

국가수준 학업성취도 평가의 초등학교 표집방법 비교¹⁾

민 경 석(세종대학교 부교수)*

박 인 용(한국교육과정평가원 부연구위원)**

김 경 희(한국교육과정평가원 선임연구위원)

《 요 약 》

이 논문에서는 국가수준 학업성취도 표집평가의 중요성과 기능 확대가 지속적으로 논의됨에 따라 기존 2단계 층화군집표집방법과 표집학교의 모든 학생을 표본으로 선정하는 '학교군집표집'방법, 기존 표집방법에 시·도별 최소 학교 수 조건을 추가한 '최소학교수표집'방법을 비교하였다. 구체적으로, 표본 대표성과 국가 및 시·도교육청 수준에서의 표본 통계치 정확성 측면에서 세 표집방법을 통계적으로 비교하였다. 연구결과로서, 표집 가중치를 적용한 경우, 세 표집방법 모두 모집단을 적절히 대표하는 표본분포를 산출하고 있었으며, 국가수준에서의 성취수준 비율과 척도점수 평균 추정치가 모집단의 모수치와 유사하였으며, 표집오차가 작아 표본 통계치의 정확성이 높았다. 반면 시·도교육청 수준의 성취수준 비율과 척도점수 평균 추정치는 대규모 시·도교육청에서는 세 표집방법 모두에서 정확성이 높게 나타난 반면, 소규모 시·도교육청에서는 '최소학교수표집'을 통한 표본 통계치의 정확성이 두 표집방법에 비해 상대적으로 높게 나타났다.

주제어: 국가수준 학업성취도 평가, 표집방법, 표집오차

1) 본 연구는 한국교육과정평가원(2014)에서 수행한 '국가수준의 기초학력 점검을 위한 초등학교 학업성취도 평가 방안' III장에서 사용한 원자료와 내용을 일부 포함하였으며 분석내용, 결과해석, 결론 등을 수정·보완한 것임.

* 제1저자, minkyungseok@sejong.ac.kr

** 교신저자, iypark@kice.re.kr

I. 서론

국가수준 학업성취도 평가의 전수시행은 2008년 이전 표집시행에서의 교육과정 질 관리에 더하여 개별 학생과 학교의 성취도 평가 결과를 기반으로 기초학력 향상을 위한 정책적 목적을 달성하고자 도입되었다. 국가수준 학업성취도 평가 전수시행에서는 일부 과목에 대한 표집시행을 병행하고 있으며,²⁾ 이를 기반으로 검사 동등화를 수행하여 학력의 추이를 점검하고, 서답형 문항의 채점기준을 결정하고 있다(김경희 외, 2011).

1998년 이후 국가수준 학업성취도 평가 시행의 다양한 변화 과정 속에서 학교, 학생의 시험 부담을 완화하고 각종 학력평가(예, 교과학습 진단평가, 기초학력 향상도평가 등)를 단순화하도록 유도하기 위해 2013년부터 중학교 학업성취도 평가 대상 교과가 3개 교과(국어, 영어, 수학)로 축소되었으며, 초등학교의 학업성취도 평가가 폐지되었다(교육부, 2013). 그러나 일선에서는 국가수준 학업성취도 평가 대상 교과의 축소와 초등학교 평가의 폐지로 인해 국가수준의 학력 점검 기능이 약화되고, 기초학력 나아가 전반적인 학력저하 현상을 초래한다는 우려가 제기되고 있다(김경희 외, 2014). 이에 따라 국가수준 학업성취도 평가의 발전 방안으로 교육 책무성과 정보공시를 위한 표집평가로의 전환이 고려되고 있으며, 특히 의무교육 과정에서 기초학력 진단을 위하여 초등학교 6학년을 대상으로 하는 국가수준 학업성취도 평가의 표집시행이 제안되고 있다(김경성 외, 2013; 김경희 외, 2014). 즉, 학교·학생의 시험 부담 완화와 동시에 국가수준의 학업성취 추이 파악을 위해 표집시행을 통한(특히 초등학교급에서) 학력수준의 정확한 진단과 예측이 요구되고 있으며, 기초학력 점검을 위해 과학적 절차(예, 표집방법, 표집 규모 등)에 기반한 표집시행이 보다 강조되고 있다(김경성 외, 2013; 김경희 외, 2014; 민경석, 2008).

현행 국가수준 학업성취도 평가의 표집규모는 검사 동등화를 위해 검사지 유형별 학생 수를 기준으로 결정된다. 즉, 검사지 유형별 목표 표본 수가 2,000명으로, 2012년까지 5종의 검사지를 포함하는 초등학교에서는 약 10,000명의 학생을 표본 크기로 설정하였다(김경희 외, 2014). 이와 같이 국가수준 학업성취도 평가의 표집은 학력추이 점검을 위한 측정이론의 제한적 목적으로 시행되어 왔기 때문에, 표집자료의 국가 및 시·도교육청 수준에 대한 대표성, 표본 통계치의 정확성 등에 대한 논의가 활성화되지 못했다(김현철, 1998; 민경석, 2008). 또한, 국가수준 학업성취도 평가를 통한 기초학력 점검에 대한 필요성에 따라 제안되는 초등학교 표집시행을 위해서는 기존 약 1.7%의 표본(10,000여명의 학생)의 모집단 대표성과 국가 및 시·도 수준의 표본 통계치에 대한 정확성 측면에서의 점검이 필요하다. 즉, 학교/학생 표집규

2) 2011, 2012년 초등학교 사회, 과학, 2013년 중학교 사회, 과학은 표집평가로 실시됨(김동영 외, 2013).

모와 절차에 대한 통계적 검증은 전국 및 교육청별 국가수준 학업성취도 평가 결과의 신뢰성과 정책적 활용에 있어 중요한 의미를 갖는다.

이 연구에서는 학업성취도 평가에서의 검사 동등화를 위한 초등학교 표집절차와 규모에 따른 국가 및 시·도교육청 수준 표본 통계치의 정확성을 평가하고(예, 표집오차), 국가수준 학업성취도 평가 표집시행의 중요성과 표집평가의 기능 확대가 지속적으로 논의됨에 따라 기존의 표집 방법(2단계 층화군집표집)과 학교 및 학생의 표본수를 증가시키는 두 가지 표집절차를 제시하고 각 방법별로 산출되는 표본 통계치 및 그 정확성을 비교하였다. 이 연구에서 제시한 대안적 표집절차 중 첫 번째는 ‘학교군집’ 표집방법으로 기존의 학업성취도 평가 표집방법과 학교 표집은 동일하지만, 학교 내 해당 학년의 모든 학급과 학생을 표본으로 포함시키는 방법으로 표집 학생 수가 크게 증가한다. 두 번째 방법인 ‘최소학교수’ 표집은 시·도 내 표집학교 수를 확대하는 방법으로, 기존 2단계 층화군집표집방법에서 시·도교육청별 표집학교 수의 하한계(예, 15개 학교)를 추가 조건으로 설정한다.

이 연구의 목적은 현행 국가수준 학업성취도 평가의 검사 동등화를 위한 표집방법에서 모집단 대표성과 표본 통계치의 특성을 평가하며, 다양한 수준(국가, 교육청)의 학업성취 정보의 추정을 위한 세 가지 표집절차(2단계 층화군집표집, 학교군집표집, 최소학교수방법)의 방법론적 비교에 제한되며, 국가수준 학업성취도 평가의 표집시행으로 전환의 필요성을 주장하는 것에 있지 않음을 밝힌다.

II. 학업성취도 평가 표집절차와 표집오차

국가수준 학업성취도 평가에서는 검사 동등화와 일부 교과에 표집시행을 위하여, 2단계 층화군집표집(two-stage stratified cluster sampling) 절차를 활용한다(김경희 외, 2011). 2단계 층화군집표집은 국제 학업성취도 비교연구(Joncas, 2008a; OECD, 2005)에서 활용되는 방법으로, 2단계 표집(multi-stage sampling), 층화표집(stratified sampling), 그리고 군집표집(cluster sampling)으로 구성된 상대적으로 복잡한 표집절차이다. 2단계 표집은 두 번의 단계에 걸쳐서 표본을 표집하는 방법으로 첫 번째 단계에서는 학교, 두 번째 단계에서 표집 학교 내 학급을 추출한다. 층화표집은 평가 대상의 학업성취도에 영향을 주는 특정 변인별로 구분하여, 집단별로 할당된 표본을 추출하는 것을 의미한다. 국가수준 학업성취도 평가의 표집에서는 외층과 내층의 두 개의 층화변수(stratification variables)를 이용한다. 마지막으로 군집표집은 평가 대상인 학생을 직접 추출하는 것이 아니라 이들이 모인 학급을 최종 표집대상으로 설정하는 것을 의미한다. 1단계 표집단위인 학교는 표집 가능한 학생 수에 비례한 확률

(Probability Proportionate to Size: PPS)에 의하여 표집하며, 2단계 표집단위인 학급은 표집된 학교에서 한 학급을 무선 표집한다.

국가수준 학업성취도 평가의 표집절차는 모집단을 정의하고, 표집틀(sampling frame)을 설정한 후 외층별 학교 표본을 할당하며, 내층을 고려하여 표집틀을 재정렬한다. 재정렬된 표집틀에서 표집 구간 및 선정 숫자에 따라 표집학교와 대체학교를 선정한 후에 학급을 표집한다. 국가수준 학업성취도 평가의 표집을 위한 목표모집단(target population)은 국가교육과정이 운영되지 않는 비정규학교와 평가 실시가 어려운 도서벽지 학교나 평가 대상학년의 학생 수가 15명 미만인 학교 및 신체적, 지적, 언어적 장애가 있는 학생을 제외한 우리나라 전체 학교 및 학생이다.

학교 표집절차에서 외층변수(explicit stratification variables)는 시·도교육청이 활용되고, 내층변수(implicit stratification variables)로 도시화(대도시, 중소도시, 읍면지역), 학교설립유형(국공립, 사립)을 구분한다. 이때 내층변수는 외층변수에 내재한 것으로 내층변수의 조합은 각 외층변수 내에서 학교순서를 결정하는 역할을 한다. 학급 및 학생의 표집은 학교 표집에서 선정된 학교 내에서 1개 학급을 임의추출하여 소속 학생 모두를 표본대상에 포함시킨다. 그러므로 개별 학교 내 학급규모가 동일하다면(혹은 유사하다면), 동일 학교 학생의 표집확률은 모든 학생에게 동일하게 된다. 만약 표집된 학교의 해당 학년 재적학생수가 표집최소기준(30명)을 미달하는 경우 전체학생을 조사대상으로 포함시킨다.

표본의 크기는 일반적으로 표집오차 수준에 따라 결정되며, 표집오차는 모집단의 특성과 표집방법과 관련된다(Cochran, 1977; Lohr, 2010). 표집방법의 가장 단순한 형태인 임의표집을 가정하고 학업성취도 평가에서의 표집규모를 기존 10,000명으로 설정할 때, 이에 따른 교과별 척도점수(200점 평균, 30점 표준편차)의 오차수준은 약 ± 0.60 에 해당한다.³⁾ 95% 신뢰수준에서의 오차범위 ± 0.60 은 교과별 척도점수의 표준편차 30점과 비교하여 매우 작은 값으로 10,000명의 학생 수는 국가수준에서의 교과별 척도점수 평균을 추정하는 데 규모가 큰 표본이라 할 수 있다. 물론 여기서 계산된 오차범위 ± 0.60 은 임의표집절차에 기반한 것으로 최종 단계에서 군집표집을 포함하는 2단계 층화군집표집방법을 고려한다면, 최종적인 오차 범위는 0.60보다 커질 것으로 예상된다.

국가수준 학업성취도 평가의 표집결과 혹은 표본 통계치 추정이 국가수준뿐만 아니라 시·도 교육청별로 제공되며, 표본 학생 수가 가장 작은 제주교육청(145명)의 95% 오차범위는 약 ± 4.84 로 계산된다. 이러한 제주 교육청의 오차범위는 국가수준 오차범위의 약 8배에 해당하며, 서울교육청(1,626명)의 95% 오차범위 ± 1.45 와 비교할 때도 매우 큰 범위라고 할 수 있다. 결국 표집에서 중요하게 고려할 점은 표본의 균일한 비율이 아니라 절대 크기으로써(Lohr, 2010),

3) 임의표집을 가정하여 오차수준은 $Z_{\alpha/2} \sqrt{\left(\frac{N-n}{N}\right) \frac{s^2}{n}}$ 으로 $Z_{\alpha/2}$ 는 1.96, N은 572662, n은 10000, s가 30일 경우 약 0.60으로 계산됨(<표 2> 참조).

신뢰로운 시·도교육청별 표본 통계치 산출을 위해서는 500 혹은 1,000명과 같이 표집규모의 하한선을 설정할 필요가 있다. 예를 들어, 제주 교육청을 과대표집(over-sampling)하여 표본 수(임의표집 기준)를 400명(TIMSS 표본의 최소 기준, Joncas, 2008b)으로 하는 경우 95% 오차범위는 ± 2.84 로 척도점수 표준편차(30점)의 약 10% 수준이며, 평균에 대한 표집오차는 95% 오차범위의 절반 수준인 1.42로 표준편차의 5%에 해당한다.

국가수준 학업성취도 평가의 표집방법은 학교 규모, 반복 표집 여부, 체계적 표집절차에 따른 학교 표집확률의 결정, 층화 변수의 활용 측면에서 고유한 특성을 갖는다. 이상의 표집절차적 특성은 표집오차와 직접적으로 관련되어, 표집오차를 최소화시키는 역할을 하는 반면, 복잡한 표집절차의 특성으로 표집오차를 계산하는 단순한 수식이 산출될 수 없다(Lohr, 2010). 이와 유사한 이유로 국제 학업성취도 비교연구인 TIMSS와 PISA는 반복표집방법(resampling methods)을 이용하여 표본 통계치의 표집오차를 산출한다. 구체적으로 TIMSS는 Jackknife 방법(Joncas, 2008b)을 활용하며, PISA는 BRR(Balanced Repeated Replication, OECD, 2002)방법을 이용한다. 두 가지 반복표집방법은 서로 유사함에도 불구하고, BRR 방법이 상대적으로 표집오차 추정에 안정적이라는 특성에 근거하여(OECD, 2012), 이 연구에서는 PISA에서 활용되는 BRR 방법 중 Fay의 교정공식(OECD, 2005)을 이용하여 국가수준 학업성취도 평가의 표본자료에 대한 표집오차를 산출하였다.

Ⅲ. 연구 방법

1. 모집단 및 표집절차

이 연구에서는 2011년 초등학교 국가수준 학업성취도 평가 전수자료(5,850개교, 584,268명)를 활용하였으며, 초등학교 6학년 모집단 중 학교수준에서 국가교육과정이 운영되지 않는 특수학교, 공민학교, 각종학교, 접근이 어려운 도서벽지 학교 및 평가 대상 학생수가 15명 미만인 학교, 학생수준에서 신체적, 지적, 언어적 장애 학생이 제외된 대상을(학교 4,431개교, 학생 572,662명)을 목표모집단으로 설정하였다.

초등학교급의 표집방법 및 표집규모에 따른 표본 통계치의 정확성을 평가하기 위하여 세 가지 표집절차 방법을 비교하였다. 첫째, 2단계 층화군집표집은 현행 국가수준 학업성취 평가에서 활용하고 있는 방법으로 시·도교육청으로 모집단을 층화하고, 재학 학생 수에 비례하여 교육청별 표본학교를 선정(도시화 및 학교유형을 고려)하고, 학교 내 1개 학급을 표본으로 포함하는 방법이다. 둘째, 학교군집표집은 학생 표본 수를 확대하기 위하여, 2단계 층화군집표집과 동일하게 학교를 선정하고, 선정된 학교에 재학하는 모든 학생을 표본으로 포함한다. 셋째, 최소학

교수표집은 표본의 교육청 단위 대표성을 높이기 위하여 교육청별 최소 표본학교 수를 15개로 설정하며, 2단계 층화군집표집 절차와 동일하게 학교를 설정하고 각 학교에서 1개 학급을 표본으로 선정하는 방법이다. 2011년 국가수준 학업성취도의 초등학교 목표모집단과 세 가지 표집 방법에 따른 학교 및 학생 분포는 <표 1>, <표 2>와 같다.

<표 1> 모집단과 표집방법별 표집학교 분포

시도	목표모집단			2단계		학교군집		최소학교수	
	학교 수	비율(%)	ICC ⁴⁾	학교 수	비율(%)	학교 수	비율(%)	학교 수	비율(%)
서울	591	13.34	0.08	59	17.40	59	17.40	59	15.69
부산	280	6.32	0.06	21	6.19	21	6.19	21	5.59
대구	209	4.72	0.06	18	5.31	18	5.31	18	4.79
인천	213	4.81	0.06	19	5.60	19	5.60	19	5.05
광주	141	3.18	0.06	13	3.83	13	3.83	15	3.99
대전	131	2.96	0.08	12	3.54	12	3.54	15	3.99
울산	109	2.46	0.03	9	2.65	9	2.65	15	3.99
경기	1,042	23.52	0.07	85	25.07	85	25.07	85	22.61
강원	174	3.93	0.05	10	2.95	10	2.95	15	3.99
충북	158	3.57	0.07	11	3.24	11	3.24	15	3.99
충남	255	5.75	0.05	14	4.13	14	4.13	15	3.99
전북	205	4.63	0.05	12	3.54	12	3.54	15	3.99
전남	246	5.55	0.05	12	3.54	12	3.54	15	3.99
경북	261	5.89	0.06	16	4.72	16	4.72	16	4.26
경남	335	7.56	0.07	23	6.78	23	6.78	23	6.12
제주	81	1.83	0.03	5	1.47	5	1.47	15	3.99
전국	4,431	100.00	0.07	339	100.00	339	100.00	376	100.00

<표 2> 모집단과 표집방법별 표본학생 분포

시도	모집단		2단계		학교군집		최소학교수	
	학생 수	비율(%)	학생 수	비율(%)	학생 수	비율(%)	학생 수	비율(%)
서울	98,937	17.28	1,626	16.92	12,071	19.97	1,626	15.45
부산	35,377	6.18	562	5.85	3,307	5.47	562	5.34
대구	30,791	5.38	509	5.30	3,372	5.58	509	4.84

4) ICC(Intra-class correlation)은 전체 분산 중에 학교 간 분산이 차지하는 비율로 이 연구에서는 국어 척도점수를 통해 산출하였다.

시도	모집단		2단계		학교군집		최소학교수	
	학생 수	비율(%)	학생 수	비율(%)	학생 수	비율(%)	학생 수	비율(%)
인천	32,535	5.68	505	5.26	3,448	5.71	505	4.80
광주	20,932	3.66	372	3.87	2,246	3.72	423	4.02
대전	19,084	3.33	343	3.57	2,121	3.51	415	3.94
울산	14,758	2.58	242	2.52	1,513	2.50	397	3.77
경기	147,146	25.70	2,563	26.68	16,028	26.52	2,563	24.36
강원	16,354	2.86	276	2.87	1,224	2.03	395	3.75
충북	17,794	3.11	287	2.99	1,620	2.68	389	3.70
충남	22,838	3.99	387	4.03	2,068	3.42	411	3.91
전북	21,017	3.67	349	3.63	1,943	3.22	416	3.95
전남	20,561	3.59	352	3.66	1,834	3.03	413	3.93
경북	27,750	4.85	456	4.75	2,653	4.39	456	4.33
경남	39,344	6.87	634	6.60	4,022	6.66	634	6.03
제주	7,444	1.30	145	1.51	961	1.59	408	3.88
전국	572,662	100.00	9,608	100.00	60,431	100.00	10,522	100.00

〈표 1〉에 제시된 학교 분포에서 2단계 층화군집표집과 학교군집표집은 동일한 학교 표본분포를 보이며(339개 학교), 소규모 교육청(예, 울산, 강원, 충북, 제주교육청 등)에서 매우 작은 수의 학교를 포함한다. 이에 반하여 최소학교수표집(376개 학교)은 교육청당 학교 수의 하한계를 설정함에 따라 전체 표집학교 수가 약간 증가하며, 학교 분포에서 대규모 교육청(예, 서울, 경기교육청 등)의 학교 수 비율이 낮아지고 소규모 교육청의 비율이 높아지는 특성을 보인다.

ICC의 경우 전체 학교를 대상으로 할 때 0.07로 매우 낮은 수준이었는데, 이는 초등학교를 대상으로 국어 척도점수를 활용하여 산출된 결과로 이해될 수 있다. 시·도교육청 수준에서는 대규모 교육청이 소규모 교육청에 비해 ICC가 높았으나 그 정도는 매우 미미하였다. 대전, 충북 등은 학교 간 차이가 서울과 유사한 정도임에도 2단계 층화군집표집에서 상대적으로 매우 적은 수의 학교가 표집되었다.

〈표 2〉의 학생 분포에서 학교군집표집(60,431명)은 2단계 층화군집표집(9,608명)보다 학생 수가 6배 이상 증가한다. 최소학교수표집(10,522명)은 기존 표집방법보다 약 900명 학생이 증가하며, 교육청별 학생 분포에서도 〈표 1〉의 학교 분포와 유사하게 대규모 교육청의 학생 비율이 낮아지고 소규모 교육청의 비율이 높아지는 경향을 보인다.

2. 표본 가중치

〈표 1〉과 〈표 2〉에 제시된 바와 같이 표본 학교와 학생의 모집단 대표성(표집 확률)은 교육 청별로 다르게 나타난다. 이에 따라, 표본의 모집단에 대한 대표성을 보다 적절히 반영하고, 표본 통계치의 정확성을 높이기 위하여 표본 가중치가 통계치 산출에 이용되었다(조지민, 정혜경, 2013; Joncas, 2008b). 국가수준 학업성취도 평가는 층화변수를 활용하여 학교를 선정한 후 최종단계에서 학급이 표집됨에 따라 학교 가중치와 학급 가중치가 식(1), 식(2)와 같이 설정된다.⁵⁾

$$\text{학교 가중치: } BW_{sc}^{hj} = \frac{M_h}{n_h m_{hj}} \quad \text{식(1)}$$

$$\text{학급 가중치: } BW_{cl}^{hj} = \frac{C_j}{c_j} \quad \text{식(2)}$$

여기서, BW_{sc}^{hj} 는 h층 j 학교의 가중치, n_h 는 h층의 표집학교 수, m_{hj} 는 h층 j 학교의 전체 학생 수, M_h 는 h층 모든 학교의 전체 학생 수를 의미한다. 학교 표집에서 선정된 학교 내의 모든 학급이 표본에 포함될 확률이 동일하다는 전제 하(임의표집)에서 BW_{cl}^{hj} 는 h층 j 학교에서 표집된 학급의 가중치, C_j 는 j 학교의 전체 학급 수, c_j 는 j 학교에서 표집된 학급 수를 의미한다. 예를 들어, 학교 당 1개 학급이 선정되는 경우 표집학급의 가중치는 해당 학교의 전체 학급 수가 된다. 최종 표집 가중치는 학교 가중치, 학급 가중치의 곱으로 나타낼 수 있으며, 식(3) 같이 산출할 수 있다.

$$\text{최종 표집 가중치: } W^{hij} = BW_{sc}^{hj} \times BW_{cl}^{hj} \quad \text{식(3)}$$

식(3)의 최종 표집 가중치에 활용되는 학교 가중치와 학급 가중치는 학교와 학급에 따라 다르지만, 동일 학급 내 학생은 동일한 가중치 값을 갖는다.⁶⁾

3. 표본 통계치와 표집오차

국가수준 학업성취도 평가의 주요 결과는 성취수준과 척도점수로 보고된다. 먼저 성취수준은

5) 표본에 포함된 학교, 학급, 학생은 모두 조사에 참여한 것으로 가정하며, 참여 여부에 따른 대체학교에 대한 조정 가중치는 고려하지 않았다.

6) 단, 학교군집표집에서는 학교 내 모든 학급과 학생이 표본에 포함되므로, 식(1)의 학교 가중치만이 이용되었다.

‘평가 대상 학년급 학생들이 성취할 것으로 기대하는 기본 내용의 이해 정도’를 나타내는 것으로 세부적으로 우수학력, 보통학력, 기초학력, 기초학력 미달의 4단계로 개별학생들에게 보고된다(김경희 외, 2014). 이에 따라 국가수준, 교육청별 성취수준은 각 수준에 해당하는 학생 비율로 평가된다. 두 번째, 국가수준 학업성취도 평가의 척도점수는 각 교과별 200점 평균, 30점 표준편차로 2010년에 재설정되었으며, 개인, 집단 수준의 성취도는 2010년 척도점수를 기준으로 매년 검사 동등화가 적용되어 연도 간 점수를 비교할 수 있다. 이에 따라, 이 연구에서는 세 가지 표집방법의 비교를 위하여 국가수준 및 교육청별 성취수준 비율과 척도점수 평균을 표본 통계치로 산출하여 모집단과 비교하였다.

모집단 통계치(모수)와 달리, 표본 통계치는 추출된 표본에 따라 변하는 값이며, 표본 통계치의 정확성은 표집오차의 크기에 따라 평가될 수 있다(Lohr, 2010). 전통적 표집방법(임의표집, 층화표집, 군집표집 등)에 따른 표본 통계치의 표집오차에 대한 수리식이 명확한 반면, 국가수준 학업성취도 평가와 같이 층화, 다단계, 군집표집과 이에 따른 가중치를 설정하는 복잡한 표집에서는 표집오차 산출을 위한 분석적 접근보다는 경험적 방법에 의한 표집오차가 계산된다(예, TIMSS, PISA). 이 논문에서는 세 가지 표집절차에 따른 표집오차를 산출하기 위하여 국제 학업성취도 비교 연구인 PISA에서 활용하는 경험적 표집오차 산출 절차인 Fay의 BRR방법을 이용하였으며, 구체적인 표집오차 산출 과정은 다음의 다섯 단계로 구성된다(OECD, 2002; 2005; 2012). 첫째, 표본에서 외층변수와 내층변수에 근거하여 학교 짝을 만든다. 둘째, 각 학교 짝에 대하여 순서대로 1부터 H의 숫자를 부여하여 가상의 층화변수를 만든다. 셋째, 가상의 층화변수 내의 두 개 학교 중 임의로 하나의 학교에는 1, 나머지 학교에는 0.5의 분산단위(variance unit)를 부여한다. 넷째, 반복계산을 위한 분산단위 변수의 개수는 4의 배수이며, 가상 층화변수와 동일하거나 커야한다. 다섯째, 분산단위 변수와 최종 가중치를 곱하여 반복표집 가중치를 설정한다. 이에 따라 표본 통계치에 대한 표집분산은 식(4)와 같다.

$$\sigma_{(\hat{\theta})}^2 = \frac{1}{G(1-k)^2} \sum_{j=1}^G (\hat{\theta}_{(j)} - \hat{\theta})^2 \quad \text{식(4)}$$

식(4)에서 $\hat{\theta}$ 은 전체 표본의 통계치를, $\hat{\theta}_{(j)}$ 는 j번째 반복에 의한 통계치를 의미하며, k는 Fay의 교정계수 0.5를 나타낸다. 식(4)로 계산된 표집분산의 제곱근을 구하여 통계치의 표집오차가 산출된다. 이 연구에서는 표본 가중치와 표집오차를 산출하기 위한 SPSS 명령문을 작성하였으며, IBM SPSS Statistics 21 Standard에 이를 적용하였다.

IV. 연구 결과

이 장에서는 앞서 논의된 세 가지 표집방법의 모집단 대표성과 표본 통계치를 비교하였다. 먼저 가중치를 이용한 학교, 학생 표본분포에 대한 모집단과의 유사성을 평가하였다. 또한 평가 결과 보고에 주요하게 활용되는 국가수준과 시·도교육청별 성취수준 비율 및 척도점수 평균의 표집오차에 따른 표본 통계치의 정확성을 비교하였다.

1. 표본의 모집단 대표성

세 가지 표집방법에 따라 표본 가중치를 부여한 후 학교 및 학생 분포를 <표 3>과 <표 4>에 제시하였다.

〈표 3〉 모집단과 표집방법별 표집 가중치를 적용한 학교 분포

시도	목표모집단		2단계			학교군집			최소학교수		
	학교 수	비율(%)	학교 수	비율(%)	비율 차이*	학교 수	비율(%)	비율 차이*	학교 수	비율(%)	비율 차이*
서울	591	13.34	603	13.33	-0.01	603	13.33	-0.01	603	12.56	-0.78
부산	280	6.32	258	5.70	-0.62	258	5.70	-0.62	258	5.37	-0.95
대구	209	4.72	206	4.55	-0.17	206	4.55	-0.17	206	4.29	-0.43
인천	213	4.81	198	4.37	-0.44	198	4.37	-0.44	198	4.11	-0.70
광주	141	3.18	136	3.01	-0.17	136	3.01	-0.17	146	3.05	-0.13
대전	131	2.96	140	3.09	0.13	140	3.09	0.13	158	3.30	0.34
울산	109	2.46	116	2.55	0.09	116	2.55	0.09	123	2.57	0.11
경기	1,042	23.52	1,049	23.18	-0.34	1,049	23.18	-0.34	1,049	21.85	-1.67
강원	174	3.93	244	5.39	1.46	244	5.39	1.46	278	5.78	1.85
충북	158	3.57	269	5.95	2.38	269	5.95	2.38	252	5.25	1.68
충남	255	5.75	306	6.75	1.00	306	6.75	1.00	349	7.26	1.51
전북	205	4.63	188	4.14	-0.49	188	4.14	-0.49	236	4.92	0.29
전남	246	5.55	202	4.47	-1.08	202	4.47	-1.08	300	6.25	0.70
경북	261	5.89	235	5.20	-0.69	235	5.20	-0.69	235	4.90	-0.99
경남	335	7.56	315	6.96	-0.60	315	6.96	-0.60	315	6.56	-1.00
제주	81	1.83	61	1.35	-0.48	61	1.35	-0.48	95	1.98	0.15
전국	4,431	100.00	4,525	100.00	-	4,525	100.00	-	4,802	100.00	-

* 비율차이는 표집방법별 비율과 목표모집단 비율의 차이임(표집방법별 비율 - 목표모집단 비율)

〈표 4〉 모집단과 표집방법별 표집 가중치를 적용한 학생 분포

시도	목표모집단		2단계			학교군집			최소학교수		
	학생 수	비율(%)	학생 수	비율(%)	비율 차이*	학생 수	비율(%)	비율 차이*	학생 수	비율(%)	비율 차이*
서울	98,937	17.28	97,920	17.15	-0.13	98,937	17.28	0.00	97,920	17.14	-0.14
부산	35,377	6.18	35,393	6.20	0.02	35,377	6.18	0.00	35,393	6.20	0.02
대구	30,791	5.38	30,667	5.37	-0.01	30,791	5.38	0.00	30,667	5.37	-0.01
인천	32,535	5.68	32,410	5.68	0.00	32,535	5.68	0.00	32,410	5.67	-0.01
광주	20,932	3.66	21,062	3.69	0.03	20,932	3.66	0.00	21,153	3.70	0.04
대전	19,084	3.33	19,031	3.33	0.00	19,084	3.33	0.00	19,150	3.35	0.02
울산	14,758	2.58	14,818	2.60	0.02	14,758	2.58	0.00	14,820	2.59	0.01
경기	147,146	25.70	146,695	25.69	-0.01	147,146	25.70	0.00	146,695	25.68	-0.02
강원	16,354	2.86	16,494	2.89	0.03	16,354	2.86	0.00	16,501	2.89	0.03
충북	17,794	3.11	17,826	3.12	0.01	17,794	3.11	0.00	17,877	3.13	0.02
충남	22,838	3.99	22,464	3.93	-0.06	22,838	3.99	0.00	22,489	3.94	-0.05
전북	21,017	3.67	20,828	3.65	-0.02	21,017	3.67	0.00	20,861	3.65	-0.02
전남	20,561	3.59	20,797	3.64	0.05	20,561	3.59	0.00	20,738	3.63	0.04
경북	27,750	4.85	27,999	4.90	0.05	27,750	4.85	0.00	27,999	4.90	0.05
경남	39,344	6.87	39,113	6.85	-0.02	39,344	6.87	0.00	39,113	6.85	-0.02
제주	7,444	1.30	7,455	1.31	0.01	7,444	1.30	0.00	7,444	1.30	0.00
전국	572,662	100.00	570,971	100.00	-	572,662	100.00	-	571,227	100.00	-

* 비율차이는 표집방법별 비율과 목표모집단 비율의 차이임(표집방법별 비율 - 목표모집단 비율)

〈표 3〉에 제시된 모집단과 표집방법별 가중치가 적용된 학교 분포를 살펴보면, 학교 표본이 동일한 2단계 층화군집표집과 학교군집표집의 학교 분포는 동일하다. 반면, 최소학교수표집은 교육청별 표집학교 수에 대한 최소기준을 설정함에 따라, 목표모집단과 비교하여 대규모 교육청을 포함한 8개의 교육청(서울, 대전, 대구, 인천, 광주, 경기, 경북, 경남)에서 표집학교 비율이 상대적으로 낮아지고 8개의 소규모 교육청(대전, 울산, 강원, 충북, 충남, 전북, 전남, 제주)의 표집학교 비율이 높아지는 경향을 보인다.

학교 및 학생 가중치를 설정한 후의 학생 분포는 목표모집단과의 비율 차이 절대값의 평균이 약 0.3%p 정도로 세 가지 표집방법에서 모두 목표모집단과 매우 유사한 분포로 나타났다. 특히, 학교 가중치만을 이용하며, 학생 수가 가장 큰 학교군집표집은 모집단의 학생 수와 정확히 일치하는 결과를 보인다. 모집단 분포와의 유사성과 전체 표집학교 및 학생 수를 고려할 때, 이러한 결과는 세 가지 표집방법 중, 현재 국가수준 학업성취도 평가에서 활용하고 있는 2단계 층화군집표집이 세 가지 절차 중 가장 효율적인 표집방법(최소 표본 학교 및 학생 수)임을 보여준다.

2. 국가수준의 표본 통계치 정확성

세 가지 표집방법에 따른 국가수준의 성취수준 비율, 척도점수 평균 추정치와 이에 따른 표집 오차를 <표 5>와 <표 6>에 제시하였다.

<표 5> 모집단과 표집방법별 국가수준의 성취수준 비율 및 표집오차

과목	성취수준	목표모집단	2단계				학교군집			최소학교수		
		비율(%)	비율(%)	비율 차이*	표집 오차	비율(%)	비율 차이*	표집 오차	비율(%)	비율 차이*	표집 오차	
국어	우수	40.79	42.45	1.66	0.75	40.98	0.19	0.60	42.05	1.26	0.71	
	보통	42.10	40.80	-1.30	0.59	41.75	-0.35	0.34	41.07	-1.03	0.56	
	기초	16.46	16.09	-0.37	0.50	16.60	0.14	0.41	16.19	-0.27	0.49	
	미달	0.65	0.66	0.01	0.08	0.67	0.02	0.05	0.69	0.04	0.07	
수학	우수	21.47	22.37	0.90	0.61	21.86	0.39	0.49	21.87	0.40	0.60	
	보통	58.53	58.00	-0.53	0.52	57.92	-0.61	0.32	58.19	-0.34	0.51	
	기초	18.96	18.52	-0.44	0.52	19.10	0.14	0.44	18.83	-0.13	0.50	
	미달	1.05	1.11	0.06	0.12	1.12	0.07	0.08	1.12	0.07	0.11	
영어	우수	71.39	71.48	0.09	0.76	71.21	-0.18	0.66	70.84	-0.55	0.72	
	보통	17.02	16.64	-0.38	0.48	17.09	0.07	0.35	17.13	0.11	0.47	
	기초	10.87	11.28	0.41	0.45	10.95	0.08	0.37	11.43	0.56	0.43	
	미달	0.71	0.60	-0.11	0.08	0.75	0.04	0.05	0.59	-0.12	0.08	

* 비율차이는 표집방법별 비율과 목표모집단 비율의 차이임(표집방법별 비율 - 목표모집단 비율)

<표 6> 모집단과 표집방법별 국가수준의 척도점수 평균 및 표집오차

과목	목표모집단		2단계				학교군집				최소학교수			
	평균	표준 편차	평균	평균차*	표준 편차	표집 오차	평균	평균차*	표준 편차	표집 오차	평균	평균차*	표준 편차	표집 오차
국어	201.70	23.05	202.20	0.50	23.00	0.38	201.73	0.03	23.15	0.33	202.02	0.32	23.06	0.36
수학	200.79	27.26	201.45	0.66	27.35	0.47	200.80	0.01	27.52	0.43	201.07	0.28	27.35	0.46
영어	205.06	29.14	205.16	0.10	29.14	0.54	204.88	-0.18	29.32	0.49	204.79	-0.27	29.25	0.50

* 평균차는 표집방법별 평균과 목표모집단 평균의 차이임(표집방법별 평균 - 목표모집단 평균)

표집방법별 국가수준의 성취수준 비율 추정치를 살펴보면, 목표모집단에서의 비율과 표집방법별 비율 차이의 절대값 평균이 2단계 군집표집, 학교군집표집, 최소학교수 표집에서 각각 0.52%, 0.19%, 0.41%였으며 척도점수 평균 차이의 절대값 평균은 각각 0.42점, 0.07점, 0.29점으로 산출되었다. 전반적으로 가장 많은 학생 수를 포함하는 학교군집표집의 표본 통계

치가 목표모집단과 가장 유사하였으며, 성취수준 비율의 경우 모집단에서의 비율이 큰 우수학력과 보통학력에서 모집단 비율과 차이가 상대적으로 크게 나타났다.

표집오차 또한 세 가지 방법 중 학교군집표집이 가장 작은 수준이었으나 표본 통계치와 표집오차 모두 방법 간 차이는 작게 나타났다. 또한, 성취수준별 결과에서 모집단에서의 비율 자체가 매우 작고 학교 간 편차 또한 상대적으로 작은 기초학력 미달 학생 비율에 대한 표집오차가 작게 나타났다. 이러한 결과는 세 가지 표집방법 간 표집오차에서 큰 차이가 없고, 국가수준의 통계치를 추정하는 데 세 방법 모두 적절한 방법임을 보여준다. 즉, 현재 국가수준 학업성취도 평가에 적용되고 있는 2단계 층화군집표집방법이 국가수준의 통계치를 추정하는 데 적절히 기능하고 있음을 보여준다.

3. 시·도교육청 수준의 표본 통계치 정확성

국가수준 학업성취도 평가는 국가수준의 학업성취도뿐만 아니라 단위 학교의 교수·학습과 교육과정의 장학을 담당하는 시·도교육청 수준의 학업성취도 정보가 중요하게 다루어져 왔으며, 매년 학업성취도 관련 보도자료를 통해 시·도교육청 수준의 성취수준 비율 추이 등을 보고하고 있다. 이에 따라 시·도교육청 수준에서의 표본 통계량의 정확성을 비교하였으며, 세 가지 표집방법에 따른 시·도교육청별 성취수준 비율과 척도점수 평균의 추정치를 <표 7>과 <표 8>에 제시하였다.⁷⁾

<표 7> 국어 교과의 모집단과 표집방법별 시·도교육청 수준의 성취수준 비율 및 표집오차

구분		목표모집단	2단계			학교군집			최소학교수		
		비율(%)	비율(%)	비율 차이*	표집 오차	비율(%)	비율 차이*	표집 오차	비율(%)	비율 차이*	표집 오차
서울	우수	42.07	42.91	0.84	2.00	41.03	-1.04	1.64	42.91	0.84	2.00
	보통	39.83	37.95	-1.88	1.34	40.19	0.36	0.81	37.94	-1.89	1.34
	기초	17.26	18.25	0.99	1.35	18.00	0.74	1.02	18.25	0.99	1.35
	미달	0.84	0.90	0.06	0.23	0.79	-0.05	0.14	0.90	0.06	0.23
부산	우수	44.90	48.73	3.83	1.71	47.35	2.45	1.88	48.73	3.83	1.73
	보통	41.15	38.51	-2.64	1.12	40.53	-0.62	1.29	38.51	-2.64	1.11
	기초	13.50	12.26	-1.24	1.82	11.63	-1.87	1.00	12.26	-1.24	1.83
	미달	0.46	0.50	0.04	0.17	0.49	0.03	0.24	0.50	0.04	0.17

7) 이 연구에서는 2011년 당시의 16개 시·도교육청(세종시 제외)에 대한 분석을 실시하였다. 또한 제한된 지면을 고려하여 국어 교과에 대한 시·도교육청 결과를 제시였다. 본문에는 제시되지 않은 영어와 수학 교과의 결과는 국어 교과와 유사함을 밝힌다.

구분		목표모집단	2단계				학교군집			최소학교수		
		비율(%)	비율(%)	비율 차이*	표집 오차	비율(%)	비율 차이*	표집 오차	비율(%)	비율 차이*	표집 오차	
대구	우수	48.00	45.47	-2.53	3.88	47.55	-0.45	2.61	45.47	-2.53	3.88	
	보통	39.99	42.02	2.03	3.69	39.76	-0.23	1.62	42.02	2.03	3.68	
	기초	11.61	11.62	0.01	1.46	12.09	0.48	1.02	11.62	0.01	1.46	
	미달	0.40	0.89	0.49	0.30	0.61	0.21	0.20	0.89	0.49	0.30	
인천	우수	39.83	43.67	3.84	3.58	42.18	2.35	2.29	43.67	3.84	3.60	
	보통	43.32	41.87	-1.45	2.83	41.43	-1.89	1.18	41.87	-1.45	2.85	
	기초	16.24	14.02	-2.22	1.56	15.80	-0.44	1.71	14.02	-2.22	1.56	
	미달	0.62	0.44	-0.18	0.32	0.60	-0.02	0.11	0.44	-0.18	0.31	
광주	우수	36.17	34.99	-1.18	4.55	34.05	-2.12	2.00	35.29	-0.88	3.66	
	보통	43.26	42.33	-0.93	3.41	42.99	-0.27	1.43	42.19	-1.07	2.55	
	기초	19.75	21.84	2.09	2.95	21.40	1.65	2.12	21.79	2.04	2.98	
	미달	0.82	0.84	0.02	0.45	1.56	0.74	0.43	0.72	-0.10	0.40	
대전	우수	47.86	46.93	-0.93	4.07	46.17	-1.69	3.80	44.53	-3.33	2.02	
	보통	39.48	40.24	0.76	2.99	38.64	-0.84	1.49	41.90	2.42	2.57	
	기초	12.17	11.91	-0.26	2.04	14.18	2.01	2.33	12.32	0.15	1.26	
	미달	0.49	0.93	0.44	0.69	1.01	0.52	0.38	1.24	0.75	0.17	
울산	우수	39.79	41.14	1.35	4.95	40.55	0.76	2.56	40.40	0.61	3.03	
	보통	42.42	39.06	-3.36	4.42	41.96	-0.46	1.26	40.35	-2.07	3.30	
	기초	17.14	19.80	2.66	2.21	17.08	-0.06	1.57	18.73	1.59	1.83	
	미달	0.64	-	-	-	0.42	-0.22	0.16	0.52	-0.12	0.37	
경기	우수	37.62	40.63	3.01	1.31	38.04	0.42	1.04	40.63	3.01	1.31	
	보통	42.89	41.58	-1.31	1.08	43.46	0.57	0.61	41.58	-1.31	1.08	
	기초	18.77	17.10	-1.67	1.02	17.84	-0.93	0.78	17.09	-1.68	1.02	
	미달	0.72	0.70	-0.02	0.16	0.66	-0.06	0.07	0.70	-0.02	0.16	
강원	우수	37.82	34.96	-2.86	2.40	34.34	-3.48	3.12	36.36	-1.46	2.09	
	보통	42.18	44.56	2.38	3.48	43.06	0.88	2.50	43.92	1.74	2.68	
	기초	19.20	19.30	0.10	1.89	21.35	2.15	3.25	18.94	-0.26	2.00	
	미달	0.80	1.18	0.38	0.68	1.25	0.45	0.38	0.79	-0.01	0.46	
충북	우수	41.12	50.77	9.65	6.01	48.68	7.56	5.18	45.95	4.83	4.57	
	보통	44.73	38.69	-6.04	4.13	41.35	-3.38	3.63	43.48	-1.25	3.41	
	기초	13.78	10.55	-3.23	2.42	9.88	-3.90	1.94	10.57	-3.21	2.28	
	미달	0.38	-	-	-	0.08	-0.30	0.06	-	-	-	

구분		목표모집단	2단계				학교군집			최소학교수		
		비율(%)	비율(%)	비율 차이*	표집 오차	비율(%)	비율 차이*	표집 오차	비율(%)	비율 차이*	표집 오차	
충남	우수	39.20	42.21	3.01	2.19	39.19	-0.01	2.82	41.61	2.41	2.59	
	보통	44.36	38.57	-5.79	2.50	42.10	-2.26	1.81	39.35	-5.01	2.51	
	기초	15.90	18.97	3.07	3.91	18.21	2.31	3.31	18.81	2.91	3.54	
	미달	0.55	0.25	-0.30	0.25	0.50	-0.05	0.24	0.23	-0.32	0.23	
전북	우수	39.35	38.26	-1.09	3.19	37.72	-1.63	1.22	36.71	-2.64	2.72	
	보통	43.33	44.16	0.83	2.56	42.64	-0.69	0.97	43.82	0.49	2.27	
	기초	16.58	16.71	0.13	2.25	18.71	2.13	1.48	18.34	1.76	2.34	
	미달	0.74	0.87	0.13	0.49	0.93	0.19	0.28	1.12	0.38	0.35	
전남	우수	35.69	34.79	-0.90	4.35	33.53	-2.16	3.44	34.95	-0.74	3.75	
	보통	44.70	44.91	0.21	3.80	44.52	-0.18	1.98	44.94	0.24	3.28	
	기초	18.91	19.18	0.27	4.19	21.60	2.69	3.05	18.96	0.05	3.49	
	미달	0.70	1.12	0.42	0.39	0.35	-0.35	0.13	1.15	0.45	0.40	
경북	우수	43.76	44.47	0.71	1.23	43.91	0.15	2.50	44.47	0.71	1.23	
	보통	41.35	40.83	-0.52	0.91	40.28	-1.07	0.95	40.83	-0.52	0.91	
	기초	14.30	14.27	-0.03	1.63	15.32	1.02	1.85	14.27	-0.03	1.63	
	미달	0.58	0.43	-0.15	0.30	0.50	-0.08	0.30	0.43	-0.15	0.30	
경남	우수	43.57	43.47	-0.10	4.00	44.25	0.68	3.16	43.47	-0.10	4.00	
	보통	42.97	42.73	-0.24	2.89	41.59	-1.38	1.94	42.73	-0.24	2.89	
	기초	13.09	13.59	0.50	1.64	13.71	0.62	1.47	13.59	0.50	1.64	
	미달	0.37	0.22	-0.15	0.22	0.45	0.08	0.13	0.22	-0.15	0.22	
제주	우수	41.17	54.79	13.62	3.63	46.07	4.90	2.00	45.17	4.00	3.48	
	보통	41.59	36.98	-4.61	3.59	41.23	-0.36	0.65	39.67	-1.92	3.24	
	기초	16.71	7.53	-9.18	1.35	12.37	-4.34	1.37	13.90	-2.81	1.64	
	미달	0.52	0.70	0.18	0.56	0.33	-0.19	0.13	1.26	0.74	0.61	

* 비율차이는 표집방법별 비율과 목표모집단 비율의 차이임(표집방법별 비율 - 목표모집단 비율)

〈표 8〉 국어 교과의 모집단과 표집방법별 시·도교육청 수준의 척도점수 평균 및 표집오차

구분	목표모집단		2단계				학교군집				최소학교수			
	평균	표준 편차	평균	평균차*	표준 편차	표집 오차	평균	평균차*	표준 편차	표집 오차	평균	평균차*	표준 편차	표집 오차
서울	201.86	24.34	201.87	0.01	24.78	1.10	201.40	-0.46	24.41	0.94	201.87	0.01	24.78	1.10
부산	204.12	22.38	206.03	1.91	22.61	1.21	205.63	1.51	21.90	1.16	206.03	1.91	22.61	1.22
대구	205.83	22.04	204.98	-0.85	21.75	1.72	205.33	-0.50	22.33	1.40	204.98	-0.85	21.75	1.72

구분	목표모집단		2단계				학교군집				최소학교수			
	평균	표준 편차	평균	평균차*	표준 편차	표집 오차	평균	평균차*	표준 편차	표집 오차	평균	평균차*	표준 편차	표집 오차
인천	201.42	22.51	203.58	2.16	22.11	1.54	202.15	0.73	23.00	1.38	203.58	2.16	22.11	1.55
광주	198.96	23.44	197.83	-1.13	24.31	2.90	197.06	-1.90	24.76	1.86	197.87	-1.09	24.06	2.48
대전	205.54	22.49	204.94	-0.60	24.08	2.36	203.61	-1.93	23.50	2.36	203.53	-2.01	24.71	0.78
울산	201.17	23.16	200.31	-0.86	22.23	1.80	201.35	0.18	22.11	1.32	200.49	-0.68	23.06	1.09
경기	199.80	23.21	201.29	1.49	22.76	0.70	200.41	0.61	23.00	0.60	201.29	1.49	22.76	0.70
강원	200.09	24.02	197.99	-2.10	23.51	1.48	197.62	-2.47	24.17	1.84	198.39	-1.70	23.63	1.40
충북	202.69	20.87	206.58	3.89	20.20	2.21	205.84	3.15	20.20	1.80	205.35	2.66	19.80	1.74
충남	201.41	22.12	202.02	0.61	22.34	1.69	201.72	0.31	22.52	1.73	201.84	0.43	22.10	1.55
전북	201.10	22.81	199.67	-1.43	22.57	1.27	200.21	-0.89	23.39	0.65	198.62	-2.48	23.31	1.42
전남	198.97	22.63	197.29	-1.68	23.16	2.23	198.24	-0.73	23.09	2.00	197.22	-1.75	23.14	1.81
경북	203.43	22.70	202.70	-0.73	22.26	0.80	202.86	-0.57	22.93	1.46	202.70	-0.73	22.26	0.80
경남	203.90	21.38	203.83	-0.07	20.96	1.68	203.69	-0.21	21.47	1.38	203.83	-0.07	20.96	1.67
제주	201.50	22.72	207.67	6.17	20.82	2.11	204.88	3.38	21.48	1.37	202.37	0.87	23.23	1.43

* 평균차는 표집방법별 평균과 목표모집단 평균의 차이임(표집방법별 평균 - 목표모집단 평균)

〈표 7〉에 제시된 시·도교육청별 성취수준 비율의 비교에서, 현행 2단계 충화군집표집은 소규모 교육청(예, 울산, 충북, 충남, 제주교육청 등)에서는 상대적으로 성취수준 비율을 과대, 과소 추정하는 경향이 나타났다. 특히 일부 교육청(울산, 충북)에서는 표집학교 수 및 학생 수가 적어, 모집단 비율 자체가 작은 기초학력 미달 학생 비율이 추정되지 못하는 결과로 이어졌다. 이에 반하여 표집된 학교의 전체 학생을 표본대상으로 하는 학교군집표집에서는 성취수준 비율이 모든 시·도교육청에서 추정되었으며, 상대적으로 비율 추정의 정확성이 높아지는 경향(작은 표집오차)을 보였다. 특히 전반적으로 우수와 보통학력 학생의 비율 추정에 있어 표집오차가 크게 줄어들었다. 그러나 소규모 교육청 및 기초학력 미달 학생 비율에 대한 과대, 과소 비율 추정은 여전히 나타났다.

시·도교육청 단위 최소 표집학교 수를 설정하여 학교 당 1학급을 표본으로 추출한 최소학교수표집은 2단계 충화군집표집에서 누락된 소규모 교육청의 기초학력 미달 학생 비율을 일부 추정하며(예, 울산 교육청), 또한 소규모 교육청의 성취수준 비율을 세 가지 방법 중 상대적으로 가장 정확히 추정하는 것으로 나타났다. 이는 2단계 충화군집표집, 학교군집표집과 비교하여 소규모 교육청에 보다 많은 학교를 표본에 포함함에 따라 나타난 결과라고 할 수 있다. 즉, 현행 2단계 충화군집표집에 비해 한 학교의 모든 학생을 표집하는 학교군집표집은 우수와 보통학력 학생 비율 추정의 정확성을 높이고, 최소학교수표집은 기초학력 및 기초학력 미달 학생 비율 추

정의 정확성을 높여준다. 이는 최소학교수표집이 학업성취도 평가 표집시행에서의 기초학력 점검 기능을 강화하기 위한 기초자료를 제공하는 데 적절한 방법임을 의미한다. 또한, 현재 시·도교육청 수준에서 기초학력과 기초학력 미달 학생 비율은 모집단에서의 비율 자체가 작으며, 이 또한 특정 학교에 분포되어 있는 것이 아니라 전반적으로 퍼져 있음을 고려할 때 학교군집표집과 최소학교수표집의 결과는 표집과정에서 시·도교육청 내의 학교 간 분산과 학교 내 분산을 고려해야 함을 의미한다.

〈표 8〉의 시·도교육청 수준의 척도점수 평균 추정치를 보면, 세 가지 표집방법에서 모두 대규모 시·도교육청에 대한 정확한 추정결과를 산출한다. 그러나 〈표 7〉의 시·도교육청별 성취수준 비율 추정 결과와 유사하게 소규모 교육청, 특히 제주의 경우에는 표집된 학교 수가 상대적으로 많은 최소학교수표집에서 시·도교육청 척도점수 평균 추정에 있어 보다 정확성을 높이고 있는 것으로 나타났다.

V. 결론 및 논의

이 논문은 국가수준 학업성취도 평가의 추정 결과의 측정이론적 안정성을 검증함과 동시에 검사 동등화를 위한 표집에 적용되는 2단계 층화군집표집의 적절성을 학교군집표집, 최소학교수표집과 비교하여 평가하였다. 이 연구에서 제안된 학교군집표집은 기존의 방법과 비교하여 표집된 학교에 소속된 모든 학생을 표본에 포함하며, 최소학교수표집은 층화변수로 이용되는 시·도교육청별 최소 표집학교 수(15개 학교)를 설정하였다.

연구결과를 요약하면 먼저, 학교와 학생의 단순 표본분포에서 2단계 층화군집표집(339개 학교, 9,608명 학생)과 비교하여 학교군집표집(339개 학교, 60,431명 학생)은 학생 수가 6배 이상 증가하며, 최소학교수표집(376개 학교, 10,522명 학생)은 학교와 학생 수에 있어 약간의 증가를 보인다(〈표 1〉, 〈표 2〉 참조). 세 가지 표집방법은 서로 다른 표집 확률로 학교와 학생을 선택하기 때문에, 이를 반영한 가중치를 적용하면, 모집단에 대한 대표성을 판단할 수 있는 교육청별 학생 분포에서 큰 차이를 보이지 않으며, 분포 형태 또한 모집단과 매우 유사하게 나타났다(〈표 4〉 참조). 학업성취도 평가의 주요 결과인 성취수준 비율과 척도점수 평균 추정치 비교에서 국가수준의 표본 통계치 정확성을 확인한 결과, 세 가지 표집방법에서 모두 모집단과 유사한 추정치를 산출하였다. 그러나 국어 교과의 우수학력 성취수준 비율에서 모집단은 40.79%, 2단계 층화군집표집은 42.45%, 학교군집표집은 40.98%, 최소학교수표집은 42.04%으로 전국수준임에도 불구하고 상대적으로 큰 편차를 보인다(〈표 5〉 참조). 물론 표집오차를 반영한 95%의 신뢰구간을 이용하면, 세 가지 방법 모두 모집단 비율을 포함한다는 점에서 통계적 정

확성에 오류가 없으나, 2단계 층화군집표집과 최소학교수표집이 상대적으로 큰 오차를 보인다. 또한 모든 통계치에서 일관되게, 표본 수(학생 수)의 크기에 따라, 학교군집표집, 최소학교수표집, 2단계 층화군집표집 순서로 전국수준 통계치의 표집오차가 커졌다.

시·도교육청 수준에서 표본 통계치의 정확성을 확인한 결과, 세 가지 표집방법에서 모두 대규모 시·도교육청의 통계치는 안정적으로 추정하는 반면, 작은 규모의 시·도교육청에서는 2단계 층화군집표집과 학교군집표집이 과대 혹은 과소 추정되어 편기된 결과를 보였다(〈표 7〉, 〈표 8〉 참조). 특히, 이러한 표본 통계치의 편기(biasness)는 2단계 층화군집표집에서 가장 강하게 나타나며, 시·도교육청당 최소 학교 수를 설정한 최소학교수표집에서 가장 약하게 나타났다. 구체적으로, 이 논문의 2011년 자료 포함되지 않은 최소 규모 교육청인 세종시(2013년 현재 23개 초등학교, 3,489명 학생, 교육통계서비스, 2014)를 예로 들면, 현재 2단계 층화군집표집(1.7% 표본)에서는 약 60명의 학생이 표본 학생 수로 설정되며, 이는 2개 학교(학교당 1개 학급, 학급당 평균 30명 학생)만이 표본에 포함됨을 의미한다. 이때 학교군집표집을 이용한 학생 수를 확대하기보다는 15개 표본학교와 같이 교육청당 최소 학교 수를 설정하는 것이 소규모 교육청별 자료 산출에 보다 정확성을 높일 것이다. 또한 표본 수가 가장 적은 2단계 층화군집표집에서 일부 교육청의 기초학력 미달 학생 비율이 추정되지 못하는 경우가 자주 나타난다. 결국, 표집학생의 규모는 학교군집표집이 가장 큼에도 불구하고, 개별 시·도교육청 수준의 통계치, 특히 기초학력 또는 기초학력 미달 학생과 관련된 통계치 산출을 위해서는 시·도교육청별 최소 학교 수를 설정한 최소학교수표집이 가장 효율적인 결과를 제공한다.

모집단 추정을 위한 표본 통계치의 정확성을 나타내는 표집오차에 영향을 미치는 요인은 표집규모, 모집단 분산, 표집절차 등으로 구분될 수 있는데, 이 논문의 결과를 통해 국가수준 학업성취도 평가의 표집과 관련되어 다음 세 가지 측면을 논의하고자 한다. 먼저, 표집규모의 경우 현행 2단계 층화군집표집방법은 국제 비교연구인 TIMSS, PISA에서 활용하고 있는 방법으로 국가수준을 대표하는 자료를 수집하기 위해서는 매우 안정적인 것으로 평가된다. 특히 국제 비교연구의 표집규모인 150개 학교, 5,000여명의 학생과 비교하여 현행 국가수준 학업성취도 평가의 표집규모는 2배 이상으로 상대적으로 정확히 국가수준의 학업성취도를 추정할 수 있다. 그러나 국제 학업성취도 비교연구는 주로 국가 간 비교에 주안점을 두는 반면, 국가수준 학업성취도 평가는 국가수준뿐만 아니라 단위 학교의 장학에 담당하는 교육청 수준의 정보를 요구한다. 이에 따라 검사 동등화라는 제한적 목적으로 이용되는 현행 2단계 층화군집표집방법은 국가수준의 척도점수 추정에는 안정적이지만, 시·도교육청 수준 정보(특히, 성취수준 비율)를 제공하기 위해서는 학교 수 및 학생 수를 확대하는 방안(예, 학교군집표집, 최소학교수표집)을 고려할 필요가 있다.

두 번째, 모집단 분산이라는 측면에서 2단계 층화군집표집은 학교 간 분산과 학교 내 분산을 고려할 필요가 있다(Raudenbush & Bryk, 2002). 특히, 소규모 시·도교육청의 경우 학교 규모 및 소재 지역이 다양하며, 연구결과 모집단에서 시·도교육청 내 학교 간 차이는 대규모

교육청과 거의 유사한 반면 2단계 층화군집표집과 학교군집표집에서의 표집학교 수는 상대적으로 매우 적어 모집단에서의 학교 간 차이를 반영하는 데 한계가 있을 수 있다. 이러한 점을 고려할 때 동일학교 내 표본 학생 수를 확대하기(학교군집표집)보다는 표집학교 수를 확대하여 다양한 학교를 표본에 포함하는 방안(최소학교수표집)이 교육청별 통계치 산출에서 보다 정확한 결과를 제공한다. 이 연구에서 최소 학교 수를 15개 학교로 설정한 결과, 기초학력과 기초학력 미달 학생 비율 추정에 있어 표집오차가 줄어든 것을 확인할 수 있었으나, 여전히 성취수준별 비율 추정에 있어 그 정확성이 떨어지는 시·도교육청이 존재하였다. 이러한 점은 최소 학교 수의 시·도교육청별 일괄 적용 보다는 시·도교육청별로 학교 간 분산 및 학교 내 분산을 확인하여 기준을 달리해야 할 가능성을 보여주며, 표집시행의 편이성 등을 고려하여 시·도교육청별 일괄 적용을 시도할 경우에도 적정 최소 학교 수에 대한 검증이 필요함을 보여준다.

마지막으로 표집절차 측면에서 실제 표집을 수행할 때에는 현행 2단계 층화군집표집에서 포함된 층화표집과 군집표집이라는 두 가지 상반된 특성을 고려해야 한다. 일반적으로 결과변수(학업성취도)에 영향을 미치는 층화변수의 통제는 표집오차를 줄이는 반면, 동일 집단에 소속된 표본을 반복 추출하는 군집표집은 표집오차를 증가시키는 경향을 보인다(Cochran, 1977; Lohr, 2010). 그러나 현실적 평가 상황을 고려할 때 군집표집은 시행의 행정적, 인적 비용의 효율성을 높인다는 장점을 갖는다. 또한 단위 학교의 교육과정 운영에 있어 임의표집된 몇몇 학생을 선택하기보다는 학급단위 혹은 학교전체에 일괄적으로 평가를 시행하는 것이 수월할 것이다. 이에 따라 국가수준 학업성취도 평가의 표집 결과의 정확성을 높이기 위해서는(특히 교육청 수준에서) 학교 다양성을 반영한 표집절차를 고려해야 할 것이다. 또한 학교 수준과 학생 수준의 표집범위에 대한 논의는 통계적 결과인 표집오차뿐만 아니라 조사 시행 비용이라는 경제적 효율성(신민웅, 신기일, 2000; 신민웅, 오상훈, 2001)이 고려될 필요가 있다.

참 고 문 헌

- 교육부(2013). **2013 국가수준 학업성취도 평가 기본계획(안)**. 세종: 교육부.
- 교육통계서비스(2014). 2013년 유초중등통계. 출처: <http://kess.kedi.re.kr/index> (2014년 5월 29일 검색).
- 김경성, 김도기, 김성식, 김준엽, 상경아, 양길석(2013). **국가수준 학업성취도 평가 개선 연구**. 교육부 수탁과제 2013-1. 세종: 교육부.
- 김경희, 김완수, 최인봉, 김희경, 상경아, 신진아, 박인용, 김성훈(2011). 2011년 국가수준 학업성취도 평가 기술보고서. 한국교육과정평가원 연구보고 RRE 2011-2-5.
- 김경희, 김성숙, 시기자, 김수진, 김완수, 신진아, 박인용, 구남옥, 구슬기, 노은희, 이인호, 김도남, 김부미, 우석진(2014). 국가수준의 기초학력 점검을 위한 초등학교 학업성취도 평가 방안. 한국교육과정평가원 연구보고 CRE 2014-2.
- 김동영, 김도남, 김영란, 김현정, 이정우, 서민철, 조운동, 이광상, 이인호, 심재호, 배주경, 정기문, 최원호, 박영신(2013). 2013 국가수준 학업성취도 평가 출제 연구. 한국교육과정평가원 연구보고 RRE 2013-4.
- 김현철(1998). 층화군비추출법과 층화추출법을 사용한 교육조사에서 표본의 효율적인 할당방법과 그 예시. **교육학연구**, 36, 149-169.
- 민경석(2008). 국가수준 학업성취도 평가의 표집규모와 표집오차. **한국교육**, 35(1), 171-190.
- 신민웅, 신기일(2000). 층화 2단 표본 추출시 최적 집락의 크기 결정. **응용통계연구**, 13, 207-224.
- 신민웅, 오상훈(2001). 층화 이단계 표본 추출시 최적 선택율. **응용통계연구**, 14, 429-437.
- 조지민, 정혜경(2013). PISA 표집 설계에 따른 모수 및 분산 추정. **교육평가연구**, 26(4), 875-896.
- Cochran, W. G. (1977). *Sampling technique* (3rd ed.). New York, NY: Wiley.
- Joncas, M. (2008a). TIMSS 2007 Sample design. In Olson, J. F., Martin, M. O., & Mullis, I. V. S. (Eds.). *TIMSS 2007 Technical report*, 77-92. IEA.
- Joncas, M. (2008b). TIMSS 2007 Sample weights and participation rates. In Olson, J. F., Martin, M. O., & Mullis, I. V. S. (Eds.). *TIMSS 2007 Technical report*, 153-192. IEA.
- Lohr, S. L. (2010). *Sampling: Design and analysis* (2nd ed.). Boston, MA: Brooks/Cole.
- OECD (2002). *PISA 2000 Technical report*. PISA/OECD.
- OECD (2005). *PISA 2003 Data analysis manual*, SPSS Users. PISA/OECD.

OECD (2012). *PISA 2009 Technical report*. PISA/OECD.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.

• 논문접수 : 2014-04-30/ 수정본접수 : 2014-06-01/ 게재승인 : 2014-06-13

ABSTRACT

Comparison of Sampling Methods for National Assessment of Educational Achievement in Elementary Schools

Kyung-Seok Min

(Associate Professor, Sejong University)

In-Yong Park

(Associate Research Fellow, Korea Institute for Curriculum and Evaluation)

Kyunghee Kim

(Senior Research Fellow, Korea Institute for Curriculum and Evaluation)

This study evaluated three sampling methods on a representativeness perspective and compared accuracy of national and provincial sample statistics. Three sampling methods included 1) two-stage stratified cluster sampling method already adopted in NAEA, 2) school-cluster sampling method selecting all student in sampled schools, and 3) two-stage stratified cluster sampling with a minimum number of schools selecting one class per a sampled school under the condition of minimum number of schools in 16 provinces. The results showed that all sample distributions with sampling weights based on each sampling method sufficiently represented the population distribution. And all sampling methods also yielded accurate national sample statistics. However, the accuracy of provincial sample estimates depended on the size of a province (i.e. the number of sampled schools and students). That is, In large provinces, three sampling methods provided accurate sample statistics. whereas, in small provinces with small number of schools and students, the third method (two-stage stratified cluster sampling with a minimum number of schools) yielded more accurate estimates (i.e., less sampling errors) than those of other sampling methods.

Key Words : National Assessment of Educational Achievement(NAEA), sampling method, sampling error