韓國數學教育學會誌 시리즈 A <數學教育>
https://doi.org/10.7468/mathedu.2018.57.4.413
2018. 11. 제 57권, 제 4호, 413-431.

J. Korean Soc. Math. Ed. Ser. A:
The Mathematical Education
November. 2018, Vol. 57, No. 4, 413-431.

‘닫힌 상자’에서의 복원추출에 의한 모비율 추측 활동수업 개발 및 적용

이기돈(경인고등학교 교사)

Ⅰ. 서론

통계적 사고의 핵심 요소로서 대표성과 더불어 변이성이 주목받고 있다(고은성, 이경화, 2010; 이은희, 김원경, 2015; 이형숙, 이경화, 김지원, 2010; 한채린, 이경원, 김도연, 배미선, 권오남, 2018; Pfannkuch & Wild, 2004; Wild & Pfannkuch, 1999). 자료를 대표하는 수치를 파악하는 것 못지않게 그 자료의 변이적 특성을 이해하는 것이 통계적 사고의 주요한 부분이라는 것이다. 한편 통계적 추정이나 검정의 토대가 되는 표본 및 표집에 대한 논의도 활발하다. 이러한 논의에서도 표본 대표성과 표집 변이성이 주요하게 다루어진다(고은성, 2012; 탁병주, 구나영, 강현영, 이경화, 2014; 탁병주, 구나영, 강현영, 이경화, 2017).

표본은 모집단의 특성을 파악하기 위해 추출하는 것이기 때문에 그에 대한 교육적 논의는 모수의 추정에 대한 이해를 염두에 두게 된다. 표본 대표성이나 표집 변이성도 표본 통계량을 기반으로 모수를 추정하는 과정에서 중요하게 고려해야 하는 개념들이다. 국내의 연구들(구나영, 탁병주, 강현영, 이경화, 2015; 박민선, 2015; 이경화, 지은정, 2005; 이정연, 이경화, 2017)은 표본 및 표집을 현재보다 이른 시기부터 교수학습 할 수 있다는 점과 이를 통한 비형식적 통계적 추리 능력의 함양에 관심을 가져왔다. 통계적 추정은 모집단, 표본과 표집, 대표성과 변이성, 확률분포 등과 같은 여러 가지 개념들의 복합적이고 계량화된 이해를 필요로 한다. 보다 이른 시기부터 표본 대표성과 표집 변이성을 경험하고 이를 통해 비형식적 추리 능력을 함양하는 것은 표준적인 통계학의 원리에 따라 계량적으로 통계적 추정을 교수학습하는 데 있어 도움이 될 것으로 생각된다.

한편 통계적 추정이나 검정에서 표집 변이성의 파악은 표본 통계량의 확률분포에 대한 이해를 필요로 한다. 연구들(Arnold, Pfannkuch, Wild, Regan, & Budgett, 2011; Saldanha & Thompson, 2002; Wild, Pfannkuch, Regan, & Horton, 2011)은 이러한 이해를 지원하기 위해 하나의 모집단으로부터 같은 크기의 표본을 여러 번 추출해 보는 것과 같은 활동을 제안하였다.

본 논문에서는 2018년 현재 고등학교 2・3학년 학생들이 학습하는 2009 개정 교육과정의《확률과 통계》과목에서 표본비율로부터 모비율을 구간추정하는 내용의 교수학습에 대해 논의한다. 이 과목에서 모비율은 복원추출에 의한 표집이라는 이론적 가정 하에 중심극한정리에 기반하여 표본비율과 표본의 크기로부터 적절한 신뢰도를 갖는 신뢰구간을 계산함으로써 계량적으로 추정된다. 교과서에서는 모평균 추정과 관련된 표본평균의 확률분포를 이해하는데 도움이 되는 몇 가지 활동들을 제시하고 있지만 모비율 추정에 대해서는 그렇지 못한 편이다. 본 연구의 목적은 형식적 모비율 추정의 이해를 지원하는 활동 중심의 수업을 개발하고 실제로 적용하여 교수학습적 측면의 시사점을 논의하는 것이다. 이를 위해 다음과 같은 연구과제를 설정하였다.

첫째, 모비율 추정의 이해를 표본 대표성과 표집 변이성의 측면에서 지원하는 활동수업을 개발한다.

둘째, 개발된 수업을 실제로 적용하여 그 결과를 표본 대표성과 표집 변이성의 측면을 중심으로 분석한다.

셋째, 분석 결과가 모비율 추정의 이해를 지원하는 활동수업의 교수학습에 주는 시사점을 논의한다.

모비율 추정은 모평균 추정과 달리 2015 개정 교육과정에서는 일반선택 과목인《확률과 통계》가 아니라 전

* 접수일(2018년 10월 15일), 수정일(2018년 11월 10일), 게재확정일(2018년 11월 21일)
* ZDM분류 : D74, U64
* MSC2000분류 : 97D40, 97U60
* 주제어 : 모비율 추정, 활동수업, ‘닫힌 상자’, 복원추출, 물리적 시뮬레이션
* 논문의 아이디어와 결과 분석에 도움을 주신 박지현 선생님과 이규희 선생님께 감사드립니다.

문교과인《심화수학Ⅱ》에서 다루어지기 때문에 이 내용을 학교수학에서 학습하는 학생들은 현재보다 줄어들 것으로 예상된다. 그러나 매주 발표되는 대통령 지지율과 같이 모비율 추정은 모평균 추정 못지않게 생활 속에서 흔하게 접하게 되는 상황으로서 그와 관련된 통계적 소양의 교육은 창의적 체험활동 등을 통해서도 이루어질 수 있다고 생각된다. 본 연구는《심화수학Ⅱ》의 교수학습뿐 아니라 그러한 교육활동에도 참고가 될 수 있을 것이다.

Ⅱ. 이론적 배경

1. 모비율 추정에서의 표본: 준비례적 축소판으로서의 표본과 복원추출에 의한 표본

Saldanha와 Thompson(2002)은 '모집단의 준비례적 축소판(a quasi-proportional, small-scale version of the population)'이라는 표본 개념을 교수학습의 지향점으로 주목하였다. 하나의 모집단에서 얻어지는 같은 크기의 여러 표본들의 특성들은 모집단의 특성을 준비례적으로 축소하여 이어받는 것으로서 모집단과 유사하지만 다양할 수 있다는 것이다. 이러한 개념은 비슷한 조건에서 여러 세트의 표본들이 추출될 수 있다는 인식에 기반한다. 이때 '축소판'은 표본이 모집단의 특성을 준비례적으로 이어받은 부분집합들 중 하나라는 의미로서 비복원추출에 의한 표집을 가정하는 것으로 보인다. 실제로 Saldanha와 Thompson(2002)은 물론 이들의 표본 개념을 참고하는 여러 연구들(고은성, 2012; 박민선, 고은성, 2014; 이경화, 지은정, 2005; 탁병주 외, 2014)도 비복원추출에 의한 표본을 바탕으로 논의하였다.

이러한 표본 개념은 표준적인 통계학의 원리에 따라 모비율을 구간추정할 때는 보완이 필요하다. 모비율의 신뢰구간의 계산은 표본비율의 확률분포를 묘사하는 중심극한정리에 근거하여 이루어진다. 중심극한정리에 의하면 표본비율은 표본의 크기가 충분히 클 때 근사적으로 특정한 정규분포를 따른다. 이에 대한 2009 개정 교육과정의《확률과 통계》교과서 총 9종의 설명은 모두 유사한데 다음과 같이 요약할 수 있다.

모비율이 p인 모집단에서 크기가 n인 표본을 택할 때, 표본비율 $\hat{p}=\dfrac{X}{n}$를 생각하자.

이때, 확률변수 X는 어떤 사건이 일어날 확률이 p인 시행을 n번 독립시행하였을 때, 그 사건이 일어난 횟수이므로 이항분포 $\mathrm{B}(n,p)$를 따른다. – (*)

따라서 $\mathrm{E}(X)=np,\ \mathrm{V}(X)=npq$ (단, $q=1-p$)

$$\mathrm{E}(\hat{p})=\mathrm{E}\left(\frac{X}{n}\right)=\frac{1}{n}\mathrm{E}(X)=p$$

$$\mathrm{V}(\hat{p})=\mathrm{V}\left(\frac{X}{n}\right)=\frac{1}{n^2}\mathrm{V}(X)=\frac{pq}{n}$$

이항분포의 정규분포 근사에 의해

$X \sim \mathrm{N}(np, npq)$ 이고 $\hat{p}=\dfrac{X}{n}\sim \mathrm{N}\left(p,\dfrac{pq}{n}\right)$

따라서 $Z=(\hat{p}-p)/\sqrt{\dfrac{pq}{n}}\sim \mathrm{N}(0,1)$

또한 $Z=(\hat{p}-p)/\sqrt{\dfrac{\hat{p}\hat{q}}{n}}\sim \mathrm{N}(0,1)$ (단, $\hat{q}=1-\hat{p}$)

이때 $\mathrm{P}\left(-1.96\sqrt{\dfrac{\hat{p}\hat{q}}{n}}\le \hat{p}-p\le 1.96\sqrt{\dfrac{\hat{p}\hat{q}}{n}}\right)=0.95$ – (**)

이로부터 모비율 p에 대한 신뢰도 95%의 신뢰구간은

$\hat{p}-1.96\sqrt{\dfrac{\hat{p}\hat{q}}{n}}\le p\le \hat{p}+1.96\sqrt{\dfrac{\hat{p}\hat{q}}{n}}$ – (***)

(*)과 같이 확률변수 X를 n회의 독립시행에서 특정 사건이 일어난 횟수로 보기 위해서는 밑줄 친 부분에서의 표본이 복원추출에 의한 것이어야 한다. 즉, 모비율 추정의 원리에 대한 교과서의 설명은 복원추출에 의한 표집을 가정한다. 그러나 실제 문제 상황의 추정 과정에서는 복원추출이 아닌 비복원추출에 의한 표본을 사용한다. 비복원추출에 의한 표집을 가정하여 표본 통계량의 확률분포를 유도하는 것은 매우 복잡하고 어려운 일이다. 반면 모집단의 크기가 표본의 크기에 비하여 충분히 큰 경우 비복원추출에 의한 표본의 성질이 복원추출에 의한 것과 유사하다. 때문에 복원추출에 의한 표본 통계량의 확률분포를 비복원추출에서의 확률분포의 근사분포로 보고 실제 모비율 추정 상황에 적용하여 추정한다(김우철 외, 1996).

교과서에서는 생각열기 등에서 추적 장치를 부착한 연어들의 회귀율을 통해 방류한 전체 연어의 회귀율을 추측하는 것(황선욱 외, 2014)과 같이 비복원추출 상황을

제시하면서도 모비율 추정의 원리에 대한 설명에서는 복원추출을 가정한다. 그러나 이러한 불일치를 보완하는 설명을 서술하지는 않는다. 비복원추출에 의한 표본이라는 실제와 복원추출에 의한 표본이라는 이론 사이의 틈을 조화시키는 교수학습적 조치가 필요하다.

2. 표본비율의 대표성과 큰 수의 법칙

모집단의 준비례적 축소판으로서의 표본 관점에서는 표본이 모집단의 특성을 준비례적으로 이어받을 것이기 때문에 표본비율이 모비율과 정확히 같지는 않아도 유사할 것이라고 추측할 수 있다. 반면 복원추출에 의한 표본의 관점은 표본이 모집단의 특성을 이어받는다는 것을 직접적으로 내포하지 않는다.

관심 있는 성질 A를 가진 개체의 비율(모비율)이 p인 모집단에서 크기가 n인 표본을 복원추출한다고 하자. 이러한 표집은 n회의 독립시행으로 볼 수 있다. 이 독립시행을 이루는 i번째 시행에서 선택된 개체가 성질 A를 가지면 $X_i = 1$, 그렇지 않으면 $X_i = 0$이라고 하자. 이때 표본비율 $\hat{p} = \dfrac{X}{n} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$은 성질 A를 가진 개체가 선택된 횟수의 전체 시행 횟수에 대한 비율, 즉 상대도수이다. 상대도수는 한 번의 시행에서 성질 A를 가진 개체가 선택될 사건의 가망성(chance)을 해석하는 한 가지 방법으로서 다른 한편으로 그 가망성을 수학적 확률, 즉 모비율 p로 해석하면 표본비율과 모비율이 유사할 것이라고 추측할 수 있다(이기돈, 2018a). 이러한 추측은 경험을 통해 가망성을 파악하는 자연스러운 사고에 기반한 것으로서 전체 시행 횟수 n이 증가하면 더 큰 신뢰를 준다(Sedlmeier, 1999).

한편 ‘큰 수의 약한 법칙’은 임의의 ϵ (>0)에 대하여

$$\lim_{n \to \infty} \mathrm{P}\left(\left| \frac{X_1 + \cdots + X_n}{n} - p \right| \geq \epsilon \right) = 0$$

이라고 서술된다. 즉, 표본비율 $\hat{p} = \dfrac{X}{n}$는 모비율 p에 확률수렴한다. 그러나 이것은 ‘$N > n$인 시행 횟수 N과 n에 대하여 $\left| \dfrac{X}{N} - p \right| < \left| \dfrac{X}{n} - p \right|$’를 의미하는 것은 아니다. 이것은 큰 수의 법칙의 이해에서 흔히 발생하는 오개념으로서 교수학습 시에 유의할 필요가 있다(신보미, 이경화, 2006; 최인용, 조한혁, 2017).

3. 모비율 추정에서의 표집 변이성과 중심극한정리

위에서 언급한 바와 같이 큰 수의 법칙은 표본의 크기가 클수록 표본비율이 모비율의 더 좋은 추정량이라는 것을 보장하지 못한다(Sedlmeier, 1999). 매우 큰 크기의 표본을 추출하여 그 표본비율의 값을 모비율로 삼는다고 하여도 그것이 반드시 실제 모비율의 값에 가까운 것은 아니다. 하지만 무작위 표집의 경우 중심극한정리에 의해 1절의 (**)와 같이 표본비율이 모비율과 가까울 확률을 계산할 수 있다. 이것은 무작위로 여러 세트의 표본들을 표집할 때 그 표본비율들의 변이가 실제 모비율을 포함하는 그 구간에서 나타날 확률이다. 이에 따라 하나의 값이 아니라 모비율이 존재할 만한 구간에 관심을 갖고, (***)와 같이 해당 확률만큼의 신뢰도 하에 그 구간을 추정할 수 있다.

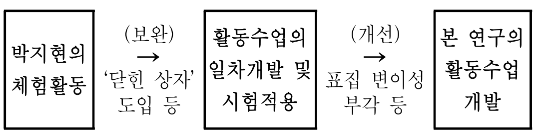
Ⅲ. 연구 방법

1. 모비율 추측 활동수업의 개발 과정

박지현(2016)은 체험 중심의 고등학교 통계 교수학습 자료개발 연구과제를 수행하면서, 상자 속에 들어 있는 세 개의 공 중 몇 개의 공이 빨간색인지를 한 번에 공 하나씩을 꺼내어 관찰하고 다시 상자에 집어넣는 실험을 반복함으로써 추측하는 활동을 개발하고 체험부스 운영 등의 형태로 적용하였다. 이와 같이 두 가지 이상의 색 공들이 들어 있는 항아리나 상자에서 무작위로 하나의 공을 꺼내어 색을 확인하는 상황 설정은 실제 확률적 문제 상황을 단순화하여 표현한 것으로서 ‘항아리 모델(urn model)’이라고 알려져 있다. 이 모델은 ‘Polya의 항아리 모델(Polya urn model)’과 같이 경제학, 물리학, 유전학 등에서 확률적 문제 상황들을 표현하고 연구하는 도구로 활용되어 왔다(DasGupta, 2010; Xue, 2006). 특히 항아리에서 하나의 공을 반복하여 꺼낸 결과로부터 항아리 안의 실제 색공 개수의 비율을 알아내고자 하는 것과 같이, 동일한 시행을 여러 번 반복한 결과로부터 그 원인을 탐구하는 것은 Thomas Bayes 등에 의해 연구되었던 ‘역확률(inverse probability) 문제’에 포함된다(Dale, 2012). 본 연구에서의 활동수업은 항아리 모델에서의 역확률 문제를 교수학적으로 변환한 박지현(2016)의 체험활동을 모티브로 개발되었다. 특히 복원추출에 의한 표

집 활동을 구현하는 데 있어 항아리 모델을 활용한 점을 차용하였다.

본 연구에서의 활동수업은 두 단계로 개발되었다([그림 1]). 첫 번째 단계에서는 박지현(2016)의 체험활동을 보완하여 교과서의 모비율 추정의 이해를 지원하는 활동수업을 일차적으로 개발하고 이를 시험적으로 적용하였다. 박지현(2016)의 체험활동은 복원추출에 의한 표본비율로부터 모비율을 추측하는 것이지만 실제 모비율 추정 상황의 모형화, 모비율 추정의 이해의 지원, 수업에의 적용 등에 있어서는 부족한 측면이 있었다. 이와 관련하여 본 연구에서는 '닫힌 상자'의 도입, 이에 따른 공의 색을 관찰하는 방법의 구안, 상자 안에 넣는 공의 개수의 조절, 양자 대결 구도의 설정 등의 개선책을 포함한 활동수업을 개발하였다. 시험적 적용은 연구자가 2017학년도 2학기《확률과 통계》과목을 가르쳤던 서울 구로구에 위치한 어느 고등학교 2학년 인문사회과정 3개 학급을 대상으로 모비율 추정을 비롯한 이 과목의 모든 내용의 지도가 끝난 후 학기말 수업 시간을 이용해 이루어졌다. 학생들은 대체로 이 활동수업에 흥미를 갖고 참여하였다. 또, 표본비율의 대표성에 기반하여 모비율의 값을 맞히거나 틀리는 과정에서 이미 학습한 모비율 추정에 대한 이해가 깊어질 수 있다는 것을 알 수 있었다(경상남도교육청, 2018).



[그림 1] 활동수업의 개발 과정
[Fig. 1] The development process of this activity lesson

두 번째 단계에서는 첫 번째 단계에서 개발한 수업을 기반으로 모비율 추측 과정에서 발생하게 되는 표집 변이성을 부각시키는 교수학습 장치를 마련하여 활동수업을 개선하였다. 이후 서술되는 활동수업의 내용과 적용 등은 두 번째 단계까지를 통해 개발된 활동수업에 대한 것이다. 이 활동수업의 내용은 표준적인 통계학의 원리에 따른 모비율 구간추정을 교수학습하기 전과 후 어느 때든 적용하여 그 이해를 지원할 수 있을 것으로 보인다. 본 연구에서는 그러한 교수학습이 이루어지기 전의 학생들을 대상으로 적용하였다.

2. 모비율 추측 활동수업의 적용 및 분석 방법

1) 적용 대상 및 환경

위와 같이 개발된 활동수업은 서울의 강남구에 위치한 어느 고등학교의 수학과학 통합 영재학급 1·2학년 학생들을 대상으로 2018년 8월 13일 연구자의 교수를 통해 적용되었다. 1학년(19명) 수업은 오전에, 2학년(17명) 수업은 오후에 각각 100분 정도로 진행되었다. 1학년은 2015 개정 교육과정의 적용을 받는 학생들로서 학교에서 아직《확률과 통계》과목을 학습하지 않았고, 2학년은 2009 개정 교육과정의 적용을 받는 학생들로서 1학기에《확률과 통계》과목을 '확률의 뜻과 활용'까지 학습하였으나 모비율 추정을 비롯한 '통계' 부분은 아직 다루지 않은 상태였다.

이론적 배경에서 살펴본 바와 같이 표본비율의 대표성과 표집 변이성은 모비율을 확률적이고 계량적으로 구간추정하게 되는 동기를 제공한다. 때문에 본 연구에서는 두 가지 특성에 대한 경험이 형식적 모비율 추정의 이해를 지원한다고 보았다. 그리고 개발한 활동수업이 복원추출에 의한 표집에 기반하여 표본비율의 대표성에 대한 성찰의 계기를 제공하고 표집 변이성을 부각시킴으로써 그러한 경험과 이해를 지원할 것으로 기대하였다. 이에 따라 이 활동수업을 적용하였을 때 학생들이 실제로 그 두 가지 특성을 경험할 수 있는지를 중심으로 그 결과를 파악하고자 하였다.

2) 자료 수집 및 분석 방법

이 활동수업에서는 학생들로 하여금 활동지(<부록> 참조)를 작성하도록 하였다. 이때 작성한 내용 중 일부는 발표를 통해 다른 학생들과 공유하였다. 이를 바탕으로 총 2종의 연구 자료를 수집하였다. 자료1은 활동지 3번 항목(추측하기)에 대한 각 모둠들의 발표를 녹취하여 전사한 것이다. 이 자료는 복원추출에 의한 표본비율을 바탕으로 추측한 모비율의 값과 그렇게 추측한 이유를 내용으로 한다. 자료1의 분석에서는 학생들이 작성한 활동지 기록도 참고하였다. 자료2는 활동지 4번 항목(활동 되돌아보기)에 대해 학생들이 작성한 기록으로서 수업 종료 후 활동지를 수거하여 수집하였다.

이렇게 수집한 총 2종의 자료들을 '질적 내용분석(Qualitative Content Analysis)'의 방법으로 분석하였다.

Krippendorff(2004, 최성호, 정정훈, 정상원, 2016에서 재인용)와 Elo & Kyngäs(2008, 최성호 외, 2016에서 재인용), 그리고 홍서영과 서태열(2014)을 참고하여 분석 결과의 타당도를 높이기 위해 단위화, 코딩, 범주화라는 내용분석 과정을 설정하여 실시하였다. 단위화와 관련하여, 자료1 전체를 하나의 분석 단위로 설정하되, 학생들이 발표하고 있는 중에 연구자가 그렇게 추측한 이유 등을 재차 물어 반성적 사고를 유도한 경우에는 그 지점을 기준으로 전과 후를 각각 하나의 분석 단위로 설정하였다. 자료2는 활동지 4번 항목의 첫 번째 소항목 4.가.(궁금한 점)와 두 번째 소항목 4.나.(느낀 점)를 각각 하나의 분석 단위로 설정하였다. 범주화에 있어서는, 이 활동수업의 적용을 통해 표본비율의 대표성과 표집 변이성에 대한 경험이 가능한지를 파악하고자 하였기 때문에 그 두 가지 특성에 대한 자각 또는 인지의 양상을 주 범주로 설정하였다. 그리고 표본비율의 대표성에 대해 이론적 배경에서 살펴본 내용과 이 활동수업에서 표집 변이성을 부각시키는 교수학습 장치의 내용에 따라, 코딩 이전에 일부 하위 범주를 설정함으로써 개념 지향적(concept-driven) 접근 방법을 적용하였다. 더불어 범주에 따른 자료 코딩 과정에서 자료의 내용에 따라 기존 범주를 수정하거나 새로운 범주들을 추가하는 방식으로 자료 지향적(data-driven) 접근 방법도 활용하였다. 또, 분석 결과의 신뢰도를 높이기 위해 Schreier(2012)를 참고하여 연구자를 포함한 2명의 수학교육 전문가가 독립적으로 일차 코딩 및 범주화를 수행한 후 협의를 통해 하위 범주를 통일하였다. 이후 통일된 하위 범주를 기준으로 각자 새로이 코딩을 수행한 후 합의의 계수가 1이 될 때까지 충분히 협의하였다.

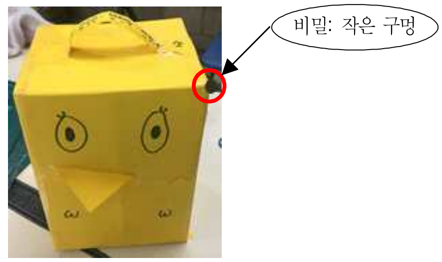
한편 연구자는 수업을 진행한 교사로서 수업 직후 실제로 실행된 수업 내용, 과정, 결과, 분위기, 느낌 등을 자유롭게 수업일지로 작성하였다. 그리고 이것을 자료1과 자료2의 분석 및 해석에 참고하였다.

Ⅳ. 결과 분석 및 논의

1. 모비율 추측 활동수업의 개발 결과

1) 수업의 개요

이 활동수업에서 학생들에게 부여되는 주 과제는 색지로 완전히 포장되어 열거나 내부를 들여다 볼 수 없는 종이 상자(이하 '닫힌 상자', [그림 2])에 들어 있는 주황색 탁구공 개수의 전체 탁구공 개수에 대한 비율을 알아맞히는 것이다. 학생들이 활동 과정에서 전체 탁구공 개수가 10개라는 것을 알게 되기 때문에 주황색 탁구공 개수를 맞히면 사실상 과제를 해결하게 된다.



[그림 2] '닫힌 상자'의 예
[Fig. 2] An example of 'Closed Box'

이 수업의 도입 부분에서는 과제 및 그 의미를 제시하고, 전개 부분에서는 '닫힌 상자'를 제작하여 과제를 해결하고 그 결과를 발표하며, 정리 부분에서는 이상의 활동을 반성적으로 살핀다([그림 3]).

도 입	·'닫힌 상자'와 그 비밀 소개 및 과제 도입 ·현실 문제 상황과의 유사점 및 차이점 언급
↓	
전 개	·모둠별로 '닫힌 상자' 제작(팀별로 지정된 개수의 주황색과 흰색 탁구공 사용) ·상대팀 '닫힌 상자'에서의 복원추출에 의한 표집 ·상대팀 '닫힌 상자'의 주황색 공 개수의 비율 추측 ·추측 결과 및 이유 발표
↓	
정 리	·활동 되돌아보기(궁금한 점, 느낀 점) 작성

[그림 3] 활동수업의 주요 흐름
[Fig. 3] Main flow of this activity lesson

각 부분의 구체적인 내용 및 방법은 아래와 같다.

2) 도입, 전개, 정리의 내용 및 방법

도입 부분에서는 '닫힌 상자'와 그 비밀을 소개하면서 학생의 질문을 통해 주 과제를 도입하고(이기돈, 2018b), 현실 문제 상황과의 유사점 및 차이점을 언급한다. 우선,

기 제작된 '닫힌 상자'의 실물을 소리가 나도록 흔들어 제시함으로써 그 안에 주황색과 흰색 탁구공 총 10개가 들어 있지만 색지로 완전히 포장되어 열거나 내부를 들여다 볼 수 없다고 알려준다. 이런 상황에서 궁금한 질문을 생각해보도록 한다. 기대하는 질문은 '그 상자 안의 두 가지 색 각각의 탁구공의 개수 또는 그 비율은 얼마인가요?'이다. 그리고 한 꼭짓점에 작은 구멍이 뚫려 있어서 상자를 위로 들어 올려 무작위로 하나의 탁구공이 그 꼭짓점에 위치하도록 한 다음 그 구멍을 통해 탁구공(이후 '공')의 색을 관찰할 수 있다는 '닫힌 상자'의 비밀([그림 2])을 공개한다. 이제, '닫힌 상자'의 비밀을 통해 궁금한 질문을 해결할 수 있다는 방식으로 이 활동수업의 주 과제를 도입한다. 그리고 이것이 대통령 지지율이나 생산품의 불량률 조사와 같이 무작위로 추출된 표본의 정보에 근거하여 모비율을 추측한다는 점에서 현실 문제 상황과 유사하지만, 과제는 복원추출에, 현실 문제 상황은 비복원추출에 의해 표집한다는 차이가 있고, 그럼에도 모집단의 크기가 클 경우 복원추출에 의한 표집에서도 비복원추출과 마찬가지로 동일한 대상이 다시 추출되지 않을 가능성이 높다는 점을 언급한다.

전개 부분에서는 모둠별로 '닫힌 상자'를 제작한 후 다른 모둠이 제작한 상자 안의 주황색 공(이후 '색공') 개수의 전체 공 개수에 대한 비율(이후 '색공 개수의 비율')을 알아맞히는 과제를 수행하고 표본비율 및 추측한 색공 개수의 비율과 그 이유를 발표한다. 보다 구체적인 내용, 방법, 순서는 다음과 같다.

・2인 1모둠(또는 1인 1모둠) 구성

・5개 모둠(1조~5조)을 홍팀으로, 나머지 5개 모둠(6조~10조)을 백팀으로 배정

・상대팀 '닫힌 상자' 안의 색공 개수의 비율을 맞힐 경우 모둠별로 간단한 강화물이 제공됨을 공지

・모둠별로, 팀별로 지정된 개수의 색공(예를 들어, 홍팀은 8개, 백팀은 4개)을 포함한 총 10개의 공을 종이 상자에 넣은 후 조립(각 모둠은 자기 상자에 넣는 색공의 개수가 상대팀에게 노출되지 않도록 유의)

・모둠별로, 조립한 종이 상자를 색지로 포장하여 밀봉

・모둠별로, 밀봉한 종이 상자의 한 꼭짓점에 작은 구멍을 뚫어 '닫힌 상자' 완성

・홍팀의 각 모둠과 백팀의 각 모둠이 짝을 지어 '닫힌 상자' 교환(예를 들어, i조와 (i+5)조가 교환)

・모둠별로, 교환받은 상대팀 '닫힌 상자'에 뚫린 구멍을 통해 한 번에 하나씩 무작위로 선택된 공의 색 관찰([그림 4])

・모둠별로 이 시행을 원하는 횟수만큼 실시(활동지 1번 항목)한 후 그 결과를 요약(활동지 2번 항목)하고 이를 바탕으로 상대팀 '닫힌 상자' 안의 색공 개수의 비율 추측(활동지 3번 항목) - (※)

・모둠별로 홍팀과 백팀이 번갈아 가며 표본비율 및 추측한 색공 개수의 비율과 그 이유 발표(동시에 모든 모둠의 표본비율과 추측 비율을 교실 전면 화면에 공시)

・'닫힌 상자' 안의 색공 개수의 비율의 참값 공개(모둠별로 '닫힌 상자'를 파손 및 개봉하여 참값 확인도 가능) 및 맞힌 모둠에 강화물 제공

[그림 4] 뚫린 구멍을 통해 무작위로 선택된 공의 색을 관찰하는 학생

[Fig. 4] A student watching the colour of a ball chosen at random

이때 (※)은 이영하, 이은호(2010)와 구나영 외(2015)를 참고하여 통계적 탐구 과정을 자료 수집, 요약 및 분석, 해석 및 추론으로 보고 이 활동수업의 과제에 맞추어 구체화한 것이다. 이 과정에서 학생들이 표본비율의 대표성을 자각 및 성찰할 수 있을 것으로 기대하였다.

정리 부분에서는 개인별로 활동지 4번 항목(활동 되돌아보기)을 작성하도록 하여 이상의 활동을 반성적으로 살펴보도록 한다. 아마 학생들은 실질적으로 동일한 상대팀의 다섯 '닫힌 상자' 각각에 포함된 색공 개수의 비율을 추측하기 위해 자기팀 다섯 모둠에서 추출한 표본의 비율들이 모두 같지는 않다는 것을 경험하게 될 것이다. 4번 항목 중 4.가.(궁금한 점)에서는 학생들이 그 경

험을 통해 모비율 추측 시 무작위 추출 과정에서 발생하는 표집 변이성의 곤란함이나 적절한 처리 방법 등을 언급하게 될 것으로 기대하였다. 활동지 4번 항목의 작성이 종료된 후 그러한 기대가 있었음을 학생들에게 이야기하였다. 또, 모비율과 정확히 같기는 어렵다는 표본비율의 대표성의 한계와 무작위 추출에 의한 표집 변이성을 고려하여 확률적이고 계량적으로 모비율을 구간추정하는 원리에 대해 추후《확률과 통계》등에서 학습할 수 있음을 언급하였다. 한편 4.나.(느낀 점)를 통해서는 이 활동에 대해 4.가. 외에도 있을 수 있는 여러 가지 성찰들을 유도하고자 하였다.

3) 수업의 주안점

(1) 미지의 모집단과 현실 문제 상황을 모사하는 '닫힌 상자'의 사용

탁병주 외(2014)는 표본 개념을 역사적으로 분석하면서 그 등장 배경으로 사회적 요구와 미지 세계의 추측에 대한 갈망을 꼽았다. 표본을 추출하여 모집단의 정보를 추측하는 것은 전수 조사에 따르는 사회적 비용을 고려한 것이기도 하지만, 한편으로 선거에서의 출구조사와 같이 미지의 세계를 알고자 하는 심리에도 기인한다는 것이다. 위 활동수업에서는, 색공을 포함한 10개의 공이 들어 있지만 색지로 밀봉하여 파손하지 않는 이상 그 안에 있는 공을 볼 수 없을 뿐만 아니라 꺼낼 수도 없게 만든 '닫힌 상자'를 통해 미지의 세계로서의 모집단을 모사하고 궁금증을 유발하고자 하였다. 한편, 이러한 장치는, 대통령 지지율이 참값은 존재하지만, 그것을 알기 어려운 것과 마찬가지로 분명히 존재하는 상자 안의 색공 개수의 비율의 참값을 쉽게 알 수 없도록 하는 대신 꼭짓점에 작은 구멍이 뚫려 있다는 점에서, 부득불 표본비율에 의해 모비율을 추정해야 하는 현실 문제 상황과 그 구조가 닮아 있다. 이와 같이 이 활동수업에서는 박지현(2016)이 활용한 상자에 밀폐성을 보완하여 미지의 모집단과 현실 문제 상황의 구조를 되도록 그대로 경험할 수 있는 방식의 교수학적 변환을 시도하였다.

(2) 크기가 큰 모집단에서의 복원추출에 의한 표집과 확률적으로 동형인 표집 활동

표본비율에 근거하여 모비율을 추정하는 현실 문제 상황을 교수학적으로 변환할 때 실제 활용되는 추정 과정에서처럼 비복원추출에 의한 표집을 반영할 수 있다. 예를 들어, 박민선, 고은성(2014)은 초등학교 4학년 학생들을 대상으로 내부가 보이지 않는 상자에 검정색과 흰색을 합쳐 총 300개의 구슬을 넣고 각 색깔별 구슬의 개수를 추측하도록 하였다. 그리고 이를 위해 무작위로 75개의 구슬을 비복원추출에 의해 꺼낸 것으로 보인다. 비복원추출에 의한 표집 활동은 비교적 이른 시기의 학생들을 대상으로 비형식적 통계적 추리 능력을 함양시키는데 있어 자연스러운 방법이라고 할 수 있다. 비례 개념을 가지고 있는 수준의 학생이 표본을 모집단의 준비례적 축소판으로 인식한다면 표본의 각 색깔별 구슬 개수에 적절한 수를 곱하여 모집단의 구슬 개수들을 추측할 수 있을 것이다. 박민선, 고은성(2014)의 학생들도 그러한 방법으로 모집단의 구슬 개수들을 추측할 수 있었다.

그러나 이론적 배경에서 살펴본 것과 같이 모비율을 확률적이고 계량적으로 구간추정하는 표준적인 통계학의 원리는 복원추출에 의한 표본을 활용하기 때문에 이러한 내용을 학습하는 단계의 학생들에게는 그러한 표본의 도입도 필요하다. 위 활동수업은 이러한 단계의 학생들에게 복원추출에 의한 표집 활동의 기회를 제공함으로써 그러한 모비율 추정의 이해를 지원하고자 하였다. 동시에 모집단의 크기가 클 경우, 복원추출에 의한 표본도 실제 추정 상황에서 활용되는 비복원추출에 의한 표본과 마찬가지로 동일한 대상을 여러 번 포함하지 않을 가능성이 크다는 점을 언급함으로써 이론과 실제 사이의 틈을 조화시키고자 하였다.

한편, 이 활동수업에서 사용하는 모집단이 크기가 10이라는 점에서 실제 문제 상황에서의 모집단의 크기와 괴리가 있다는 지적이 있을 수 있다. 이 활동수업에서 상자 안에 들어 있는 전체 공의 개수를 적게 설정한 것은 아래에서 서술되듯이 표본비율의 대표성에 대한 성찰의 계기를 효율적으로 제공하기 위한 것이었다. 그러나 복원추출에 의한 표본비율로부터 모비율을 추측하는 이와 같은 교수학습 활동에서 상자 안에 들어 있는 전체 공의 개수 자체가 반드시 실제 문제 상황에서의 모집단의 크기를 나타내는 것은 아니다. 한 세트의 표본을 복원추출하기 위한 각각의 시행에서 색공이 추출될 확률은 상자 안에 들어 있는 공의 개수가 아니라 색공 개수의

비율에 의해 결정되기 때문에 전체 공의 개수가 적은 상자를 사용하더라도 모집단의 크기가 큰 실제 문제 상황과 확률적으로 동형인 표집을 구현할 수 있다. 예를 들어, 박민선, 고은성(2014)은 총 300개의 구슬 중에 검정색을 200개, 흰색을 100개 사용하였는데, 복원추출에 의한 표집 활동에서는 검정색 2개와 흰색 1개만을 넣은 상자가 그 경우와 확률적으로 동형인 상황을 제공한다. 요컨대 본 연구의 활동수업은 모집단의 크기가 반드시 10이 아니라 10의 배수인 경우들을 포함하여 상정한 것이다. 이러한 유연성은 이 활동수업이 모비율을 추측하는 것이기 때문에 자연스럽게 얻어지는 것으로서 비슷한 방식으로 모평균 추측 활동을 할 때는 그러한 유연성이 발휘되기 어렵다.

(3) 효율적인 물리적 시뮬레이션 활동

교수학습 상황에서 표본비율에 근거하여 추정한 모비율의 신뢰구간을 모비율의 참값과 비교해 볼 수 있다면 그 추정이 얼마나 타당한지 직접 확인할 수 있다는 점에서 도움이 될 것이다. 하지만 대통령 지지율과 같은 현실 문제 상황은 모비율의 참값을 알기 어렵기 때문에 그것을 그대로 교수학습 콘텐츠로 활용하기 어려운 측면이 있다. 이러한 난점과 관련하여 윤수경(2018)은 후에 모비율의 참값이 알려지는 실제 전교 학생회장 선거에서 출구조사를 통해 모비율을 추정하는 수업을 설계 및 적용하였다. 그러나 어느 특정한 시점에 한 번 실시되는 학생회장 선거의 시기적 제약 때문에 이러한 방법을 널리 적용하기는 어려울 것으로 보았다. 이러한 어려움과 관련하여 현실 문제 상황을 모사하면서도 모비율의 참값을 공개되지 않도록 미리 설정해 두었다가 후에 참고할 수 있는 표집 시뮬레이션 활동이 주목된다.

Wild et al.(2011)은 비형식적 통계적 추리를 지원하는 컴퓨터 시뮬레이션 활동을 개발 및 적용하면서 학생들이 그러한 시뮬레이션의 의미를 이해하지 못하는 경우가 있다는 것을 인지하고 손으로 직접 해 보는 물리적 시뮬레이션 활동을 먼저 도입하였다(Arnold et al., 2011). 이처럼 물리적 시뮬레이션은 낯선 소프트웨어를 통해 가상 세계에 친숙해져야 하는 컴퓨터 시뮬레이션의 단점을 보완할 수 있을 것으로 생각된다. 한편, 확률적 과정을 포함하는 컴퓨터 시뮬레이션은 인식론적 측면에서도 난점이 지적되어 왔다. 컴퓨터가 구현하는 무작위는 실제 무작위와는 구별되는 그럴듯한 유사 무작위로서 학생들이 그것의 실제성을 믿지 않을 수 있다는 것이다(Pratt, 2005). 실세계의 무작위를 가상적으로 구현하기 위해서는 프로그래머가 컴퓨터로 하여금 실세계의 해당 무작위 요소를 반영하도록 프로그래밍 해야 하기 때문에 그 무작위가 실세계의 무작위 자체는 아니라는 주장은 타당해 보인다. 한편, Pratt(2005)는 실세계의 무작위와 마찬가지로 컴퓨터가 생성하는 패턴을 예상하거나 조절할 수 없다는 것을 학생들이 경험하면서 그 무작위성을 받아들이게 된다고 하였다. 그러나 컴퓨터 시뮬레이션이 실세계만큼이나 무작위적인 것으로 보인다고 해서 실제로 실세계를 모사한 것은 아닐 수 있다. 예를 들어, 가상의 동전을 여러 번 던졌을 때 앞면이 뒷면보다 아주 조금 더 많이 나오도록 설정된 컴퓨터 시뮬레이션을 경험한다면 실세계의 공평한 동전을 던질 때의 무작위를 모사한 것으로 보이겠지만 실제로 그런 것은 아니다. 때문에 타당한 교수학습 활동을 위해서는 실세계의 무작위를 모사하는 컴퓨터 시뮬레이션에 대해서도 무작위적인 것으로 보인다는 것을 인지하는 것과는 별개로 실세계의 무작위를 실제로 모사한다는 것에 대한 믿음의 행위가 요청된다. 이러한 행위는 동전을 실제로 던지는 것과 같이 심리적으로 무작위적인 것으로 인정될 수 있는 물리적 시뮬레이션 활동에서는 필요하지 않다.

이러한 점에서 모사하고자 하는 문제 상황과 닮아 있는 물리적 시뮬레이션을 먼저 활용하는 것이, 보다 직접적이고 지적으로 정직한 교수학적 변환이라고 생각된다. 본 연구에서의 활동수업은 복원추출에 의한 표본비율로부터 모비율을 추정하는 문제 상황의 구조를 그대로 보존하는 물리적 표집 시뮬레이션을 통해 그러한 교수학적 변환을 시도하였다. 다만 이 활동수업에 이어질 수 있는 확률적이고 계량적인 모비율 구간추정 교수학습과 관련하여 중심극한정리를 시뮬레이션으로 확인하고자 할 때는 되도록 많은 세트의 표본을 적은 시간에 관찰할 수 있는 컴퓨터 시뮬레이션도 적극적으로 활용할 필요가 있을 것이다.

한편 이 활동수업에서의 시뮬레이션은 ‘닫힌 상자’ 안에 넣는 공의 개수가 비교적 적기 때문에 상자를 제작하거나 추후 모비율의 참값을 확인할 때 효율적이다. 이것

은 300개의 구슬을 상자에 넣어 준비해야 하고 표본 추출 후에는 모비율의 참값의 확인을 위해 적어도 100개의 구슬을 세어 보아야 하는 박민선, 고은성(2014)에서의 활동과 비교된다.

(4) 표본비율의 대표성에 대한 성찰의 계기 제공

이론적 배경에서 살펴 본 것처럼 비슷한 상황에서 여러 번의 시행을 통해 얻은 상대도수, 즉 표본비율이 모비율과 유사할 것이라는 추측은 경험에 기반하여 가망성을 파악한다는 점에서 자연스러운 사고라고 할 수 있다. 때문에 이 활동수업에서 학생들은 표본비율의 대표성을 어렵지 않게 자각할 것으로 기대된다. 그러나 모비율을 하나의 값이 아닌 구간으로 추정해야 할 필요성은 표본비율의 대표성에 대한 자각이 그것이 정확히 모비율과 일치하는 것은 아니라는 것까지를 포함할 때 인지될 수 있다고 생각된다. 본 연구의 활동수업에서는 ‘닫힌 상자’ 안에 10개의 공을 넣음으로써 모비율이 0.1, 0.2, ⋯ , 0.9와 같이 소수점 아리 첫 번째 자리의 소수가 되도록 설정하였다. 표본비율의 값은 대부분 정확히 소수점 아래 첫 번째 자리까지만으로 나오지는 않을 것이므로 학생들은 자신들의 경험에 의한 표본비율이 모비율과 정확히 같을 수는 없다는 것을 인식하고 모비율의 참값이 무엇일지 고민하게 될 것이다. 이 활동수업은 이러한 설정을 통해 그러한 성찰의 계기를 제공하고자 하였다.

(5) 표집 변이성의 부각

표집 변이성은 하나의 모집단에서 표집한 같은 크기의 여러 표본 집합들이 다양하다는 것을 함축한다(탁병주 외, 2014). 박지현(2016)의 체험활동을 보완하여 일차적으로 개발했던 활동수업에서는 상자 안의 색공 개수의 비율을 각 모둠별로 결정하여 ‘닫힌 상자’를 제작한 후 두 모둠씩 짝이 되어 서로가 제작한 ‘닫힌 상자’ 안의 색공 개수의 비율을 알아맞히는 활동을 하였다. 이러한 활동에서는 하나의 모집단에서 표집한 여러 표본 집합들을 비교할 수 있는 기회를 제공하기 어려웠다. 이를 개선하기 위해 본 연구의 활동수업에서는 모둠들을 홍팀과 백팀의 두 팀으로 나누고 팀별로 ‘닫힌 상자’ 안에 넣을 색공의 개수, 즉 색공 개수의 비율을 교사가 결정하여 상대팀 모르게 알려주었다. 이로써 같은 팀에 속한 다섯 모둠들은 서로 다른 상자들을 동시에 활용하면서도 확률적으로 동일한 모집단을 모사한 ‘닫힌 상자’로부터 표본을 추출하게 된다. 때문에 같은 팀에 속한 다섯 모둠이 추출한 표본들은 사실상 하나의 모집단으로부터 표집한 다섯 세트의 표본들로 볼 수 있다. 이 활동수업에서는 발표를 통해 그 표본비율의 값들을 비교할 수 있는 기회를 제공함으로써 표집 변이성을 부각시키고자 하였다. 다만, 모둠 활동의 자율성 측면에서 추출해야 하는 표본의 크기를 지정하지는 않았기 때문에 같은 크기의 표본들에 대한 표본비율이 비교되는 것은 아니다. 그러나 이를 통해 이 활동수업에 이어질 수 있는 확률적인 방식의 모비율 구간추정 교수학습 장면에서 크기가 같은 표본들에 대해서도 그 비율들이 변이성을 갖는다는 직관이 보다 쉽게 인지될 수 있을 것으로 보았다.

2. 모비율 추측 활동수업의 적용 결과 및 해석

1) 모둠별 모비율 추측 결과

1학년과 2학년 수업에서 모둠별로 추출한 표본의 비율과 추측한 모비율의 값을 정리하면 [표 1]과 같다.

[표 1] 모둠별 표본비율과 추측한 모비율
[Table 1] Sample ratios and guess values of the population ratios by groups

학년/표본비율/모비율의 추측값 및 참값	홍팀					백팀				
	1조	2조	3조	4조	5조	6조	7조	8조	9조	10조
1학년의 표본비율	0.39	0.43	0.40	0.35/0.37	0.52/0.45/0.43	0.78	0.76	0.74	0.90	0.75/0.797
1학년이 추측한 모비율	0.4	0.4	0.4	0.4	0.4	0.8	0.8	0.8	0.8	0.8
2학년의 표본비율	0.38	0.39	0.28	0.44	0.00	0.75	0.70	0.72	0.72	0.69
2학년이 추측한 모비율	0.4	0.3	0.3	0.4	0.0	0.7	0.7	0.7	0.7	0.7
모비율의 참값	0.4					0.8				

총 20개 중 12개의 모둠에서 한 세트의 표본을 추출하여 얻은 표본비율의 값을 소수점 아래 두 번째 자리에서 반올림하여 모비율의 값을 추측하였다.[1] 이에 비해 1

1) 12개의 모둠에는 2학년 6조도 포함된다. 이 모둠은 주황색이 아닌 흰색 공 기준의 표본비율 0.25를 얻었다. 때문에 [표 1]과 같이 주황색 공을 기준으로 할 때, 표본비율의 값으로는

학년 4조, 5조, 8조, 10조는 처음 얻어진 표본비율의 값에 의심을 품고 새로운 시행을 추가하거나(4조), 다른 모둠의 시행 결과를 추가하거나(10조), 기 시행했던 결과 중에서 일부만을 사용하거나(5조), 기 시행했던 결과를 여러 구간으로 나누어 여러 세트의 표본들처럼 바라보거나(8조) 하여 구성한 새로운 표본을 통해 모비율을 추측하였다. 또, 2학년 2조는 연달은 시행에서 상자를 충분히 흔들지 않아 이전 시행의 결과가 그대로 관찰되는 경우가 많았을 것이라고 보고 반올림이 아닌 버림의 방법으로 추측하였다.[2] 1학년 9조와 2학년 10조는 자신들의 실험 결과 이외에도 다른 모둠의 실험 결과나 기타 교사의 행동 등에서 얻은 부수적인 정보를 함께 참고하여 추측하였다. 한편, 2학년 5조는 교환받은 '닫힌 상자'를 제작한 10조가 교사의 안내를 어기고 상자 안에 흰색 공만 10개를 넣었기 때문에 표본비율의 값으로 0.00을 얻어 모비율의 값을 0.0으로 추측하였다.

2) 표본비율의 대표성에 대한 자각 및 성찰

학생들의 발표 및 활동지 서술에 대한 질적 내용분석의 결과를 정리하면 [표 2]와 같다. 표본비율의 대표성에 대한 자각 양상은 7가지 범주로 요약할 수 있었다. 표본비율 및 모비율의 추측값과 그 이유를 발표하는 수업 장면에서, 총 20개 중 19개 모둠이 표본비율의 값과 유사한 값을 모비율로 삼았다고 발표하였지만, 교사가 재차 묻기 전까지는 그 이상의 이유를 언급하지 않았다. 이러한 발표 사례들은 '표본비율의 대표성에 대한 단순 자각(ㅁ)'으로 코딩되었다. 교사의 질문을 통해, 보다 반성적으로 생각한 이후에는 그중 9개의 모둠들은 표본비율이 '여러 번의 경험에 의한 것이므로(ㄱ)' 모비율에 가까울 것이라는 취지의 답변을 하였다. 이때 표본비율이 '항상 모비율에 가까운 것은 아니라는 점에 대한 인식(ㄴ)'이 드러나는 경우도 있었지만, '표본의 크기가 클수록 모비율에 가깝다(ㄷ)'고 말하기도 하였다.

ㅁ, ㄱ, ㄴ, ㄷ으로 코딩된 발표 사례들은 표본비율의 대표성에 대해 이론적 배경에서 살펴 본 내용들을 잘 반영한다. 표본비율을 구한 학생들은 자연스럽게 그것과 가까운 값을 모비율로 추측할 수 있었는데(ㅁ), 그것은 여러 번의 경험에 대한 신뢰(ㄱ)에서 비롯된 것이었다. 하지만 그러한 신뢰가 표본의 크기가 클수록 더 모비율에 가깝다(ㄷ)는 오류로 이어지기도 했던 것이다. 이러한 오류는 이론적 배경에서 살펴 본 대로 확률적이고 계량적인 모비율 구간추정 교수학습과 관련하여 중심극한정리를 도입하는 발판이 될 수 있을 것이다.

[표 2] 발표 및 활동지 서술의 질적 내용분석 결과
[Table 2] The result of QCA of students' presentations and activity appreciation

주 범주	하위 범주
표본비율의 대표성에 대한 자각 양상	· 표본비율은 여러 번의 경험에 의한 것이므로 모비율에 가까움(ㄱ) · 표본비율은 표본의 크기가 클 때 항상 그렇지는 않지만 모비율에 가까움(ㄴ) · 표본비율은 표본의 크기가 클수록 모비율에 가까움(ㄷ) · 표본은 모집단의 부분으로서 부분을 통해 전체 모집단의 비율 추측 가능(ㄹ) · 표본비율의 대표성에 대한 단순 자각(ㅁ) · 표본비율의 값이 어떤 모비율 값을 가리키는지에 대한 갈등(ㅂ) · 표본비율의 대표성에 대한 의심(ㅅ)
표집 변이성의 인지 양상	· [무작위] 표집에 의한 변이성 및 이에 따른 모비율 추측의 난점 인지(a) · [무작위] 표집에 의한 변이성의 단순 인지(b) · 표집 과정의 편의(bias) 가능성 인지(c)

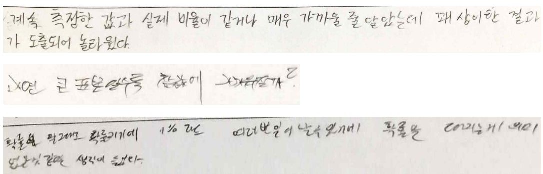
'표본비율의 값이 어떤 모비율 값을 가리키는지에 대한 갈등(ㅂ)'은 얻어진 표본비율의 값이 소수점 아래 첫 번째 자리로 떨어져야 하는 모비율의 참값과 정확히 같을 수 없을 때 발생하는 갈등을 의미한다. 이러한 갈등은 얻어진 표본비율의 소수점 아래 두 번째 자리의 숫자가 4나 5와 같을 때 주로 관찰되었다. 실제로 1학년 4조, 5조, 8조, 10조 등 4개 모둠은 각 모둠이 표본비율의 값으로 얻었던 0.35, 0.45, 0.74, 0.75([표 1])가 모비율을 추측하기에 모호하였기 때문에 단순히 반올림하여 추측하지는 않았다는 취지로 발표하였다.[3] 그러나 2학년 2조와 10조는 얻어진 표본비율의 값이 0.39와 0.69와 같이 소수

0.75(=1-0.25)를 얻은 셈이지만 모비율의 값을 0.8이 아닌, 0.25를 반올림한 0.3을 1에서 뺀 0.7이라고 추측하였다.

2) 실제로는 버림은 2학년 2조의 의도를 반영하지 못한다. 이전의 시행의 결과가 주황색 공이었을 때만 연달은 시행의 결과에 영향을 미친다면 버림의 방법을 사용할 수 있을 것이다. 그러나 흰색 공이었을 때도 그럴 수 있다는 점이 간과되었다.

점 아래 두 번째 자리의 숫자가 4, 5, 6 등이 아니었음에도 어떤 값을 모비율로 취할지 고민하였다고 하였다. 이러한 발표 사례들은 이 활동수업이 ‘1.3) 수업의 주안점’에서 기대했던 표본비율의 대표성에 대한 성찰의 계기를 제공할 수 있음을 시사한다.

표본비율의 대표성 자각에 대한 이러한 양상은 활동지 서술을 통해서도 유사하게 드러났다. 더불어 활동지 서술에서는 ‘표본비율의 대표성에 대한 의심(ㅅ)’도 3건의 사례에서 관찰할 수 있었다([그림 5]).



[그림 5] 표본비율의 대표성에 대한 의심
[Fig. 5] Doubts about the representativeness of sample ratios

이 중 첫 번째 사례는 얻어진 표본비율의 값이 0.90이었지만 모비율을 0.8로 추측한 1학년 9조([표 1]) 학생의 반응이다. 이 모둠은 처음에는 모비율을 표본비율과 같은 0.9로 추측하였지만, 발표 시간이 되기 전 모비율 추측 활동 단계에서 다른 모둠의 추측값이 모두 0.8인 것을 보고 추측값을 0.8로 수정하여 발표하였었다. 즉, 표본으로부터 원래 얻어진 표본비율과 추측한 모비율의 값이 모두 0.9였던 것으로서, 모비율의 참값과 “꽤 상이한([그림 5])” 표본비율을 얻은 것에 대한 놀라움을 관찰할 수 있다. 이 놀라움은 표본비율의 값이 모비율과 무시할 수 없는 차이가 있을 수 있다는 생각으로 이어졌을 것으로 보인다. 두 번째 사례에서의 의심은 2학년 4조 학생B의 반응으로서 발표 장면에서 그 연원을 찾을 수 있다([Episode(2학년 4조)]). 학생B는 ‘큰 수의 법칙’의 권위에 의지해 가장 많은 횟수의 시행(90회)을 하였기 때문에 가장 정확한 값을 얻었을 것이라고 하였지만, 그것이 자연스럽게 느껴지는지를 재차 묻는 교사의 질문에는 작은 소리로밖에 대답하지 못하였다. 사실 이 모둠이 90회 시행하여 얻은 표본비율 0.44는 참값과의 오차가 0.04로서 먼저 발표한 50회의 시행에서 0.38의 표본비율을 얻은 2학년 1조에 비해 더 정확한 값을 얻었다고 할 수 없었다. 학생B는 그러한 상황에서 교사의 반복되는 질문을 통해 표본의 크기가 클수록 모비율과의 오차가 줄어든다는 생각에 대해 성찰하면서 표본비율의 대표성에 의문을 품게 되었을 것으로 보인다. 한편, 세 번째 사례의 1학년 2조 학생은 얻어진 표본비율의 값 0.43이 0.5보다는 0.4에 더 가깝기 때문에 “가까운 쪽이 그나마 좀 더 확률이 있(자료1)”다고 발표하였었다. 이것은 표본비율에 따라 추측한 0.4가 틀릴 확률이 “1%라도 여러 번 일어날 수 있기에 확률을 따지는 게 의미 없는 것 같다([그림 5])”는 활동지 서술을 설명해 준다. 표본비율의 대표성을 확률 개념을 동원하여 의심하고 있지만 틀리거나 맞힐 확률의 정량적 차이의 중요성까지 인식하고 있지는 못한 것으로 보인다.

[Episode(2학년 4조)]
학생A: . . . 90번을 던졌는데 . . . 0.44의, . . . 4개라고 . . ..
. . .
교사: . . .. 가장 많이 시행을 한 거 같아요. 90번.
학생B: 네. 가장 정확하다고 할 수 있습니다.
학생들: (웃음)
교사: . . . 시행하면 할수록 더 오차가 줄어 들 . . . 거예요?
학생A, B: 네.
교사: . . . 자연스럽게 느껴져요?
학생A: 네.
학생B: 큰 수의 법칙에 의해서.
. . .
교사: 거부감이 없나요? 괜찮아요?
학생 A, B: (작은 소리로) 네.
(2018.8.13.)

처음 얻어진 표본비율의 값에 의심을 품고 곧바로 반올림을 적용하는 것 이외의 방법을 모색했던 모둠 중 1학년 4조, 5조, 8조, 10조 등은 전술했듯이 표본비율의 값이 소수점 아래 두 번째 자리의 숫자가 4나 5로서 모비율의 참값에서 상당히 벗어났을 것이기 때문에 ‘표본

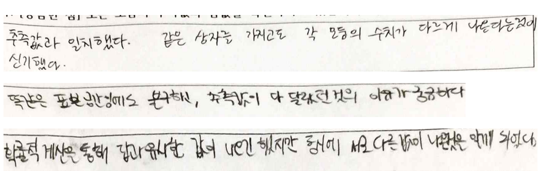
3) 1학년 5조는 원래 총 50회의 시행에서 0.52의 표본비율을 얻었다. 그런데 31회부터 40회까지의 시행에서 색공이 8회나 나온 것을 이상히 여겨 그 구간을 제외함으로써 두 번째 표본비율 0.45를 얻었다. 그러나 0.45는 소수점 아래 두 번째 자리의 숫자가 5로서 모호했기 때문에 또 다시 문제가 됐다. 결국에는 1회부터 30회까지의 시행만으로 계산한 세 번째 표본비율 0.43을 바탕으로 모비율을 추측하였다([표 1], 자료 1).

비율의 값이 어떤 모비율 값을 가리키는 지에 대한 갈등(ㅂ)'을 경험했다. 이때 이 모둠들이 그 갈등을 해결하기 위해 기타의 방법들을 동원하여 표본비율을 수정했던 행위에는 그 이면에 원래의 표본비율의 대표성에 대한 의심이 잠재되어 있었을 것으로 추측된다. [그림 5]의 처음 두 가지 사례들은 이와 같이 표본비율의 값이 모비율의 참값에서 상당히 벗어나는 경우 다른 표본들과의 비교나 반성적 사고를 유발하는 교사의 질문 등을 통해 표본비율의 대표성에 대한 의심이 표면화될 수 있음을 보여준다. 그리고 [그림 5]의 표면화된 의심들은 모비율을 그것이 존재할 만한 구간으로 추정하게 되는 동기나 그러한 구간을 확률적으로 추정하게 되는 동기를 제공함으로써 확률적이고 계량적인 모비율 구간추정 교수학습을 지원할 수 있을 것으로 보인다.

한편, '표본은 모집단의 부분이고 부분을 통해 전체 모집단의 비율을 추측하는 것이 가능할 것(ㄹ)'이라는 발표 사례 한 건과 그 이유를 고민해 보았다는 활동지 서술 사례 한 건이 관찰되었다. 그러나 이 활동수업에서의 표본은 복원추출에 의한 것이기 때문에 ㄹ과 같은 반응에서의 '부분'은 비복원추출에 의한 '모집단의 준비례적 축소판'으로서의 표본에서의 '축소판'과는 다른 개념이어야 한다. 복원추출에 의한 표본의 경우에도 부분과 전체의 관계로 표본비율의 대표성을 자각하는 인지심리에 대해서는 추가적인 연구가 필요한 것으로 보인다.

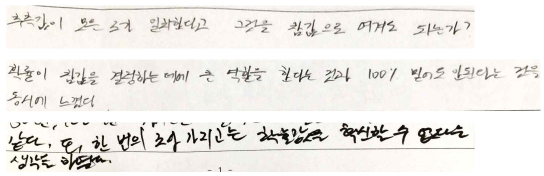
3) 표집 변이성의 인지

표집 변이성의 인지 양상은 3가지 범주로 요약할 수 있었다([표 2]). 이 활동수업에서는 발표와 그 내용의 공시를 통해 같은 팀에 속한 다섯 모둠들이 사실상 하나의 모집단에서 표집한 표본의 비율들을 비교할 수 있는 기회를 제공하였다. 이러한 활동 후에 작성한 활동지 서술 중 [그림 6]과 같은 3건의 사례에서 '표집에 의한 변이성의 단순 인지(b)'를 관찰할 수 있었다.



[그림 6] 표집 변이성의 인지
[Fig. 6] Perception of the sampling variability

이러한 인지는 표본비율로부터 모비율을 추측하는 상황에서는 추출되는 표본에 따라 그러한 추측 결과가 다양할 수 있다는 난점에 대한 인지로 확장될 수 있을 것이다. 실제로 [그림 7]과 같이 3건의 활동지 서술 사례에서 그러한 확장된 인지(a)를 관찰할 수 있었다.



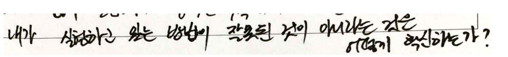
[그림 7] 표집 변이성에 따른 모비율 추측의 난점 인지
[Fig. 7] Perception of the Difficulty in estimating the population ratios due to the sampling variability

첫 번째 사례와 두 번째 사례는 1학년의 어느 한 학생이 각각 활동지의 4.가.(궁금한 점)와 4.나.(느낀 점) 항목에 대해 작성한 반응이다. 실제로 1학년은 각 팀별로 모비율의 "추측값이 모든 조가 일치([그림 7], [표 1])"했음에도 이 학생은 그렇다고 해서 그 값이 반드시 참값은 아닐 것이라는 의문을 제기하였다. 모비율의 추측값은 모두 일치하였지만, 표본비율의 값은 모두 달랐고([표 1]) 더 많은 세트의 표본을 추출한다면 1학년 9조의 표본비율과 같이 어떤 표본비율에 따라서는 얼마든지 모비율을 다르게 추측할 수도 있다는 것을 인지한 것으로 보인다. 그래서 "[통계적] 확률([표본비율])이 [모비율의] 참값을 결정하는 데에 큰 역할을([그림 7])" 하는 것이 사실이지만 동시에 "100% 믿어서도 안 된다([그림 7])"는 것을 느꼈을 것이다. 또 다른 1학년 학생도 같은 맥락에서 "한 번의 조사 가지고는 [수학적] 확률([모비율]) 값을 확신할 수 없다([그림 7])"고 생각한 것으로 보인다.

표집 변이성으로 인한 모비율 추측에서의 난점 인지는 표본비율 및 모비율의 추측값과 그 이유를 발표하는 수업 장면을 통해서도 나타났다. 실제로 1학년 8조는 이 모둠을 이루는 두 학생이 각각 25회씩, 총 50회의 복원추출 시행을 실시하였는데, 두 학생의 결과가 매우 달라 모비율 추측의 난점을 인지하였다는 취지로 발표하였다. 또, 전술했듯이 1학년 9조는 자기 모둠의 실험에서 다른 모둠에 비해 색공이 지나치게 많이 나왔다는 것을 인지하고 부수적인 방법을 사용하여 추측값을 수정하였다.

활동지 서술과 발표에서 나타난 이상의 몇몇 a와 b 코딩 사례들은 모비율이 존재할 만한 구간을 확률적으로 추정하게 되는 동기 또는 그러한 확률을 구하는 방법을 설명하기 위한 보조 자료 등으로 활용될 수 있을 것이다. 이것은 이 활동수업이 ‘1.3) 수업의 주안점’에서 기대했던 표집 변이성의 부각을 통해 확률적인 방식의 모비율 구간추정 교수학습을 일정 부분 지원할 수 있음을 시사한다.

확률적이고 계량적인 모비율 구간추정은 무작위 표집에 의한 변이성을 묘사하는 중심극한정리에 근거하여 이루어진다. 한편, 표본비율의 변이성은 표본의 편의(bias)에 의한 변이를 포함할 수도 있는데 되도록 그것을 피하는 것이 무작위 표집을 위해 바람직하다. 이 활동수업을 통해서 학생들은 표집 과정의 편의 가능성을 인지(c)하기도 하였다. 이는 처음 얻어진 표본비율의 값에 의심을 품고 곧바로 반올림을 적용하는 것 이외의 방법을 모색했던 모둠 중 1학년 5조, 2학년 2조, 10조 등의 발표에서 관찰되었다. 1학년 5조는 표본의 추출 과정에서 갑자기 색공이 많이 나오는 시행 구간을 인지하여 그 부분을 제외하고 표본비율을 다시 계산하였고, 2학년 2조는 전술하였듯이 연달은 시행에서 상자를 충분히 흔들지 않아 발생했을지 모르는 편의를 고려하였으며, 2학년 10조도 표집 과정의 편의를 고려하여 모비율 추측을 위해 부수적인 정보를 적극적으로 수집하였다. 또, [그림 8]과 같이 1건의 활동지 서술 사례에서도 표집 과정의 편의 가능성에 대한 인지를 관찰할 수 있었다. 이러한 편의 가능성의 인지와 관련해서는 아래에서 추가로 논의한다.



[그림 8] 표집 과정의 편의 가능성 인지
[Fig. 8] Perception of the possibility of sampling bias

3. 모비율 추측 활동수업의 교수학습에 주는 시사점

2절에서 살펴 본 것처럼 복원추출에 의한 표본비율로부터 모비율을 추측하는 이 활동수업은 표본비율의 대표성에 대한 자각 및 성찰, 그리고 무작위 표집에 의한 변이성의 인지 등에 일정 부분 효과가 있었다. 이 활동수업에서 표본비율의 대표성은 표본비율로부터 모비율을 추측하는 과정에서, 무작위 표집에 의한 변이성은 사실상 하나의 모집단에서 표집한 표본의 비율들을 비교할 수 있는 기회를 제공함으로써 주로 부각시키고자 하였다. 수업의 실제 적용 결과, 표본비율의 대표성에 대한 단순 자각은 거의 모든 모둠(19개)에서 관찰되었고 교사의 질문을 통해 적지 않은 수의 모둠들(9개)이 그 대표성의 자연스러움이 경험으로부터 나온다는 점을 자각할 수 있었다. 특히 얻어진 표본비율의 값이 모비율의 참값에서 상당히 벗어나는 경우에도, 표본비율의 값이 어떤 모비율 값을 가리키는지에 대해 갈등하기도 하고, 다른 표본들과의 비교나 교사의 질문을 통해 표본비율의 대표성에 대해 의심을 갖기도 하면서 모비율을 확률적이고 계량적으로 구간추정하게 되는 동기가 제공될 수 있을 것으로 보였다.

한편, 무작위 표집에 의한 변이성의 인지와 관련해서는 하나의 모집단에서 추출한 표본의 비율로부터 모비율을 추측하는 과정만으로도 그것이 인지되는 장면을 1학년 8조와 9조에서 관찰할 수 있었다. 전술했듯이 8조는 모둠을 이루는 두 학생의 시행 결과가 매우 다르다는 것으로부터, 9조는 미리 살펴 본 다른 모둠의 실험 결과가 자신들의 것과 다르다는 것으로부터 모비율 추측 시의 표집 변이성의 난점을 깨달았다. 또, 9조는 다른 모둠의 학생들이 9조의 ‘닫힌 상자’를 가져가 실험한 결과 9조의 결과보다 색공이 훨씬 적게 나온 것도 부수적인 방법을 사용하여 모비율의 추측값을 수정하게 된 계기가 되었다고 하였다. 9조의 사례는 학생들끼리 친근한 교실에서는 모둠들 상호간의 자연스러운 교류를 통해 표집 변이성이 인지될 수도 있음을 보여준다.

그러나 이러한 상황들이 항상 펼쳐지는 것은 아니기 때문에 이 활동수업에서 표집 변이성을 부각시키기 위해 마련한 교수학습 장치가 중요한 의미를 갖는다. 살펴 본 것처럼 몇몇 활동지 서술 사례에서 그러한 장치가 효과적이라는 것을 확인할 수 있었다. 그러나 활동지 4번 항목에서 물었던 궁금한 점과 느낀 점에 대한 반응들은 무작위 표집에 의한 변이성의 인지에 대한 것보다는 표본비율의 대표성에 대한 자각 및 성찰에 대한 것이 주를 이루었다. 실제로 모든 학생의 설문지 서술들로 이루어진 분석 단위 총 64개 중에 3개의 중복코딩을 포함해 37개가 후자로 코딩되었는데 반해 전자로 코딩된 것은 6개에 그쳤다. 궁금한 점과 느낀 점에 대한 이러한 분석 결과가 직접적으로 다수 학생의 표집 변이성의 미인지를

의미하는 것은 아니지만, 적어도 표본비율의 대표성에 비해 표집 변이성이 더 적게 부각되었음을 시사한다고 생각된다.

그 원인 중 하나로 이 활동수업에서 1학년과 2학년이 추측한 모비율 값의 균일성을 생각해 볼 수 있다. [표 1]에서 보듯이 1학년의 홍팀은 모비율을 모두 0.4로, 백팀은 모두 0.8로 추측하였다. 또, 2학년의 홍팀은 0.3과 0.4로, 백팀은 모두 0.7로 추측하였다. 이에 비해 모둠들이 얻은 표본비율의 값들은 다양하여 [그림 7]의 첫 번째 사례에서처럼 표집 변이성이 인지되기도 하지만 그것으로부터 추측한 모비율의 값들이 균일하기 때문에 그만큼 표집 변이성이 덜 부각될 수 있다고 생각된다. 이에 대해 '닫힌 상자' 안에 넣는 총 공의 개수를 늘리는 것이 하나의 개선 방안이 될 수 있을 것이다. 예를 들어, '닫힌 상자'에 20개의 공이 들어 있다고 가정하면 얻어진 표본비율에 곧바로 반올림을 적용했던 모둠들의 모비율 추측값들은 더 다양해질 수 있다. 이것은 그 가상의 모비율 추측값들을 [표 1]의 추측값들과 함께 표시한 [표 3]에서 확인된다. 이로써 표본비율의 변이성이 모비율의 추측값들을 더 다양하게 하고 따라서 참값을 맞히는 모둠의 수가 줄어들 가능성이 높아진다. 이러한 변화가 표집 변이성을 조금 더 부각시킬 수 있을 것으로 기대된다. 다만, 상자 안에 넣을 공의 개수는 '닫힌 상자'를 제작하거나 후에 상자를 파손 및 개봉하여 모비율의 참값을 확인하고자 할 때 가중될 번거로움도 감안하여 적절히 조절하여야 할 것이다.

[표 3] 가상의 모비율 추측값

[Table 3] Guess values of the population ratios in a hypothetical situation

학년	표본비율/ 모비율/ 가상의 모비율	1조	2조	3조	4조	6조	7조	8조	9조
1	표본비율	0.39	0.43	0.40		0.78	0.76		
	추측한 모비율	0.4	0.4	0.4		0.8	0.8		
	가상의 모비율 추측값	0.40	0.45	0.40		0.80	0.75		
2	표본비율	0.38		0.28	0.44	0.75	0.70	0.72	0.72
	추측한 모비율	0.4		0.3	0.4	0.7	0.7	0.7	0.7
	가상의 모비율 추측값	0.40		0.30	0.45	0.75	0.70	0.70	0.70

한편, '2.3) 표집 변이성의 인지'에서 살펴 본 것처럼 특별한 교수학습 장치가 없어도 어떤 학생들은 표집 과정에서 비교적 자율적으로 그 편의 가능성에 대해 인지하고 그것을 조정하거나 조정하고자 하는 모습을 보였다. 교사는 수업의 도입 부분에서 활동을 안내하면서 모집단의 정보를 추측하기 위한 표집은 무작위로 이루어지는 것이 중요하고 '닫힌 상자'도 잘 흔들어 공의 색을 관찰해야 한다는 점을 주지시켰지만 더 이상 표집 과정의 편의 여부에 대해 주목하지 않았다. 그러나 이 활동수업이 지원하고자 하는 확률적인 방식의 모비율 구간추정에서 차지하는 무작위 표집의 중요성을 감안하면 표집 과정에서 발생 가능한 편의에 대해 주목시키는 교사의 보다 적극적인 관여가 필요하다고 생각된다. 예를 들어, 모비율의 추측값과 그 이유를 발표하는 수업 장면에서 어떤 모둠이 표집 과정의 편의를 조정하였음을 언급한다면 그 내용을 전체 학생들에게 주목시키고 이를 참조하면서 다른 모둠들에게도 편의의 인지나 조정 여부에 대해 질문할 수 있을 것이다.

이 활동수업은 표준적인 통계학의 원리에 따른 모비율 구간추정을 교수학습하기 전 단계의 학생들을 대상으로 적용하여 추후 이어질 수 있는 그러한 모비율 구간추정의 이해를 지원하고자 하였다. 한편 확률적이고 계량적인 모비율 구간추정을 학습하는 중이거나 이미 학습한 학생들을 대상으로 적용하여 이론적 원리를 실제로 구현해 보고 그것이 잘 작동하는지를 확인하는 방식으로 그 이해를 지원할 수도 있다고 생각된다. 이 경우 모비율을 하나의 값으로 추측하지 않고 특정 신뢰도 하에 그 신뢰구간을 추정하도록 활동 내용을 수정해야 할 것이다. 이 때 그 신뢰구간이 수치상으로 가능한 모비율의 값을 단 하나만 포함하기 위해서 필요한 표본의 크기를 구해보는 활동도 포함시킬 수 있을 것이다.

V. 결론 및 제언

본 연구에서는 표준적인 통계학의 원리에 따른 확률적이고 계량적인 모비율 구간추정의 이해를 지원하기 위한 활동 중심의 수업을 개발하고 실제로 적용하여 교수학습적 측면의 시사점을 논의하였다. 이 활동수업에서는 내부의 공을 볼 수도 꺼낼 수도 없는 완전히 밀봉된 '닫힌 상자'를 통해 미지의 모집단과 현실 문제 상황을 모

사하는 한편 복원추출에 의한 효율적인 물리적 시뮬레이션 활동을 구안하였다. 또, 이를 활용하여 표본비율의 대표성과 표집 변이성을 부각시키는 교수학습 장치를 마련하였다. 이 활동수업을 실제로 적용한 결과 표본비율의 대표성에 대한 자각 및 성찰, 그리고 무작위 표집에 의한 변이성의 인지 등에 일정 부분 효과가 있었다. 한편 표집 변이성의 보다 뚜렷한 부각을 위해 '닫힌 상자' 안에 넣는 총 공의 개수를 적절히 증가시키는 방안과 표집 과정에서 발생 가능한 편의 및 그 조정에 주목시키는 교사의 적극적 관여의 필요성 등이 논의되었다.

본 연구는 표준적인 통계학의 원리에 따른 모비율 구간추정을 교수학습하기 전 단계의 학생들을 대상으로 활동수업을 적용하여 그 이해의 지원 가능성을 확인하였지만 실제로 그러한 형식적인 모비율 추정의 교수학습이 이 활동수업 직후에 이어지지는 못하였다. 그러한 모비율 구간추정의 교수학습이 실제로 이어질 수 있는 상황에서 이 활동수업을 적용하는 것이 더 바람직할 것이다. 이를 위해 2015 개정 교육과정의《심화수학Ⅱ》교과서나 창의적 체험활동 프로그램에서 그러한 모비율 구간추정을 다룰 때 이 활동수업이 함께 활용되기를 기대한다.

한편, 이 활동수업은 크기가 큰 모집단에서의 복원추출에 의한 표집과 확률적으로 동형인 표집 활동을 물리적으로 시뮬레이션하면서도 적은 수의 공을 사용한다는 점에서 효율적이었다. 전술했듯이 이것은 모평균이 아닌 모비율 추정의 이해를 지원하기 위한 활동이었기 때문에 가능했다. 같은 방식으로 모평균 추정 이해의 지원을 위한 활동수업을 개발한다면 박민선, 고은성(2014)에서의 활동과 같이 현실 문제 상황에서의 모집단의 크기를 고려하여 상자 안에 이 활동수업에서보다 훨씬 많은 공을 넣어야 할 것이다. 또, 각 공에는 키나 몸무게와 같은 모사하고자 하는 모집단의 특성값들을 표시해야 할 것이다. 이와 같이 모평균 추정의 이해를 지원하는 물리적 시뮬레이션 활동은 매우 번거로운 일로서 효율적이지 않다. 한편 표준적인 통계학의 원리에 따른 모비율과 모평균의 추정은 모두 중심극한정리에 기반하여 이루어지지만, 모비율 추정의 경우 모집단의 확률분포가 Bernoulli 분포로서 모평균 추정에서의 모집단의 확률분포보다 간단한 편이다. 즉, 모비율 추정은 모평균 추정과 동일한 이론에 기반하면서도 그 구조와 이해가 더 단순하고 수월한 측면이 있다. 이제까지《확률과 통계》에서는 모평균 추정 이후 모비율 추정이 다루어져 왔던 바, 본 연구의 활동수업과 같이 그 이해를 지원할 수 있는 효율적인 활동이 가능하고 보다 간단하고 수월한 모비율 추정을 모평균 추정에 앞서 교수학습한 후에 동일한 이론적 기반을 매개로 모평균 추정을 다루는 학습경로를 고려해 볼 수 있다고 생각된다(Sedlmeier, 1999).

참 고 문 헌

경상남도교육청 (2018). 중등수학콘텐츠 교사지도서. 2018 중등수학콘텐츠 개발 사업 자료집.

Gyeongsangnamdo office of education (2018). *Teacher's guide to secondary mathematics content*, 2018 Secondary School Mathematics Contents Development Project Document.

고은성 (2012). 통계적 변이성 사고 요소 간의 관계 연구. 학교수학 14(4), 495-516.

Ko, E. S. (2012). The relationships among components of thinking related to statistical variability. *School Mathematics 14*(4), 495-516.

고은성, 이경화 (2010). 변이성과 변이 추론의 지도를 위한 지식. 수학교육학연구 20(4), 493-509.

Ko, E. S., & Lee, K. H. (2010). A study on knowledge for the teaching of variability and reasoning about variation, *Journal of Educational Research in Mathematics 20*(4), 493-509.

구나영, 탁병주, 강현영, 이경화 (2015). 표본 지도에 대한 고찰: 국외 교육과정 분석을 중심으로. 학교수학 17(3), 515-530.

Ku, N. Y., TAk, B., Kang, H. Y., & Lee, K. H. (2015). A study on the teaching sample: an analysis of foreign curriculum. *School Mathematics 17*(3), 515-530.

김우철, 김재주, 박병욱, 박성현, 송문섭, 이영조, 전종우, 조신섭 (1996). 일반통계학. 서울: 영지문화사.

Kim, Y., Kim, J., Park, B., Park, S., Song, M., Lee, Y., Jeon, J., & Jo, S. (1996). *General statistics*, Seoul: Youngji Publishers.

박민선 (2015). 비형식적 통계적 추리의 평가. 박사학위 논문, 서울대학교.

Park, M. (2015). *Assessment of informal statistical inference*. Doctoral dissertation, Seoul National University.

박민선, 고은성 (2014). 초등학교 4학년 학생들의 표집활동 분석. 학교수학 16(3), 503-518.

Park, M. & Ko. E. S. (2014). Fourth graders engaged in sampling: a case study. *School Mathematics 16*(3), 503-518.

박지현 (2016). 고등학교 통계교육 교수・학습자료 개발: 통통세-통계(統計)로 통(通)하는 세상 탐구. In 통계청, 통계 교수・학습 자료 개발 교사연구회 보고서 - 고등학교 (pp. 186-189).

Park, J. H. (2016). High school statistics teaching and learning material development: Tong Tong Seo - exploring the world through statistics. In Statistics Korea, *Teachers research group document for developing statistics teaching and learning materials - high school* - (pp. 186-189).

신보미, 이경화 (2006). 컴퓨터 시뮬레이션을 통한 통계적 확률 지도에 대한 연구. 수학교육학연구 16(2), 139-156.

Shin, B. M. & Lee K. H. (2006). A study on the statistical probability instruction through computer simulation. *Journal of Educational Research in Mathematics 16*(2), 139-156.

윤수경 (2018). 출구조사를 활용한 모비율의 추정 수업 설계 및 적용. 석사학위논문, 교원대학교.

Yoon, S. K. (2018). *Instruction design and its application of statistical estimation by using exit poll.* Master thesis, Korea National University of Education.

이경화, 지은정 (2005). 표본 개념의 교육적 의의와 인식 특성 연구. 수학교육학연구 15(2), 177-196.

Lee, K. H. & Ji, E. J. (2005). Pedagogical significance and students' informal knowledge of sample and sampling. *Journal of Educational Research in Mathematics 15*(2), 177-196.

이기돈 (2018a).《확률과 통계》의 시행과 두 가지 확률에 대한 고찰 및 교육적 시사점. 한국수학사학회지 31(4), 251-269.

Lee, G. D. (2018b). A Study on Experiments and Two Interpretations of Probability in《Probability and Statistics》and Its Educational Implications. *The Korean Journal for History of Mathematics 31*(5), 251-269.

이기돈 (2018b). 학생들의 질문하기에 의해 개념을 도입하는 수학수업의 제안 및 가능성 탐색. 교육과학연구 20(1), 123-153.

Lee, G. D. (2018a). A proposal of the mathematical lesson where concepts are introduced according to students' questioning, and an exploration of the possibilities of that method. *Journal of Education science 20*(1), 123-153

이영하, 이은호 (2010). 통계적 추론에서의 표집분포 개념 지도를 위한 시뮬레이션 소프트웨어 설계 및 구현. 학교수학 12(3), 273-299.

Lee, Y. H. & Lee, E. H. (2010). The design and implementation to teach sampling distributions with the statistical inferences. *School Mathematics 12*(3), 273-299.

이은희, 김원경 (2015). 국내외 통계교육 연구동향 비교분석. 수학교육 54(3), 241-259.

Lee, E. H. & Kim, W. K. (2015). A comparative analysis on research trends of statistics education between Korea and overseas. *The Mathematical Education 54*(3), 241-259.

이정연, 이경화 (2017). 중·고등학생들의 비형식적 통계적 추리의 수준 연구. 학교수학 19(3), 533-551.

Lee, J. Y. & Lee, K. H. (2017). Study on the levels of informal statistical inference of the middle and high school students. *School Mathematics 19*(3), 533-551.

이형숙, 이경화, 김지원 (2010). 초등수학영재들의 통계적 사고 특성 사례 분석. 수학교육학연구 20(3), 339-356.

Lee, H. S., Lee, K. H., & Kim, J. W. (2010). A Case study of the characteristics of mathematically gifted elementary students' statistical reasoning: focus on the recognition of variability. *Journal of Educational Research in Mathematics 20*(3), 339-356.

최성호, 정정훈, 정상원 (2016). 질적 내용분석의 개념과 절차, 질적탐구 2(1), 127-155.

Choi, S., Jung, J. H., & Jung, S. W. (2016). Concept and procedures of Qualitative Content Analysis. *Journal of Qualitative Inquiry 2*(1), 127-155.

최인용, 조한혁 (2017). 큰 수의 법칙 시뮬레이션에서 중학생의 안구 운동 분석. 수학교육 56(3), 281-300.

Choi, I. Y. & Cho, H. H. (2017). An analysis of middle school student's eye movements in the law of large numbers simulation activity. *The Mathematical Education 56*(3), 281-300.

탁병주, 구나영, 강현영, 이경화 (2014). 표본 개념에 대한 고찰: 역사적 분석을 중심으로. 학교수학 16(4), 727-743.

Tak, B., Ku, N. Y., Kang, H. Y., & Lee, K. H. (2014). A study on the concept of sample by a historical analysis.

School Mathematics 16(4), 727-743.

탁병주, 구나영, 강현영, 이경화 (2017). 중등수학 예비교사들의 통계적 소양: 표본 개념에 대한 이해를 중심으로. 수학교육 56(1), 19-39.

Tak, B., Ku, N. Y., Kang, H. Y., & Lee, K. H. (2017). Preservice Secondary Mathematics Teachers' Statistical Literacy in Understanding of Sample. *The Mathematical Education 56*(1), 19-39.

한채린, 이경원, 김도연, 배미선, 권오남 (2018). 과제의 구조화 정도에 따른 초등학생들의 통계적 변이성 이해 양상에 대한 사례 연구. 초등수학교육 21(2), 131-150.

Han, C., Lee, K., Kim, D., Bae, M. S., & Kwon, O. N. (2018). Aspects of understandings on statistical variability across varying degrees of task structuring. *Education of Primary School Mathematic 21*(2), 131-150.

홍서영, 서태열 (2014). 지리교육 연구의 대안적 연구 방법으로서 QCA(Qualitative Content Analysis)의 적용. 한국지리환경교육학회 22(3), 103-120.

Hong, S. & Seo, T. Y. (2014). Alternative methods in geography education research -application of QCA (Qualitative Content Analysis)-. *The Journal of the Korean Association of Geographic and Environmental Education 22*(3), 103-120.

황선욱, 강병개, 김영록, 윤갑진, 김수영, 송미현 외 (2014). 확률과 통계, 서울: 좋은책 신사고.

Hwang, S. W., Kang, B. G., Kim, Y. L., Yoon, G. J., Kim, S. Y., Song, M. H. et al. (2014). *Probability and statistics*. Seoul: Joheunchaeg sinsago.

Arnold, P., Pfannkuch, M., Wild, C. J., Regan, M., & Budgett, S. (2011). Enhancing students' inferential reasoning: from hands-on to "movies". *Journal of Statistics Education 19*(2), 1-32.

Dale, A. I. (2012). *A history of inverse probability: From Thomas Bayes to Karl Pearson*. Springer Science & Business Media.

DasGupta, A. (2010). *Fundamentals of probability: a first course*. Springer Science & Business Media.

Elo, S. & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing 62*(1), 107-115.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Beverly Hills, CA: Sage.

Pfannkuch, M. & Wild, C. (2004). Towards an understanding of statistical thinking. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 17-46). Dordrecht, The Netherlands: Kluwer Academic.

Pratt, D. (2005). How do teachers foster students' understanding of probability? In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 171-189). Springer Science & Business Media.

Saldanha, L. & Thompson, P. (2002). Conceptions of sample and their relationship to statistical inference. *Educational studies in mathematics 51*(3), 257-270.

Schreier, M. (2012). *Qualitative content analysis in practice*. London: Sage.

Sedlmeier, P. (1999). *Improving statistical reasoning: Theoretical models and practical implications*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Wild, C. J. & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International statistical review 67*(3), 223-248.

Wild, C. J., Pfannkuch, M., Regan, M., & Horton, N. J. (2011). Towards more accessible conceptions of statistical inference, *Journal of the Royal Statistical Society: Series A (Statistics in Society) 174*(2), 247-295.

Xue, J. (2006). A polya urn model of conformity. *Cambridge working papers in economics 0614*.

An application and development of an activity lesson guessing a population ratio by sampling with replacement in 'Closed box'

Lee, Gi Don
Kyeongin High School, Seoul
E-mail : tracer0@sen.go.kr

In this study, I developed an activity oriented lesson to support the understanding of probabilistic and quantitative estimating population ratios according to the standard statistical principles and discussed its implications in didactical respects. The developed activity lesson, as an efficient physical simulation activity by sampling with replacement, simulates unknown populations and real problem situations through completely closed 'Closed Box' in which we can not see nor take out the inside balls, and provides teaching and learning devices which highlight the representativeness of sample ratios and the sampling variability. I applied this activity lesson to the gifted students who did not learn estimating population ratios and collected the research data such as the activity sheets and recording and transcribing data of students' presenting, and analyzed them by Qualitative Content Analysis. As a result of an application, this activity lesson was effective in recognizing and reflecting on the representativeness of sample ratios and recognizing the random sampling variability. On the other hand, in order to show the sampling variability clearer, I discussed appropriately increasing the total number of the inside balls put in 'Closed Box' and the active involvement of the teachers to make students pay attention to controlling possible selection bias in sampling processes.

* ZDM Classification : D74, U64
* 2000 Mathematics Subject Classification : 97D40, 97U60
* Key words : estimating population ratios, activity lesson, 'Closed Box', sampling with replacement, physical simulation

<부록> 활동지

'닫힌 상자' 속 색공 개수의 비율 추측

팀: (　　　　) 모둠: (　　　　)조　이름: ________________

1. 자료(복원추출에 의한 표본) 수집

관찰횟수	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
관찰 값																

관찰횟수	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
관찰 값																

관찰횟수	33	34	35	36	37	38	39	40	41	42	43	44	45	46
관찰 값														

. . .

※ 관찰 값: (주황)색 공이 관찰되면 '1', 그렇지 않으면 '0'

※※ 꼭 46회를 관찰해야 하는 것은 아닙니다. 원하는 횟수만큼 관찰해도 좋아요. 칸이 부족한 경우 뒷면을 이용해 추가로 기록하셔요.

2. (1학년) [표본 자료 분석] 위에서 얻은 표본 자료에 의하면 상자에서 무작위로 선택된 한 개의 공이 색공일 확률은?

무작위로 선택된 한 개의 공이 색공일 확률 : ________

(2학년) [표본 자료 분석] 색공 개수의 비율을 추측하기 위해 위에서 얻은 표본 자료를 요약하면?

3. 추측하기

가. '닫힌 상자' 속 색공 개수의 비율은?

색공 개수의 비율 : $\dfrac{\qquad}{10}$　　(색공의 개수: (　　　) 개))

나. 그렇게 추측한 이유는?

4. 활동 되돌아보기

가. [궁금한 점] 모든 모둠의 추측값과 참값을 확인하여 보았습니다. 궁금한 질문을 적어 보아요.

나. [느낀 점] 이 활동을 하면서 어떤 생각들을 하였나요?