기 때문에, 이 방법이 다른 방법들과 대등한 성과를 가져 올 수만 있다면, 현장에서 가장 용이하게 받아들일 수 있는 방법이라고 본다.

한편, Libby & Blashfield[9]는, 복수인의 판단 결과들을 합성하는 그의 연구에서, 소수의 인원만으로도 다수의 인원으로부터 얻을 수 있는 성과의 상당부분을 얻을 수 있다고 보고했다. 이 연구결과들을 포함한 대부분의 집단구성원수(group size)의 효과에 관한 연구들은 구성원들의 개인적 평가결과들을 수리적이거나, 또는 통계적으로 무작위하게(statistically randomly) 조합하는 형태로 집단적 평가결과를 유도하는 방법을 사용하고 있다[2, 3, 6]. 그러나, 본 연구에서는 확률적 의견수렴방법(probabilistic opinion aggregation methods)들의 각각에 대해서 집단구성원수에 따른 소요인원을 직접적으로 배정하는 방법을 사용하여 확률적 의견 수렴방법과 집단구성원수의 관계에 대한 실증적인 결과를 얻고자 한다.

3. 연구방법

3.1. 연구의 대상

본 연구에서 비교하고자 하는 의견수렴방법들의 내용은 다음과 같다:

(1) 명목집단법

(Estimate-Talk-Estimate, ETE)

피실험자들은, 먼저, 다른 사람들과의 접촉(또는 상호교류)없이 개별적인 추정을 한다. 이 과정이 끝나면, 구성원들간의 상호교류(공개토론)을 허용하고, 이어서, 모든 구성원들이 개별적이고 단독적인 추정을 한번 더 수행하게 된다.

(2) 집단적 재평가법

(Estimate-Talk-Consensus, ETC)

이 방법도 명목집단법(ETE)과 거의 동일한 순서로 진행되는데, 한 가지 다른 점은, 구성원들간의 공개적 토론단계에서 합의적 성격을 가진 하나의 합의분포(consensus distribution)를 도출해야 한다는 점이다.

(3) 대표자 재평가법

(Estimate-Talk-Leader's Estimate, ETLE)

다른 사람들과의 상호교류없이 개별적인 추정을 끝낸 피실험자들은 각 집단별로 선출된 집단의 대표자에게 구성원들간의 상호교류를 통해서 정보를 제공한다. 이어서, 대표자만의 단독적인 추정이 이루어진다. 본 연구에서는 대표자에게 확률평가에 따른 권한과 더불어, 평가결과(즉, 성과(performance))에 따른 최소한의 책임을 부여하기 위해서, 각 집단구성원수(2인, 4인, 7인)마다 가장 우수한 성과를 얻은 집단에게 상금을 부여하였다. 이상과 같은 세 가지 종류의 집단적 방법들은, 각각, 2인집단, 4인집단, 그리고 7인집단들로 구성되어 있다.

3.2. 확률유도방법

Spetzler & Staël von Holstein[15, p.340]에 따르면, 확률유도(probability encoding, 또는, elicitation methods)란, “불확실한 양(uncertain quantities)에 관한 개인의 판단을 이끌어 내고, 이를 계량화하는 과정”으로 정의된다. 본 연구에서는 기법자체가 단순하고, 피실험자들이 기법을 숙지하도록 하기위한 훈련이 용이하다는 장점을 지닌 누적분포함수 및 프렉타일법(Cumulative Distribution Function & Fractiles)을 확률유도방법으로 사용하고자 한다[14].

피실험자들은, 이 기법을 사용하여, 각 질문항목에 대한 자신의 판단을 하나의 확률분포 상에서 .01, .10, .25, .50, .75, .90, 그리고 .99 프렉타일들의 형태로 표현한다. 또한, 일반적으로 볼 때, 사용될 극한 확률점(extreme points; 여기서는, .01, .10, .90, .99 프렉타일들이 해당된다)의 선택방법, 극한확률점의 갯수, 그리고 평가순서 등이 확률판정의 성과에 유의한 영향을 준다는 근거는 희박하기 때문에[11, 14], 평가순서에 제약을 두지 않고 피실험자의 자유의사에 준해서 확률판정작업이 이

루어 지도록 하였다.

한편, Hogarth[8]와 Winkler[21]에 따르면, 전문가가 아닌(naive) 평가자들도, 최소한, 평가과제가 복잡하지 않는 상황들 속에서는 평가방법에 관한 통계적 훈련(statistical training)을 수행함으로써 자신의 판단을 확률형태로 일관성있게 표현하는 데 많은 도움을 받을 수 있다고 한다. 본 연구에서는 Solomon[14]에서 소개된 일관성 점검방법을 사용하여 피실험자들 자신이 평가한 확률관정이 일관성을 갖도록 배려하였다.

3.3. 실험의 운영

본 실험은 1988년 5월 27일과 동년 6월 3일 양 일간에 걸쳐서 진행되었다. 5월 27일에는 집단적 재평가법을, 6월 3일에는 명목집단법과 대표자 재평가법을 실험하였다.

실험시간은 각 집단과정별로 약 1시간~1시간 30분정도가 소요되었으며 실험은 수업분위기 속에서 진행되었다.

(1) 피실험자

최소한의 규범적 적합성(normative goodness : 평가자가 자신의 의견들을 확률형태로 표현하는 능력을 말함[20])을 갖도록 하기 위하여, 1988년 1월 1일 현재, 확률 및 통계학 관련 과목을 3학점 이상 이수한 동국대학교 산업공학과 3학년 이상의 학생들을 실험대상으로 선택하였다. 이들 중, 3학년생(36名)의 확률 및 통계학 관련 과목 이수학점 평균은 3.9학점(최소 3~최대 10 학점)이었고, 4학년생(18명)의 경우는 9.8학점(최소 3~최대 19 학점)이었다.

그리고, 각 집단과정에 대한 실험이 시작되기 전마다 제비뽑기를 통한 임의적(random) 방법을 사용하여 피실험자들을 각 집단에 배정함으로써 참가인원규모에 따른 실험상의 제약성을 극복하고자 하였다. 이러한 3회의 임의적인 배정방법을 통해서, 54명의 피실험자들만으로도 총 162명이 참

가하는 효과를 거둘 수 있다. 그러나, 6월 3일에 2명의 피실험자들이 불참함으로 인하여 2인집단의 수가 5집단에서 4집단으로 감소하였고, 결과적으로, 총 158명이 참여하는 효과를 거둘 수 있었다.

(2) 질문자료의 선택

본 실험에서 사용된 질문자료는, 1) 피실험자들이 비교적 많은 관심을 갖고 있고, 2) 그들 주변에서 일어나는 친근한 사건들이지만, 3) 어느 정도의 모호성(uncertainty)을 지니고 있어야 한다는 기준에 입각하여 선택되었다. 예를 들면, 축구 및 야구경기의 결과를 추측하는 문제, 기상예측과 관련된 문제 등을 들 수 있다. 이렇게 선정된 질문자료들은 각 집단과정(ETC, ETE, 또는, ETLE)별로 12문제씩 배정되었는데, 질문들의 내용이 평가결과에 미치는 효과를 극대화시킬 목적으로 난수(random number)를 이용한 임의적 방법을 사용하여 질문자료를 배정하였다.

(3) 실험의 운영

집단적 재평가법(ETC)에 관한 실험이 이루어진 5월 27일에는, 실험이 시작되기 전에 확률유도방법에 관한 3페이지 분량의 자료를 모든 피실험자들에게 제공했다. 본 실험자는, 이 자료를 통하여, 피실험자들에게 누적분포함수의 성질들을 설명하고, 한 가지 예를 들면서 'fractile'의 의미를 전달하려 하였고, 실제의 평가대상 문제들과 유사한 두 가지 예제들을 활용하여 평가방법에 관한 훈련을 실시하였다. 또한, 구두설명을 통해서 확률관정의 일관성 유지방법을 설명하였다. 설명이 끝난 뒤에는, 피실험자들에게 질문시간을 제공하여 평가방법에 익숙해지도록 배려하였다. 그리고, 6월 3일에는, 평가방법에 관한 특별한 설명없이 각 집단과정별 설문지에 첨부된 서면자료만으로 평가훈련을 대신했다. 또한 일관성 유지의 중요성 때문에, 그 유지방법을 구두로 설명함과 아울러, 설문

지의 응답란에 관련내용을 일부 추가시켰다.

3.4. 성과의 정확도 측정

본 연구에서는 다음과 같은 순서로 자료를 분석하였다. 먼저, 평가된 모든 확률분포들을 6개의 확률구간들 ($(-\infty, X_{.10})$, $(X_{.10}, X_{.25})$, $(X_{.25}, X_{.50})$, $(X_{.50}, X_{.75})$, $(X_{.75}, X_{.90})$, $(X_{.90}, +\infty)$)로 분할하고, 각 분포가 질문의 정답(또는, 실제결과)를 포함하는 구간을 찾는 방식으로 구간별 정답발생빈도수를 계산한다. 위의 구간별 기대확률은, 각각, .10, .15, .25, .25, .15, .10이다. 그리고, 관찰빈도와 기대빈도간의 편차를 파악한다는 측면에서 볼 때, 비모수 통계 방법의 일종인 적합도검정(Chi Square goodness-of-fit test)의 검정통계량 χ^2 을 본 연구의 검정통계량으로 사용하는 것이 적절하다고 본다.

$$\chi^2 = \sum_{i=1}^6 \frac{(O_i - E_i)^2}{E_i}$$

O_i : 구간 i 의 관찰빈도수,

E_i : 구간 i 의 기대빈도수,

E_i =(분석대상으로 고려된 총 문항수)

×(구간 i 의 기대확률)

여기서 $i=1, 2, \dots, 6$.

이상과 같은 구간별 기대확률을 토대로 하여 구간별 관찰빈도수와 기대빈도수간의 차이를 분석함으로써 집단구성원수, 집단과정, 평가단계 등과 같은 분석항목별로 교정(calibration) 정도를 파악할 수 있다. 이 방법은 누적분포함수 및 프렉타일법이 적용될 질문의 성격이 확률분포상의 대상문제와 직접(또는 간접)으로 관련된 몇 개의 고정값들에 대응하는 확률을 평가하는 형태가 아니라 확률분포상에 존재하는 몇 개의 프렉타일들을 평가할 경우(특히, 고려된 프렉타일들간의 구간확률이 서로 다를 경우)에 적절하게 사용될 수 있는 성과 척도로 생각되어 본 연구에서 제안하여 사용하고

자 한다. 실제로, 누적분포함수 및 프렉타일법을 사용한 대부분의 선행연구들은 전자와 같은 질문 형태를 취하고 있다[7, 16, 19].

원래의 평가과정에서는 $(-\infty, X_{.10})$, $(X_{.90}, +\infty)$ 대신에 $(-\infty, X_{.01})$, $(X_{.01}, X_{.10})$, $(X_{.90}, X_{.99})$, $(X_{.99}, +\infty)$ 를 사용함으로써 총 8개의 확률구간들을 고려하였음에도 불구하고 이를 6개의 구간들로 줄인 이유는, $(-\infty, X_{.01})$ 및 $(X_{.99}, +\infty)$ 와 같은 1% 구간들을 대상으로 한 검정통계량이 의미를 가질 수 있을 만큼의 충분한 수의 자료를 얻을 수 없었기 때문이다. 또한, 본 연구를 위한 실험자료 중에는 누적확률분포의 기본성질을 잘못 이해한 것으로 보이는 평가결과들이 나타나거나 평가가 안된 문항들(35/2820)이 발견되었다. 예를 들면, 0.01 프렉타일로 평가한 값이 0.10 프렉타일로 평가된 것보다 큰 값으로 평가하는 등의 현상이 일어났다. 이들은 분석대상에서 제외되었다.

누적분포함수의 프렉타일들에 대응하는 값들을 평가하는 경우에 일어날 수 있는 현상의 하나는, 하나의 확률분포상에 놓인 둘 이상의 프렉타일들이 동일한 값으로 평가되는 것인데, 만약 그 값이 정답(또는, 실제결과)에 해당한다면, 결과적으로, 한 질문항목에서 2개 또는 그 이상의 정답구간들이 발생하게 된다. 이러한 분석상의 문제점을 해결하기 위하여, 본 연구에서는 복수의 정답구간들이 발생하면 그 구간들 중 구간확률(최소 10%, 최대 25%)이 가장 큰 구간($(X_{.25}, X_{.50})$, $(X_{.50}, X_{.75})$)을 그 질문항목의 정답구간으로 정하였다.

또 한가지 문제는 $(X_{.25}, X_{.50})$ 와 $(X_{.50}, X_{.75})$ 가 동시에 정답구간으로 나타나는 경우다. 본 연구에서는, 각 질문항목이 구간정답수 산정시에 동일한 비중을 가져야 한다는 점을 고려하여, 정답구간으로 나타난 이들 두 구간에 대하여, 각각, 1/2의 비중을 할당하였다. 즉, 구간 i 만이 정답구간에 해당한다면 구간 i 의 관찰빈도수에는 1.0 단위만큼 가산된다. 반면에, 구간 i 와 구간 j 가 동시에 정답구간에 해당한다면, 구간 i 및 구간 j 의 관찰빈도

수는 각각 1/2단위씩 가산하고 구간 i 및 구간 j의 기대빈도수 계산시에도 1/2문항씩으로 고려하여 주게 된다.

한편, 각 집단과정별로 상이한 질문들을 고려하고 있기 때문에, 집단과정들의 성과를 비교하기 위해서는, 동일한 집단과정 내에서의 개인평가단계의 성과와 집단토의후 평가단계의 성과를 직접 비교하는 방법을 사용하는 것이 적절하다고 본다. 이러한 기준에 입각할 때, F-검정통계량이 본

연구의 분석목적과 부합되는 의미를 갖기 때문에, 이를 평가단계별 성과도의 비교수단으로 이용하였다.

$$\text{성과향상도 } K = \frac{x_{ji}^2 / V_1}{x_{j2}^2 / V_2}$$

여기서 자유도 $V_1 = V_2 = 5$,

x_{ji}^2 : 집단 j의 개인평가단계에서의 x^2 값,

x_{j2}^2 : 집단 j의 집단토의후 평가단계의 x^2 값,

4. 실험결과분석

4.1. 개인적 평가와 집단적 평가

표 4.1. 개인평가단계 및 집단토의후 평가단계의 성과비교

집 단 과 정	집 단 크 기	평 가 단 계	x^2 - value	검 정 결 과
E T C	2인집단	개 인 평 가	56.16	**
		토 의 후 평 가	20.81	**
	4인집단	개 인 평 가	92.41	**
		토 의 후 평 가	20.46	**
	7인집단	개 인 평 가	340.45	**
		토 의 후 평 가	23.17	**
E T E	2인집단	개 인 평 가	128.72	**
		토 의 후 평 가	171.66	**
	4인집단	개 인 평 가	203.97	**
		토 의 후 평 가	192.23	**
	7인집단	개 인 평 가	448.65	**
		토 의 후 평 가	341.60	**
E T L E	2인집단	개 인 평 가	74.34	**
		토 의 후 평 가	6.94	기각 못함
	4인집단	개 인 평 가	100.93	**
		토 의 후 평 가	21.10	**
	7인집단	개 인 평 가	191.00	**
		토 의 후 평 가	22.49	**

$H_0 : O_i = E_i \forall i$

'***' : $\alpha = .01$ 에서 귀무가설 H_0 를 기각함($x^2 .01(5) = 15.09$).

'기각 못함' : $\alpha = .05$ 에서 귀무가설 H_0 를 기각하지 못함($x^2 .05(5) = 11.07$).

평가자들이 개인적으로 평가한 확률분포의 성과와 집단토의를 하고 난뒤 평가한 분포의 성과를

비교하므로써 집단토의의 필요성을 조사하고자 한다. 표 4.1에서 볼 수 있듯이, 한 가지 경우를 제

외한 모든 경우들에서 확률판정이 부정확하게(ill-calibrated) 이루어 졌음을 알 수 있었다. 이 점은, 특히, 비전문가들(naive subjects)을 대상으로 하는 실험에서 두드러지게 나타나는 현상들이다. Alpert & Raiffa[1]는, 프렉타일법에 관한 실험을 통하여, 비전문가들로 구성된 피실험자들의 대부분이 실제결과치들과 비교할 때, 지나치게 '밀집

(tight)'한 분포들로 평가하는 경향이 있는 것으로 보고했다. 이러한 침도(tightness) 현상은 피실험자들이 .25, .50 및 .75 프렉타일들로는 너무 인접한 값들을 선택하고, .01 및 .99 프렉타일들로는 충분히 떨어져 있지 못한 값들을 할당하기 때문이다 [10, 11].

표 4.2. 밀집성(tightness) 현상의 비교

연구 대상		구간별 정답발생비율 (%)	
		IQ 구간	Surprise 구간
Alpert & Raiffa [1]		33	41
		43	23
Selvidge [10]		60	13
		54	24
		57	22
개인 단계	ETC	25.60	39.04
	ETE	17.12	43.89
	ETLE	31.09	31.41
토의 후 단계	ETC	30.77	25.64
	ETE	19.26	43.02
	ETLE	37.50	27.08

표 4.2는 선행연구결과들에서 나타났던 침도 현상을 본 연구의 실험결과와 비교분석하기 위한 자료이다.

만약 정확하게 확률판정이 이루어지는 경우라면, .25 프렉타일과 .75 프렉타일 사이에 참값이 포함될 확률은 50%일 것이며, 참값이 .01 프렉타일 및 .99 프렉타일들의 외부에 놓일 확률은 2%일 것이다(Alpert & Raiffa[1]는 전자를 'Interquartile(또는, IQ)' 구간, 후자를 'Surprise index(또는, Shock level)' 구간이라고 불렀다). 그런데, 선행연구결과들과 마찬가지로, 본 연구의 실험결과에서도 IQ 구간과 Surprise index구간의 관찰정답확률이 기대정답확률(각각, 50%와 2%)로 부터 크게 벗어나 있음을 알 수 있다. 한 가지 흥미있는

사실은, 집단과정별로 집계된 IQ 구간확률 및 Surprise index구간확률이 개인평가단계보다는 집단토의 후 평가단계에서 보다 개선된 결과를 나타냈다는 점이다.

이 결과로 부터, 집단토의가 평가성과에 영향을 주며, 최소한, 집단토의가 유의한 효과를 가져왔다는 잠정적인 결론을 얻을 수 있다고 본다. 그러나, 이 결과만으로는, 개인평가와 집단토의 후 평가에 대한 성과비교수단으로서 미흡하기 때문에, IQ 구간 및 Surprise구간을 포함한 모든 확률구간들을 대상으로 하여 개인평가단계의 성과(χ^2 값)에 대한 집단토의 후 평가단계의 성과의 향상도(K)를 구하고 이를 표 4.3에 나타내었다. 그런데, 본 연구에서는, 집단과정(ETC, ETE, ETLE)마다 서로

다른 질문내용들을 고려하고 있기 때문에, 질문내용의 상이함에 따른 분석상의 변수요인을 배제하기 위해서 동일한 집단과정내의 개인평가단계 및 집단토의후 평가단계의 성과향상도(K)를 집단과정들간의 성과비교수단으로 이용하였다.

표 4.3에서 보는 바와같이, 명목집단법(ETE)의

2인집단의 경우(K=.75)를 제외하면, 집단토의를 통해서 보다 향상된 성과를 얻은 것으로 분석되었다. 따라서, 평균적으로 볼 때, 개인평가단계보다는 집단토의후 평가단계에서 더욱 우수한 성과를 가져왔다는 점에서 선행연구결과들과 일치된 결과를 얻었다[3, 7, 9, 18, 19, 21].

표 4.3. 개인평가단계의 성과에 대한 집단토의후 평가단계의 성과향상도(K)

집단과정	집단구성원수		
	2인집단	4인집단	7인집단
ETC	2.70	4.52	14.70
ETE	.75	1.06	1.31
ETLE	10.71	4.78	8.49

K>1.0 : 집단토의후 평가단계의 평가성도가 더 우수함,

K<1.0 : 개인평가단계의 평가성도가 더 우수함,

K=1.0 : 개인평가단계와 집단토의후 평가단계의 성도가 동등함.

4.2. 집단구성원수의 효과

표 4.4. 집단과정과 집단구성원수간의 상관관계분석

고려된 집단과정	χ^2	검정결과	Contingency Coeff.
ETC & ETE & ETLE	6.77	기각 불합	.35
ETC & ETE	.76	기각 불합	.17
ETC & ETLE	6.37	*	.35
ETE & ETLE	.56	기각 불합	.14

H₀ : 집단구성원수는 집단과정의 성과에 영향을 주지 않는다.

'기각 불합' : $\alpha=.05$ 에서 귀무가설 H₀를 기각 불합.

'*' : $\alpha=.05$ 에서 귀무가설 H₀ 기각. ($\chi^2(2)=5.99$; $\chi^2(4)=9.49$)

표 4.4는 집단구성원수(group size)와 집단과정(group process)간의 상관관계를 알아 보기 위해서 표 4.3의 성과향상도(K)를 대상으로 독립성검정(χ^2 test for two independent samples)을 수행

한 결과를 나타낸 것이다[13, pp.104-111, 195-202].

먼저, 세 가지 집단과정들(ETC, ETE, 및 ETLE)을 모두 고려한 경우에는 집단구성원수의

변동이 성과향상도에 통계적으로 유의한 효과를 갖지 않는 것으로 나타났다. 이 결과는, 표 4.3에서도 알 수 있듯이, 명목집단법(ETE)의 집단구성원수에 따른 성과향상도(K)가 다른 두 집단과정들(ETC 및 ETLE)에 비해, 둔감한 변화를 보여 주었기 때문이라고 분석되었다. 이러한 결과해석은, 명목집단법을 제외한 나머지 두 집단과정들(ETC 및 ETLE)만을 대상으로한 독립성검정으로 부터 통계적으로 유의한 결과가 얻어졌기 때문에, 타당하다고 본다.

따라서, 최소한, 집단적 재평가법과 대표자 재평가법으로부터 집단구성원수의 증가에 따른 성과향상도의 변동을 기대할 수 있으며, 집단구성원수가 집단과정에 미치는 효과를 분석하는 일이 중요한 의미를 가질 수 있다고 본다. 이제, 위와 같은 결과들을 토대로 하여, 집단구성원수가 집단과정에 미치는 효과를 좀더 자세하게 분석하고자 한다.

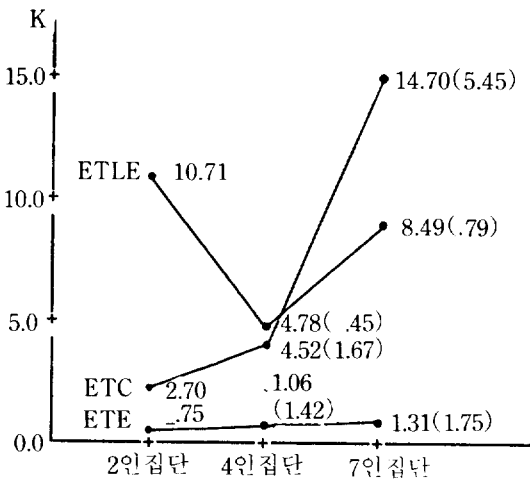


그림 4.1. 집단구성원수에 따른 성과향상도 K

()는 각 집단과정별로 2인 집단의 성과향상도를 기준으로 한 동일한 집단과정내의 4인 집단 및 7인 집단의 성과향상도들의 상대비율을 나타낸다.

그림 4.1은, 표 4.3에 나타난 성과향상도값들을 집단구성원수에 따라 분류하여 나타낸 것이다. 이 그림을 살펴 보면, 모든 집단구성원수들에 걸쳐서 집단적 재평가법(ETC) 및 대표자 재평가법(ETLE)들이 명목집단법(ETE)보다 우수한 성과를 가져왔음을 관찰할 수 있다. 집단구성원수가 증가함에 따라, 정도의 차이는 있지만, 성과가 개선되는 효과가 나타난 집단과정들은 집단적 재평가법과 명목집단법들이었다. 또한, 집단적 재평가법 및 명목집단법의 성과향상도를 비교해 볼 때, 명목집단법보다는 집단적 재평가법에서 집단구성원의 수가 증가함에 따른 성과개선효과가 더 큰(최소 1.7배, 최대 5.5배) 것으로 나타났다.

그리고, 명목집단법에서는, 실험에 포함된 모든 집단구성원수들(2인, 4인, 7인)에 걸쳐서, 성과향상도가 거의 대등하게 나타났는데, 이 결과는 2인 집단만으로도 4인 집단 및 7인 집단으로부터 얻어진 성과의 상당부분을 얻을 수 있었음을 뜻하며, Libby & Blashfield[9]의 연구에서도 이와 유사한 현상을 찾아 볼 수 있다.

그런데, 4인 집단을 고려하면, 명목집단에서는 비교적 꾸준한 성과향상을 가져온 반면에, 집단적 재평가법은 2인 집단 및 7인 집단에서의 성과와 비교할 때, 둔감하게 성과가 개선되었고, 대표자 재평가법에서는 2인 집단에 비해 급격한 성과하락 현상이 일어난 것으로 관찰되었다. 이 현상은 짝수인(4인 또는 6인) 집단들은, 동일한 크기를 갖는 두 개의 하부 집단들로 분할되는 것이 가능하기 때문에, 홀수인(3인, 5인, 또는 7인) 집단들에 비해서 논쟁(disagreement)이나 대립(antagonism)이 더욱 왕성하게 일어날 수 있다는 점에서 그 원인을 찾을 수 있을 것이다[12].

따라서, 집단구성원들이 합의에 도달해야 하는 집단적 재평가법이나 집단의 대표자가 그 집단의 의견을 대표하는 대표자 재평가법을 사용할 경우에는, 집단을 짝수인으로 구성하는 경우를, 가급적, 피하는 것이 성과저하를 예방할 수 있는 한

방편이라고 생각된다. 반면에, 명목집단법은, 비교적 성과는 떨어지지만, 구성원들간의 의견일치나 의견결충이 수반되는 집단과정들에서 일어날 수 있는 심리적 갈등요인을 배제할 수 있기 때문에 상대적으로 안정된 성과를 기대할 수 있다고 본다.

4.3. 결론

본 연구는 집단구성원수의 증감이 확률적 의견수렴방법의 성과에 미치는 효과를 파악함으로써 집단구성원수별로 적절한 의견수렴방법을 제시하는 것을 주목적으로 하고 있다.

본 연구에서 얻어진 결과들을 종합하면, 개별적 평가방법보다는 집단적 평가방법으로부터 더욱 개선된 성과를 얻을 수 있었다. 집단구성원수가 커짐에 따라 성과가 꾸준히 향상된 것으로 나타났던 방법은 명목집단법과 대표자 재평가법뿐이었다. 그러나, 명목집단법에서는 그 효과가 유의하지 않았던 것으로 나타났다. 그리고, 2인집단의 경우에는, 심리적 요인이 주는 영향을 무시할 수는 없지만, 대표자 재평가법이 가장 효과적이었다. 4인집단의 경우에는 대표자 재평가법과 집단적 재평가법으로부터 거의 대등한 성과개선 효과를 얻을 수 있었다.

를 얻을 수 있었다.

그러나, 대표자 재평가법에서는, 4인집단 및 7인집단의 성과향상도가 2인집단의 경우보다 뒤떨어진 것으로 나타났다. 따라서, 4인집단의 경우에는, 대표자 재평가법보다는 집단적 재평가법으로부터 집단구성원수의 증가에 따른 좀더 안정된 성과향상을 기대할 수 있을 것으로 본다. 7인집단에서는, 집단적 재평가법으로부터 가장 우수한 성과를 얻을 수 있었다.

그런데, 1) 현실상황의 집단이나 모임이 지속적 성격을 갖는 경우가 많은 데 반해 본 실험을 위한 모임의 성격자체가 일시적인 것이라는 점, 2) 피실험자들에게 평가성과에 따른 상금을 제공함과 아울러 평가에 신중을 기하도록 호소하는 등의 방법을 사용하여 본 실험에서 고려한 질문항목들이 피실험자들에게 현실적이고 중요한 의미를 지니도록 배려하였지만, 그 효과가 충분하였는지는 확신하기 어렵다는 점은 연구방법상의 한계점으로 지적되어야 할 것으로 본다. 따라서, 이러한 조건들을 모두 만족하는 실험상황을 구현하는 데에는 현실적으로 많은 어려움이 있다는 점도 간과할 수 없지만[16, p.140], 향후의 연구에서는 이와 같은 문제점들이 보완되어야 할 것이다.

참고문헌

1. Alpert, M., and Raiffa, H., *A Progress Report on the Training of Probability Assessors*, Unpublished Manuscript, Harvard University, 1969.
2. Ashton, A. H., and Ashton, R. H., "Aggregating Subjective Forecasts : Some Empirical Results". *Management Science*, 1985, Vol. 31, pp. 1499-1508.
3. Ashton, R. H., "Combining the Judgments of Experts : How Many and Which Ones ?". *Organizational Behavior and Human Decision Processes*, 1986. Vol. 38, pp. 405-414.
4. Berhold, M., "Procedures to Increase the Validity of Subjective Probability Estimates". *Decision Sciences*, 1975, Vol. 6, No. 4, pp. 721-730.
5. DeGroot, M. H., "Reaching a Consensus". *Journal of the American Statistical Association*, 1974, Vol. 69, pp. 118-121.

6. Einhorn, H. J., Hogarth, R. M., and Klempner, E., "Quality of Group Judgment". *Psychological Bulletin*, 1977, Vol. 84, pp. 158-172.
7. Gustafson, D. H., Shukla, R. K., Delbecq, A., and Walster, G. W., "A Comparative Study of Differences in Subjective Likelihood Estimates Made by Individuals, Interacting Groups, Delphi Groups, and Nominal Groups". *Organizational Behavior and Human Performance*, 1973, Vol. 9, pp. 280-291.
8. Hogarth, R. M., "Cognitive Processes and the Assessment of Subjective Probability Distributions". *Journal of the American Statistical Association*, 1975, Vol. 70, No. 350, pp. 271-289.
9. Libby, R., and Blashfield, R. K., "Performance of a Composite as a Function of the Number of Judges". *Organizational Behavior and Human Performance*, 1978, Vol. 21, pp. 121-129.
10. Pickhardt, R. C., and Wallace, J. B., "A Study of the Performance of Subjective Probability Assessors". *Decision Sciences*, 1974, Vol. 5, No. 3, pp. 347-363.
11. Selvidge, J. E., "Assessing the Extremes of Probability Distributions by the Fractile Method". *Decision Sciences*, 1980, Vol. 11, No. 3, pp. 493-502.
12. Shull, F. A. Jr., Delbecq, A. L., and Cummings, L. L., *Organizational Decision Making*, McGraw-Hill, New York, 1970.
13. Siegel, S., *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill, Tokyo, 1956.
14. Solomon, I., "Probability Assessment by Individual Auditors and Audit Teams : An Empirical Investigation". *Journal of Accounting Research*, 1982, Vol. 20, pp. 689-710.
15. Spetzler, C. S., and Staël von Holstein, C.-A. S., "Probability Encoding in Decision Analysis". *Management Science*, 1975, Vol. 22, pp. 340-358.
16. Staël von Holstein, C.-A. S., "Probabilistic Forecasting : An Experiment Related to the Stock Market". *Organizational Behavior and Human Performance*, 1972, Vol. 8, pp. 139-158.
17. Stone, M., "The Linear Opinion Pool". *Annals of Mathematical Statistics*, 1961, Vol. 32, pp. 1339-1342.
18. Winkler, R. L., "The Assessment of Prior Distributions in Bayesian analysis". *Journal of the American Statistical Association*, 1967, Vol. 62, pp. 1105-1120.
19. Winkler, R. L., "The Consensus of Subjective Probability Distributions", *Management Science*, 1968, Vol. 15, pp. B61-B75.
20. Winkler, R. L., and Murphy, A. H., "'Good' Probability Assessors". *Journal of Applied Meteorology*, 1968, Vol. 7, pp. 751-758.
21. Winkler, R. L., "Probabilistic Prediction : Some Experimental Results". *Journal of the American Statistical Association*, 1971, Vol. 66, pp. 675-685.
22. Winkler, R. L., and Cummings, L. L., "On the Choice of a Consensus Distribution in Bayesian Analysis". *Organizational Behavior and Human Performance*, 1972, Vol. 7, pp. 63-76.