

국가수준 학업성취도 평가의 표집설계 개선 방안 연구¹⁾

박 인 용(한국교육과정평가원 부연구위원)*

정 혜 경(한국교육과정평가원 부연구위원)**

김 완 수(한국교육과정평가원 연구위원)

<요 약>

본 연구는 국가수준 학업성취도 평가가 표집 체제로 전환됨에 따라 국가수준의 표본 통계 결과의 정확성 및 대표성을 제고하기 위한 목적으로 시뮬레이션 연구를 수행함으로써 기존의 전수 체제에서 적용되었던 표집설계에 대한 개선 방안을 제시하였다. 이를 위해 2016년 학업성취도 평가 전수 자료를 토대로 표집 방법, 표집 규모, 외층 변인, 목표 표집 학교 수 결정을 위한 기준학급 크기에 대한 조건을 다양화하여 표집 설계에 적용하고 각각에서 산출된 표본 통계와 모집단 통계를 비교하였다. 연구 결과, 군집 표집보다는 층화표집의 적정성 및 다른 구간에 비해 1.5%에서 3% 표집 시 정확도 향상의 효율성을 확인하였으며, 표집 자료의 대표성 제고 차원에서 학생 수 감소에 따라 실제 학급 당 학생 수를 반영할 수 있는 기준학급 크기 선정, 지역규모를 반영한 외층 변인의 확대 방안이 제시되었다. 도출된 시사점은 학업성취도 평가뿐만 아니라 대규모 평가에서의 표집 설계 개선에 유용한 정보로 활용될 수 있다.

주제어 : 국가수준 학업성취도 평가, 표집설계, 층화 표집, 외층 변인, 표집 규모

1) 본 연구는 2017년 한국교육과정평가원에서 수행한 “국가수준 학업성취도 평가의 표집설계 개선 방안”(박인용 외, 2017) 연구의 내용을 수정·보완한 것임.

* 제1저자, iypark@kice.re.kr

** 교신저자, hjung@kice.re.kr

I. 서론

1. 연구의 필요성

국가수준 학업성취도 평가(이하 학업성취도 평가)는 매년 학생의 학력 수준을 파악하고 변화 추이를 점검함으로써 국가 교육과정이 학교 현장에서 적용된 이후 학생에 미치는 교육적 성과를 체계적으로 점검하는데 그 목적이 있다. 따라서 학업성취도 평가 결과는 1998년 기본 계획 수립 이후 지속적으로 학교 교육의 질 관리, 기초학력 점검, 교육과정 개선, 단위학교의 교수·학습 방법 개선, 교육정책 결정 등의 기초 자료로 활용해왔다. 자료 수집을 위하여 1986년 전수평가로 시작하였으나 1998년 ‘국가수준 교육성취도 평가 기본 계획(김명숙 외, 1998)’이 수립되면서 시행의 효율성과 경제성을 위하여 표집 평가로 시행되었다. 이후 학생들의 기초학력 보장을 위한 정책이 강화되면서 2008년 전수평가로 전환된 이후 2016년까지 전수체제가 유지되었다.

그러나 2017년 6월 전국 시·도교육감 협의회는 전수체제의 학업성취도 평가가 시·도 및 학교를 서열화하고 계층화하는 자료로 남용되고 있다고 비판하면서 표집 평가로 전환할 것을 제안하였다. 이에 국정기획자문위원회는 학업성취도 평가의 원래 취지에 부합되도록 2017년 학업성취도 평가부터 표집평가 방식으로 전환하되, 표집학교 이외의 모든 학교는 자율에 따라 시행할 것을 교육부에 제안하였으며, 교육부에서는 이러한 제안을 받아들여 2017년을 기점으로 전수평가를 폐지하고, 중학교 3학년, 고등학교 2학년 표집 학교는 의무시행, 이외의 학교는 시·도교육청에서 자율적으로 시행하도록 결정하였다(교육부, 2017.06.14.). 결과적으로 2017년 6월 학업성취도 평가의 표집 시행이 전면 도입되면서 시행령 개정을 추진하여 개별 학교의 학업성취도 평가 결과를 공시하지 않고 시·도교육청별 결과 발표도 폐지되었으며, 결과적으로 국가수준의 평가 결과만을 발표하기로 하였다(교육부, 2017.06.14.).

본 연구는 학업성취도 전수 평가 체제 하에서도 별도의 표집 평가 시행을 병행해 왔기 때문에 기존의 표집 설계 방식의 개선안을 모색하는데서 출발하였다. 다만 전수평가 체제에서의 표집 시행 결과는 전수 결과의 변화 추이 파악, 전수 학생 서답형 문항 채점의 표준화 등 평가 결과 산출의 안정성 확보 차원에서 약 1.5%의 학생을 별도로 표집하고 이 결과를 동등화 및 채점 기준 검점에 활용해 왔다는 측면에서 그 목적과 취지가 일반적인 표집평가와는 차이가 있다. 따라서 본 연구는 동등화를 위한 목적으로 활용되었던 기존의 표집 설계 방식을 검토하고, 표집 방법 및 표집 규모의 조건을 다양화하여 모의자료를 생성하고 그에 따른 표본

통계 결과의 정확성과 안정성을 비교·분석해 봄으로써 표집자료에 근거한 학업성취도 평가 결과가 국가수준의 학력수준을 진단하는데 있어 타당하고 유용한 정보로 활용될 수 있도록 그 토대를 마련하는데 주요 목적이 있다. 본 연구 결과는 국가단위의 표집 평가 및 조사 연구를 수행하고자 하는 일반 연구자들에게 평가 결과의 정확성 및 시행의 효율성을 고려하여 적절한 표집 설계 방법을 선정하는데 실증적인 데이터 및 가이드라인을 제공한다는 측면에서 그 의의가 있다.

II. 학업성취도 평가 표집 설계

일반적으로 표집 조사 또는 표집 평가는 자료 수집의 용이성, 통계적 효율성, 비용 측면에서 전수 조사보다 선호된다. 표집 조사를 위해서는 표본 또는 표집 설계(sampling design)가 필요한데 조사 대상을 통계적으로 추출하는 방법이라 할 수 있다. 표집 설계에서는 목표 모집단(target population), 표집 모집단(sampled population) 관찰 단위(observation unit), 표집 단위(sampling unit), 표집 틀(sampling frame) 등에 대한 개념을 명확히 하고, 이상적인 표집은 전체 모집단의 모든 특성을 반영(mirroring)하여 모집단의 축소 버전을 생성하는 것이라 할 수 있다(Lohr, 2010, p.3). 표본 추출을 위해서는 목적과 상황에 따라 단순 임의 추출, 계통 추출, 층화 추출, 집락(또는 군집) 추출 등 다양한 방법들이 상황과 조건에 따라 적용된다(박민규, 강현철, 2016). 학업성취도 평가에서는 여러 표본추출 방식들이 복합적으로 적용된 경우로, 2단계 층화 군집 표집 방법(two-stage stratified cluster sampling)을 통해 시도, 지역규모, 학교 설립유형 및 성별유형의 비율을 반영하여 표집학교를 선정하고, 표집학교 선정 후, 학교 내 학급을 무선 표집하며 해당학급 학생을 최종 표집학생으로 선정하고 있다.

층화 표집은 모집단 특성 정보 또는 보조정보를 활용하여 서로 겹치지 않는 (외)층(stratum)을 나누고 각 층별로 독립적으로 표본을 추출하는 방식을 의미한다(Lohr, 2010, p.74). 이 표집 방식은 층내 표집 대상 간의 이질성을 줄여 모집단 대표성을 확보하는데 효율적이며, 특정 하위 모집단에 대한 추정량 산출 시 표집 단계부터 이를 반영하여 층별 표집 규모를 결정하는데 용이할 뿐만 아니라, 층화 표집 방식과 다른 표본 추출 방식과의 연계가 쉽다는 장점이 있다(IEA, 2016, p.3.12; 박민규, 강현철, 2016, p.74). 이에 학업성취도 평가에서도 2단계 층화 군집 표집을 적용하고 있으며, 1단계에서 시·도교육청을 외층변인으로 하여 각 층별로 독립적으로 표집이 이루어지며, 층별 표집 시 조사단위(학생)의 군집(cluster)인 학교를 우선적으로 표집하게 된다. 군집 표집은 표집 단위가 제공하는 정보가 상대적으로 독립적이지 않기 때문에 임의 표집에 비해 정확성(precision)이 축소되는 단점에도 불구하고(Lewis,

2017, p.17), 학업성취도 평가에서도 군집 추출을 하는 이유는 조사단위(학생)에 대한 추출틀이 직접적으로 존재하지 않기 때문이며, 반면에 교육통계연보 등의 정보를 활용하여 시도별 모집단 학교 정보가 가능하기 때문이다. 또한 군집 표집은 학생에 대한 임의 표집에 비해 조사 비용이나 시행의 편의성 측면에서 매우 효율적이라 할 수 있으며, 이런 현실적인 이유로 인해 군집 표집은 대규모 조사에서 자주 사용된다(박민규, 강현철, 2016, p.137).

좀 더 구체적으로 학업성취도 평가의 2단계 표집 설계 절차를 살펴보면 다음과 같다. 먼저 목표 모집단(국가 교육과정을 이수하는 중학교 3학년, 고등학교 2학년 재학생 전체)을 구체적으로 명시하고 이를 바탕으로 전년도 정보를 활용하여 학교 표집틀을 구성하는데, 학교 표집틀에는 학교명, 주소, 외층 및 내층 변인, 학교평균 학생 수, 해당 학년 총 학생 수 등의 정보를 포함한다.

전수 체제에서의 표집 설계에 대해 단계적으로 살펴보면, 먼저 표집 규모(1.5%) 기준 외층 변인인 시·도교육청 학생 모집단 비율에 맞추어 각 외층단위별 표집 학생 수를 확정한 후(모집단 크기 \times 외층별 비율), 외층별 필요 표집 학생 수를 기준학급 크기(35명)로 나누어 표집 학교 수를 확정한다. 다음으로 17개 시·도교육청별 학생을 표집하기 위해서는 2단계에 걸쳐 확률 추출이 일어나는데, 일차추출단위(primary sampling unit: PSU)인 학교 표집의 경우 확률비례추출(probability proportional-to-size sampling; PPS)과 계통추출(systematic sampling)을 결합한 방식을 적용하여 학교를 표집하고 있다. 이론상 PPS 방식은 조사단위 개체가 통계 추정치에 미치는 중요도인 크기(규모) 변수를 표집틀에 반영함으로써 규모가 클수록 해당 표집 단계에서 표집될 확률이 커지도록 표집되는 특징을 갖는다(박민규, 강현철, 2016, p.138). 한편 계통추출 표집은 표본추출간격(sampling interval; I)을 정하고, 임의의 출발점부터 매 번 I만큼 떨어진 간격에 위치하는 표집단위를 표본에 포함하는 방식으로, 다른 표본추출 방식과 결합되어 표본의 대표성을 높이게 된다(박민규, 강현철, 2016, p.57). 학업성취도 평가 표집의 또 다른 특징으로는 PPS와 계통추출 방식을 적용하기 전 시·도교육청별 내층변인(지역규모, 학교설립유형, 학교성별유형, 학교 규모 등)을 기준으로 학교 데이터를 정렬하게 된다. 층화변인으로 학교를 정렬하게 되면 근접학교의 특성이 유사하게 됨에 따라 결과적으로 내층변인과 계통 추출을 병합한 표집은 모든 내층변인에 소속된 다양한 특성의 학생들을 비율에 맞춰서 표집하는데 효과적이며(IEA, 2016, p.50), 계통추출 또한 설계변수(design variable)를 활용하여 모집단 데이터를 정렬한 후에 자료를 표집하는 것이 표집의 효율성을 높이는 것으로 알려져 있다(박민규, 강현철, 2016, p.57, pp. 60-61). 끝으로 일 단계에서 제시된 방식을 활용하여 표집 학교를 선정하고 나면 이 단계에서는 1개의 학급을 임의 표집하여 최종적으로 조사대상인 표집학생을 결정하게 된다.

이처럼 기존의 전수 체제 하에서의 학업성취도 평가 표집 대상 선정은 비교적 체계적이고 정교한 방식을 유지하고 있으나, 전수 평가에서의 표집의 목적이 표집 자료의 모집단 대표성

및 표집 데이터에 기반한 모수 추정의 정확성 등을 담보하는 측면이라기 보다는 전수자료 채집 및 점수 산출의 안정성 확보 및 채집의 효율성 제고를 위한 목적으로 표집을 시행하였다. 따라서 전수 평가에서 표집 평가 체제로 전환된 시점에서 기존의 표집 체제에서 유지되는 주요 요소 즉, 표집 규모, 외층 변인 선정, 기준 학급 크기에 대한 재검토가 필요하다. 즉, 전수 체제에서 표집의 목적이 동등화의 안정성을 위해 서로 다른 가교 문항으로 구성된 4종의 검사지 각각에 안정적인 모수 추정을 위한 최소 인원 확보 차원에서 검사지 당 2,000명을 산정하여 총 8,000여명의 표집 규모를 유지하여 왔다. 따라서 표집 평가 전환에 따라 선정한 표집 규모에 대한 적정성, 기존 시도교육청 이외의 외층 변인 확대 및 그에 따른 효과, 학생 인구 감소에 따라 학급당 기대되는 학생 수 즉 기준학급크기를 35명으로 적용하는 것에 대한 적정성 등에 대한 보다 면밀한 검토가 필요하겠다.

실제로 우리나라 학업성취도 평가와 표집 설계가 유사한 TIMSS(Trends in International Mathematics and Science Study) 국제 학업학성취도 평가에서도 한국의 표집 데이터 산출을 위한 외층 변인, 학급당 평균 학생 수 등에 대한 변화 연구를 수행하여 2015년 예비시행부터 외층 변인으로 도시화 정도와 학교 성별 유형을 선정하고 학급 당 평균 학생 수를 25명으로 추정하여 반영하였다(김수진 외, 2014, p.33). 또한 TIMSS 국제 본부와의 연구 결과 기존의 외층 변인으로 사용된 시도·교육청 변인을 제외하고 학급당 평균 학생 수의 축소에 따라, 기존에 적용된 실제 모집단 특성을 제대로 반영하기 위한 표집 학교 개수가 150개에서 219개가 필요하다는 연구 결과를 발표하였다(김수진 외, 2014, p.39). 그러나 TIMSS 2015 본검사에서 표집학교 수를 늘리기 보다는 가중치와 같은 통계 방법으로 대표성을 제고하는 방안을 모색함으로써 실제로 한국의 중학교 2학년 모집단에 대해 150개 학교 5,509명을 표집하여 평가를 시행하였다(김수진 외, 2015, p.40). 이처럼 학업성취도 평가와 동일하게 2단계 층화 군집 표집 방식을 적용하는 TIMSS 연구 결과는 학업성취도 평가의 표집 설계 변화를 모색하는데 있어 실증적인 정보가 된다. 더불어 OECD가 주관하는 PISA(Programme for International Student Assessment) 또한 2단계 층화 비례 설계 방식을 적용하여 학생을 표집하고 있는데, 만 15세 학생을 표집하기 위해 외층 변인으로 학교급과 계열을 구분하여 중학교, 일반계 고등학교, 비일반계 고등학교 3개의 외층을 활용하고 있으며, 부가적으로 도시화 정도(대도시/중소도시/읍면지역), 학생 성별(남학교/여학교남녀공학)을 내층변인으로 사용하여 PISA 2015 본검사 기준 우리나라의 경우 총 168개교 5,749명을 표집하였다(구자옥 외, 2015, pp.61-64). 다만 TIMSS나 PISA의 경우 우리나라 학생들의 학업성취 정도를 국제 수준에서 비교하는데 그 주요 목적이 있으므로 국가수준뿐만 아니라 교육정책 및 필요에 따라 시도교육청별 학업성취도 결과의 정확성 및 대표성까지를 고려하는 국가수준 학업성취도 평가와 비교할 때 상대적으로 표집 학생 수 규모가 작을 수 있다. 한편 미국의 유일한 국가수준 학업성취도 평가라 할 수 있는 NAEP의 경우 주기별로 국가수준의 결과만 산출할 경우와 주정부별 결과도 함께 산출할 경우에 표집 설계를 달리함으로써 평가 결과 추정의 정확성을 확보하고 있다(NECS, 2019a;

2019b). 특히 주정부별 결과 산출 까지 고려할 경우 주정부별 약 6,000여명 이상을 표집하고 학교당 63명을 기준으로 표집하고 있어 국가수준으로 볼 때 표집 규모 또한 TIMSS나 PISA에 비해 클 뿐만 아니라, 외층 변인으로 50개 주정부 정보를 활용함으로써 국가단위의 평가 결과뿐만 아니라 주정부의 평가 결과에 대해서도 추정의 정확성 및 평가 결과의 대표성을 제고하고자 시도하고 있다. 이처럼 표집 규모 및 층화 변인은 평가의 목적과 활용 정도에 따라 달라질 수 있음을 확인할 수 있으며, 학업성취도 평가의 목적과 결과 활용까지를 고려하여 표집 설계 방안을 마련할 필요가 있다.

Ⅲ. 연구 방법

1. 시뮬레이션 연구 설계

표집체계 하에서 학업성취도 평가의 제 기능을 유지할 수 있는 적정 규모의 표집방법과 표집규모를 탐색하기 위하여 기존 학업성취도 평가의 표집설계에서의 표집방법과 표집규모의 조건을 달리하여 시뮬레이션 분석을 수행하였다. 먼저 2016년 학업성취도 평가 전수자료를 활용하여, 다양한 표집 방법과 표집 규모의 조건에 따라 반복적으로 표집학생을 선정하고, 각 조건별 표집 분석 결과와 모집단 분석 결과를 비교함으로써 표집 방법과 표집 규모에 따른 결과의 정확성과 대표성을 점검하고자 하였다.

반복적으로 표집학생을 선정하기 위해 모집단으로 2016년 중학교 3학년, 고등학교 2학년 학업성취도 평가 교과별 전수자료를 활용하였다. 표집 규모는 현재 검사지 유형별 표집 학생인 2,000명을 기준으로 4개 유형의 총 약 8,000명을 표집 규모로 활용하는 현행 방식(이는 모집단 학생 수의 약 1.5%에 해당)과 더불어 표집을 통한 평가 결과의 정확성 제고를 위해 현실 적용 가능한 범위 안에서 표집 규모를 약 3%, 5%, 10%로 확대하는 방안을 고려하였다.

표집 방법은 현행 표집 방법에서 고려되는 군집 방법의 설정, 외층의 설정, 기준학급의 크기 설정을 변화시키며 표집하였다. 먼저 군집 방법의 경우 현행 방식인 학교, 학급을 고려한 다단계 임의 표집 방식을 적용하였다. 다만 학교 내 표집 학급수를 현행 방식은 1학급 고정에서 학교 내 학급 수를 늘리는 방안이 고려하였다(단, 시행 상의 어려움과 교사 설문과의 직접 연계를 위해 학교 내 학생 임의표집은 고려하지 않음).

<표 1> 군집 방법 조건

군집방법	내용	특징
임의1학급 표집	· 학교 내 1학급 임의 표집 후 학급 내 학생 전체 표집	· 표집 규모 증가 시 표집학교 수 증가(학교 간 차이 반영에 용이)
임의N학급 표집	· 학교 내 여러 학급(N학급) 임의 표집 후 학급 내 학생 전체 표집	· 표집학교 수 고정 시(1.5%에서의 학교 수), 표집 규모가 증가할수록 학교 내 학급 수 증가(학교 내 학급 간 차이 반영에 용이)
임의학교 표집	· 학교 내 학생 전체 표집	· 표집 규모 고정 시 임의1학급표집과 임의N학급표집보다는 표집 학교 수 현저히 감소

외층 설정의 경우 현행 방식에서는 시·도 교육청을 외층으로 설정하고 있으며, 이 연구에서는 시·도 교육청 외층 방식과 더불어 시·도 교육청과 지역규모의 교차 층을 외층으로 설정하여 외층의 세분화가 표집 오차에 주는 영향력을 탐색하고자 하였다.

<표 2> 외층 설정 조건

외층	내용	특징
시도외층	· 외층: 시·도 교육청, 내층: 지역규모, 성별유형, 설립유형(현행)	· 현행 방식으로 시·도교육청을 고려하여 대표성 확보
교차외층	· 외층: 시·도 교육청과 지역규모 교차외층, 내층: 성별유형, 설립유형	· 시·도교육청과 지역규모를 동시에 고려하여 대표성 확보

기준학급 크기(학급 당 또는 학교 당 기준학생)는 현행 방식의 경우 목표 표집 학교 수 결정을 위해 35명을 기준으로 하고 있다. 즉, 현행 방식은 1학교에 1학급을 무선 추출하는 방식이기 때문에 각 시도별로 표집 학교 수를 결정하기 위해 1학급 포함되는 기준 학생 수를 35명으로 설정하여 산출되는 학교 수 만큼을 표집하는 것이다. 이러한 방식은 현재 우리나라의 학생 수가 줄어들고 있으며, 외층에 따라 실제 학급 당 학생 수를 반영하기 어려운 제한점이 있어 이 연구에서는 현행 방식의 35명과 더불어 우리나라 전체 학급 당 학생 수의 평균을 활용하는 방안과 외층별로 학급 당 학생 수의 평균을 활용하는 방안으로 표집 결과의 정확성을 점검하고자 하였다.

모집단을 토대로 표집 조건에 따라 각 100회를 시행하며, 시뮬레이션 절차는 다음과 같다.

- ① 조건별 표본 추출: 모집단에서 각 표집 방법 및 표집 규모에 따라 표집 학생 반복 추출

<표 3> 시뮬레이션 변수 및 변수의 수준

시뮬레이션 변수	시뮬레이션 변수의 수준
표집규모	1.5%, 3%, 5%, 10%
외층설정 방법	시·도교육청 외층, 시·도교육청과 지역규모의 교차외층
군집방법	1학급, N학급, 학교
기준학급 크기(학급 당 학생수)	35명, 모집단 평균, 외층별 평균

② 조건별 분석 항목 산출

- 성취수준별 학생 비율: 국가수준의 분석에 대한 정확성과 안정성 점검과 더불어 학교수준에서의 성취수준 비율 분석의 정확성과 안정성을 확인하는 지표로 활용
- 전체 분산 대비 학교 간 분산 비율(ICC): 학업성취도 평가의 기초학력 점검 기능 유지와 기초학력 미달 비율이 사례 수에 보다 민감할 수 있다는 점을 고려할 때 표집 방법 간 비교가 가능한 지표임. ICC(Intra-class correlation)는 학교 간 차이를 반영하는 지표로서 활용

<표 4> 결과 분석 항목

구분	분석 항목
전체 데이터	성취수준별 학생 비율, 전체 분산 대비 학교 간 분산의 비율(ICC)
학교별 데이터	성취수준별 학생 비율

③ 결과 분석: 준거 지표에 근거하여 각 결과 분석 항목별 정확성과 안정성 비교

이상의 시뮬레이션 절차를 통한 데이터 생성 설계와 데이터 생성 조건을 요약하면 다음과 같다.

<표 5> 시뮬레이션 데이터 생성 설계

표집 규모	군집 방법	외층 및 기준학급 크기 구분					
		시도 외층			교차외층		
		35명	전체평균	외층평균	35명	전체평균	외층평균
1.5%	1학급	100세트(현행)	100세트	100세트	100세트	100세트	100세트
	N학급	100세트	100세트	100세트	100세트	100세트	100세트
	학교	-	100세트	100세트	-	100세트	100세트
3%	1학급	100세트	100세트	100세트	100세트	100세트	100세트
	N학급	100세트	100세트	100세트	100세트	100세트	100세트
	학교	-	100세트	100세트	-	100세트	100세트
5%	1학급	100세트	100세트	100세트	100세트	100세트	100세트
	N학급	100세트	100세트	100세트	100세트	100세트	100세트
	학교	-	100세트	100세트	-	100세트	100세트
10%	1학급	100세트	100세트	100세트	100세트	100세트	100세트
	N학급	100세트	100세트	100세트	100세트	100세트	100세트
	학교	-	100세트	100세트	-	100세트	100세트

2. 분석대상

이 연구에서 활용한 모집단에서의 성취수준 비율과 학교 간 차이를 보여주는 ICC는 <표 6>과 같다.

<표 6> 모집단 성취수준 비율과 ICC

학교급	학교 수	학생 수	교과목	성취수준				ICC
				우수학력	보통학력	기초학력	기초미달	
중학교	3,094	512,807	국어	37.5%	52.6%	7.9%	2.0%	9.0%
			수학	19.9%	48.3%	26.9%	4.9%	14.8%
			영어	28.0%	46.7%	21.3%	4.0%	16.7%
고등학교	1,780	456,091	국어	32.6%	50.9%	13.1%	3.4%	27.5%
			수학	30.1%	47.5%	17.0%	5.4%	28.3%
			영어	41.0%	44.4%	9.2%	5.4%	31.9%

모집단은 3,094개 중학교의 512,807명이며 국어, 수학, 영어의 기초학력 미달 학생 비율은 각각 2.0%, 4.9%, 4.0%로 나타났다. 또한, 전체분산에서 학교 간 분산(학교 간 차이)의 비율의 지표인 ICC는 각각 9.0%, 14.8%, 16.7%로 나타나 학교 간 차이가 전체 차이 대비 영어에서 가장 큰 것을 볼 수 있다. 고등학교의 경우 모집단은 1,780개 학교의 456,091명이며, 국어, 수학, 영어의 기초학력 미달 학생 비율은 각각 3.4%, 5.4%, 5.4%이었다. 또한, 전체분산에서 학교 간 분산(학교 간 차이)의 비율은 각각 27.5%, 28.3%, 31.9%로 나타나 고등학교에서의 학교 간 차이가 비교적 크게 나타난 것을 볼 수 있다.

이 연구에서는 <표 6>의 모집단에서 표집 방법과 표집 규모에 따라 학교 및 학생을 표집 하였으며, 반복적으로 추출된 학교 수는 <표 7>과 같다.

<표 7> 표집 방법 및 규모에 따른 학교 수

외 층	기준 학급크기	군집 방법	중학교				고등학교			
			1.5%	3%	5%	10%	1.5%	3%	5%	10%
시 도 외 층	35명	1학급	236	464	695	1,380	236	465	694	1,363
		N학급	236	236	236	236	236	236	236	236
	전체 평균	1학급	281	554	827	1,646	272	535	798	1,484
		N학급	281	281	281	281	272	272	272	272
		학교	55	104	152	298	41	72	102	196
	외층 평균	1학급	282	556	825	1,644	271	533	796	1,500
		N학급	282	282	282	282	271	271	271	271
		학교	56	105	154	300	38	70	102	197
	35명	1학급	242	471	699	1,386	245	476	698	1,341
		N학급	242	242	242	242	245	245	245	245
교 차 외 층	전체 평균	1학급	288	561	835	1,641	277	540	807	1,452
		N학급	288	288	288	288	277	277	277	277
		학교	65	114	161	303	50	81	110	202
	외층 평균	1학급	290	562	834	1,649	278	541	803	1,466
		N학급	290	290	290	290	278	278	278	278
		학교	65	111	160	303	48	79	107	199
	35명	1학급	242	471	699	1,386	245	476	698	1,341
		N학급	242	242	242	242	245	245	245	245

다른 조건이 동일할 경우, 전반적으로 시도외층보다는 교차외층 적용 시 보다 많은 학교가 표집되었다(예, 학급당 학생 수 35명 기준에서 표집 규모 1.5% 적용 시, 시도외층에서는 236개 학교, 교차외층에서는 242개교가 산출됨). 또한, 학급당 학생 수의 기준을 35명으로 할 경우에 비해 실제 학급당 평균 학생 수를 반영한 경우 보다 많은 학교가 표집되는 것을 볼 수 있다(예, 시도외층, 표집 규모 1.5% 적용 시, 35명 방식은 236개교가 표집되었고 전체평균 방식은 281개교가 표집됨). 다만 학급당 학생 수의 기준을 전체평균 방식과 외층평균 방식을 적용하여 산출된 표집 학교 수는 거의 유사하였다.

외층 변인과 기준학급 크기 조건이 동일하고 표집 규모만 증가시킬 경우, 학교 전체를 군집 표집하는 경우 필요로 하는 표집 학교 수가 현저히 낮은 것을 볼 수 있다. 또한, 동일 조건에서 1학급 방식과 N학급 방식을 비교하면, 표집 규모가 커지면 1학급 방식에서는 필요로 하는 학교 수도 커지지만, N학급 방식은 학교 내 표집되는 학급 수만 증가하여 학교 수는 동일하게 유지되고 있다.

표집 방법과 표집 규모에 따라 반복적으로 추출된 학생 수는 <표 8>과 같다.

<표 8> 표집 방법 및 규모에 따른 학생 수

외 층	기준 학급크기	군집 방법	중학교				고등학교			
			1.5%	3%	5%	10%	1.5%	3%	5%	10%
시 도 외 층	35명	1학급	7,152	14,063	21,072	41,774	7,345	14,481	21,613	41,673
		N학급	7,152	14,249	21,236	40,621	7,345	14,662	21,961	43,179
	전체 평균	1학급	8,513	16,785	25,076	49,513	8,452	16,660	24,841	44,947
		N학급	8,513	16,947	25,250	48,331	8,452	16,913	25,329	49,672
		학교	13,429	25,479	37,216	73,087	12,755	22,670	32,223	62,084
	외층 평균	1학급	8,521	16,786	24,931	49,417	8,412	16,532	24,714	45,343
		N학급	8,521	16,976	25,281	48,375	8,412	16,811	25,197	49,482
		학교	13,301	24,973	36,681	71,399	11,575	21,198	31,173	60,202
교 차 외 층	35명	1학급	7,318	14,256	21,170	41,920	7,606	14,805	21,695	41,104
		N학급	7,318	14,551	21,668	41,388	7,606	15,233	22,770	44,699
	전체 평균	1학급	8,711	16,984	25,270	49,405	8,615	16,799	25,103	44,152
		N학급	8,711	17,331	25,826	49,324	8,615	17,187	25,737	50,544
		학교	15,412	27,404	38,831	73,810	14,502	24,609	33,800	63,164
	외층 평균	1학급	8,701	16,899	25,086	49,369	8,602	16,734	24,874	44,493
		N학급	8,701	17,312	25,750	48,998	8,602	17,176	25,728	50,343
		학교	14,035	24,237	35,045	66,528	13,459	22,574	30,657	57,772

시·도 교육청 외층 방법을 적용할 경우 시·도 교육청별로 학생을 추출하되, 시·도 교육청별로 표집해야 할 학생 수는 모집단에서의 시·도교육청별 학생 비율에 따라 산정된다. 이러한 방식으로 시·도 교육청별 표집 학생 수가 결정되면, 해당 학생 수를 추출하기 위한 학교 수 산출이 필요하다. 학교 수 산출은 기준 학급 크기(학급 당 기준 학생 수)와 군집방법에 따라 다르게 산출되는데, 대체로 기준 학급 크기가 35명일 경우 가장 적게 표집 되었으며, 전체 평균, 외층 평균 순으로 나타났다. 또한, 군집 방법에 따라 1학급의 경우 학교 당 1학급, N학급의 경우 학교 당 여러 학급, 학교 군집 표집에서는 학교 전체 학생을 표집하는데, 대체로 군집방법으로 학교 당 1학급으로 할 경우 가장 작은 수의 학생이 표집되었고, N학급, 학교 순으로 나타났다. 표집 방식의 예를 들면, 전체 표집 규모가 1,000명이고 A시도의 학생이 전체 모집단에서 10%에 해당한다면, 시·도 교육청 외층은 시도A에서 100명을 표집하며, 이를 위해 학교 수를 결정해야 된다. 100명을 표집하기 위한 표집학교 수 결정에 있어, 기준 학급 크기를 35명으로 하고, 군집방법을 1학급으로 할 경우 각 학교에서 1개의 학급 학생만 표집되기 때문에, 표집학교 수는 3개($100 \div 35$) 학교가 되며, 각 3개의 학교에서 1학급을 임의추출하여 표집학생을 선정한다.

시·도 교육청 외층에 학급 당 학생 수 기준을 35명으로 하고 학교 당 1학급을 추출할 경우, 표집규모 1.5% 적용 시 7,152명이 표집되었으며, 3%인 경우에는 약 2배에 해당하는 14,063명이 표집되었다. 시·도교육청 외층에 학급 당 학생 수 기준을 전체 학급의 평균 학생 수로 하고 학교당 1학급을 표집할 경우, 표집규모가 1.5%인 경우에는 약 8,513명, 3%일 경우에는 16,785명이 표집되어, 학급 당 학생 수 35명 기준보다는 약간 더 많은 학생이 표집되었다. 표집 방법에 따른 학교와 학생 수를 종합적으로 정리하면, 먼저 외층 방법에서는 시도외층 대비 교차외층 적용 시 추출된 학교 수 및 학생 수가 다소 증가하였다.

기준 학급 크기의 경우 다른 조건이 동일할 경우, 현재 적용 중인 학급당 학생 수 35명 기준 표집 방법보다는 실제 학급당 평균 학생 수를 반영한 표집 방안 적용 시, 학교 수 및 학생 수 모두 좀 더 많은 표집을 필요로 하였다. 기준 학급 크기에 따른 차이가 학생 수 보다는 학교 수에서 상대적으로 크게 나타났으나, 전체평균 방식과 외층평균 방식에는 커다란 차이를 보이지 않았다. 군집 방법에서는 학교 군집 표집 시, 학교당 1학급 표집 또는 학교 당 여러 학급 표집방법에 비해 학교 수는 현저하게 작게 표집되지만, 학생 수는 학교의 전체학생을 표집 대상으로 하기 때문에 다른 두 방법에 비해 매우 크게 나타났다.

3. 준거지표

이 연구에서는 학업성취도 평가 전수 학생을 모집단으로 하여 산출된 국가수준에서의 성취 수준 비율, 학교수준에서의 성취수준 비율과 ICC를 시뮬레이션을 통해 표집 방법과 표집 규

모에 따라 반복 추출된 표집 자료에서의 결과를 비교하기 위하여 다음의 준거지표를 적용하여 대표성과 정확성을 검증하였다.

$$BIAS = (\bar{\hat{\tau}} - \tau), \quad SEE = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\tau}_r - \bar{\hat{\tau}})^2}, \quad RMSE = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\tau}_r - \tau)^2}$$

R : 반복 횟수, τ : 모집단의 결과(성취수준 비율, ICC), $\hat{\tau}$: 표집에서의 결과

BIAS는 표집 결과를 통한 추정 결과의 정확성을 보여주는 지표이며, SEE는 표집 결과를 통한 추정 결과의 안정성을 보여주는 지표이다. RMSE는 이 두 가지 지표를 종합적으로 고려한 것으로 $RMSE^2 = BIAS^2 + SEE^2$ 의 관계를 가지고 있다. 이 연구에서는 표집 방법에 따라 표집학생들을 통한 학교별 성취수준 비율의 정확성을 탐색하고자 하였으며, 이에 따라 모집단에서의 학교별 성취수준 비율과 표집 방법에 따른 학교별 성취비율의 추정의 정확성을 보고자 하였다. 이를 위해 다음과 같은 지표를 통해 학교별 정확성과 안정성을 종합적으로 확인하였다.

$$RMMSE = \sqrt{\frac{1}{R} \sum_{r=1}^R \frac{1}{N_{s_r}} \sum_{r_i}^{r_f} (\hat{\tau}_{r_i} - \tau_i)^2}, \quad R : \text{반복 횟수}, N_{s_r} : r \text{ 번째 반복에서의 학교 수}, r_i :$$

r 번째 반복의 i 학교, τ_i : i 학교의 모집단 결과(학교의 성취수준 비율), $\hat{\tau}_{r_i}$: r 번째 반복의 i 학교 결과

RMMSE는 학교별 정확성과 안정성을 종합적으로 고려한 지표로 R 은 반복 표집 횟수를 의미하며, N_{s_r} 는 r 번째 반복에서의 학교 수를 의미한다. r_i 는 r 번째 반복에서의 i 번째 학교를 의미한다. 즉, RMMSE는 모집단의 i 학교에서의 결과와 i 학교가 표집되었을 때의 산출되는 결과의 전반적인 평균 차이를 보여주고 있다.

IV. 연구 결과

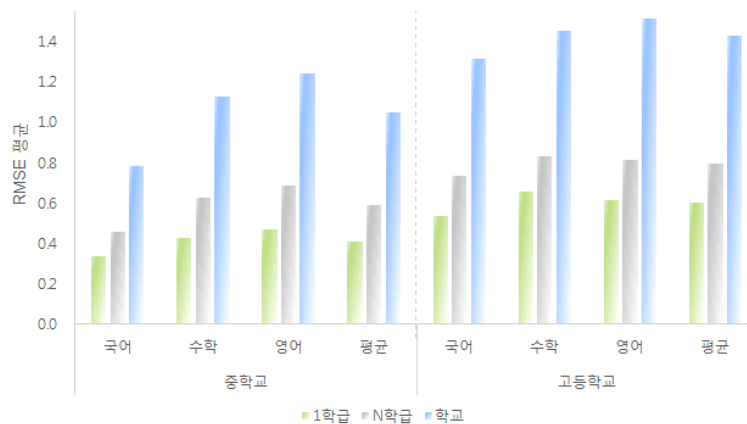
1. 군집방법에 따른 추정의 정확성

시뮬레이션을 통해 군집 방법, 즉 1개 학교의 1개 학급, 1개 학교의 여러 개 학급, 1개 학교의 모든 학급의 학생 표집한 학생을 통한 추정치의 정확성은 <표 9>, [그림 1]과 같다. RMSE는 추정의 정확성 지표인 BIAS(편의성)와 추정의 안정성 지표인 SEE(추정치 표준편

차)가 결합된 종합적인 지표이므로, RMSE를 중심으로 결과표를 제시하였다. 이에 <표 9>에 서는 모집단 분석 결과 대비 표집 조건별 추정치에 대한 RMSE(평균제곱근 오차)값을 제시하였다.

<표 9> 군집 방법에 따른 과목별 평가항목별 추정의 정확성

항목	군집	중학교				고등학교			
		국어	수학	영어	평균	국어	수학	영어	평균
성취수준 비율의 RMSE 평균	1학급	0.3352	0.4260	0.4655	0.4089	0.5348	0.6517	0.6108	0.5991
	N학급	0.4563	0.6258	0.6818	0.5880	0.7311	0.8268	0.8127	0.7902
	학교	0.7809	1.1229	1.2359	1.0466	1.3127	1.4493	1.5100	1.4240
ICC의 RMSE 평균	1학급	0.0070	0.0187	0.0086	0.0114	0.0249	0.0448	0.0201	0.0299
	N학급	0.0132	0.0274	0.0178	0.0194	0.0368	0.1084	0.0400	0.0617
	학교	0.0198	0.0346	0.0274	0.0273	0.0564	0.1676	0.0627	0.0956
학교별 성취수준 비율의 RMMSE 평균	1학급	5.3952	5.7184	5.8447	5.6528	8.2627	11.1339	8.4778	9.2915
	N학급	3.1293	3.2986	3.3758	3.2679	4.9628	6.7726	5.0932	5.6095



[그림 1] 군집 방법에 따른 성취수준 비율의 RMSE 평균

군집 방법에 따른 RMSE 평균값을 비교한 결과, 과목이나 학교급에 관계없이 학교 내 1학급 표집 방식(현행)이 다른 방법에 비해 국가수준에서의 성취수준 비율과 ICC의 RMSE가 가장 작았으나, 학교단위 군집 표집에서 RMSE 평균값이 가장 크게 나타났다. 이러한 결과는 학교 내 1개 학급 표집 방식이 국가수준의 성취수준 비율을 산출하는 데는 적합하나, 개별 학교별로 성취수준 비율을 추정 할 경우 추정의 정확성이 낮은 것을 보여준다. 또한, 전반적으로 RMSE 평균값이 중학교에 비해 고등학교에서 상대적으로 크게 나타났는데, 이는 학교 간 차이의 영향으로 중학교 결과의 정확성이 높은 것을 시사하고 있다. 학교별 성취수준 비율의 RMMSE 경우, 1학급 표집보다는 N학급 표집에서 RMSE가 상대적으로 작게 나타나, 학교 단

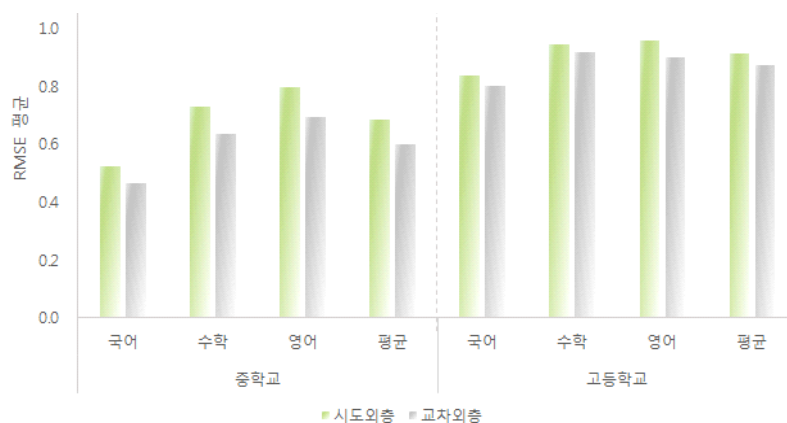
위의 성취수준 비율의 추정을 위해서는 학교 내에서 여러 학급을 표집하는 것이 유리한 것을 볼 수 있다.

2. 외층 설정에 따른 추정의 정확성

학업성취도 평가의 표집 방식에서는 외층으로 시·도교육청을 활용하여 시·도교육청별로 학생을 표집하였다. 이 연구에서는 시·도교육청 내에 있는 지역규모를 추가적으로 고려하기 위해 시·도교육청과 지역규모의 교차 외층을 통해 정확성을 점검하고자 하였다. 시·도교육청과 시·도교육청 및 지역규모의 교차 외층 설정별 표집학생을 통한 추정치의 정확성은 <표 10>, [그림 2]와 같다.

<표 10> 외층 설정에 따른 과목별 평가항목별 추정의 정확성

항목	외층	중학교				고등학교			
		국어	수학	영어	평균	국어	수학	영어	평균
성취수준 비율의 RMSE 평균	시도외층	0.5242	0.7283	0.7963	0.6829	0.8357	0.9461	0.9558	0.9125
	교차외층	0.4655	0.6331	0.6946	0.5977	0.8000	0.9157	0.8977	0.8711
ICC의 RMSE 평균	시도외층	0.0127	0.0269	0.0175	0.0190	0.0389	0.1028	0.0406	0.0608
	교차외층	0.0125	0.0252	0.0163	0.0180	0.0353	0.0947	0.0357	0.0552
학교별 성취수준 비율의 RMMSE 평균	시도외층	4.2674	4.5118	4.6045	4.4613	6.6016	8.9700	6.7764	7.4494
	교차외층	4.2599	4.5010	4.6010	4.4540	6.6171	8.9467	6.7845	7.4494



[그림 2] 외층 설정에 따른 성취수준 비율의 RMSE 평균

시·도 교육청 외층 표집 방식(현행)과 시·도 교육청과 지역규모를 반영한 교차외층 방식을 통한 분석 결과, 성취수준 비율 및 ICC 각각에 대하여 교과 및 학교급에 상관없이 교차외층

방식에서 RMSE 평균값이 작게 나타났으며, 전반적으로 RMSE 평균값이 중학교에 비해 고등학교에서 크게 나타났다. 반면, 학교별 성취수준 비율의 RMMSE 경우, 두 방식 간 커다란 차이를 보이지 않았다. 이러한 결과는 국가수준에서의 추정치를 활용하기 위해서 현행 표집 방법에서 시·도교육청과 지역규모의 교차 외층의 설정을 통해 추정치의 정확성을 보다 높일 수 있는 방안으로 고려될 수 있는 가능성을 보여주고 있다.

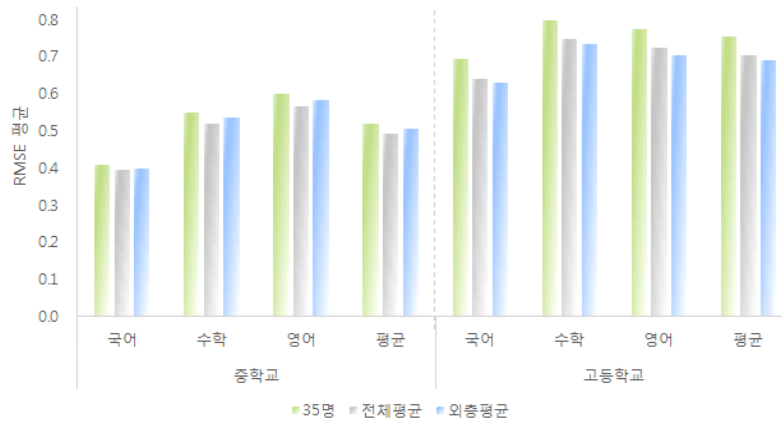
3. 학급당 기준 학생 수 산정

학업성취도 평가에서는 표집학교 선정에 위해 학급 당 기준 학생 수를 35명으로 설정하고 있다. 이 연구에서는 학급 당 기준 학생 수 현실화와 학교의 대표성 제고를 위해 학급 당 학생 수의 전체 평균, 외층별 평균으로 기준 학생 수를 설정하여 표집한 후 결과를 비교하였으며, 정확성 결과는 <표 11>, [그림 3]과 같다.

<표 11> 학급당 기준 학생 수 산정에 따른 과목별 평가항목별 추정의 정확성

항목	기준학급 크기	중학교				고등학교			
		국어	수학	영어	평균	국어	수학	영어	평균
성취수준 비율의 RMSE 평균	35명	0.4070	0.5481	0.5985	0.5178	0.6928	0.7957	0.7735	0.7540
	전체평균	0.3936	0.5178	0.5661	0.4925	0.6383	0.7454	0.7206	0.7014
	외층평균	0.3979	0.5340	0.5812	0.5044	0.6276	0.7330	0.7029	0.6878
ICC의 RMSE 평균	35명	0.0103	0.0235	0.0139	0.0159	0.0301	0.0741	0.0300	0.0447
	전체평균	0.0100	0.0231	0.0130	0.0153	0.0311	0.0779	0.0301	0.0464
	외층평균	0.0102	0.0230	0.0134	0.0155	0.0307	0.0753	0.0300	0.0453
학교별 성취수준 비율의 RMMSE 평균	35명	4.2664	4.5022	4.5877	4.4521	6.6026	8.9685	6.7704	7.4471
	전체평균	4.2611	4.5092	4.6093	4.4599	6.6097	8.9610	6.7734	7.4480
	외층평균	4.2635	4.5078	4.6112	4.4608	6.6158	8.9456	6.7977	7.4530

<표 11>과 [그림 3]의 결과를 보면, 학급당 기준 학생 수 산정 방식 간 RMSE 평균값의 차이가 거의 없음을 확인할 수 있다. 다만 매우 근소한 차이지만 중학교에서는 전체 평균 방식의 RMSE가 가장 작았고, 고등학교에서는 반대로 외층 평균 방식이 가장 작게 나타났다. 또한, 학교별 성취수준 비율의 RMMSE 또한, 세 가지 방식 간 커다란 차이를 보이지 않은 것을 확인할 수 있다.



[그림 3] 학급당 기준 학생 수 산정에 따른 성취수준 비율의 RMSE 평균

<표 11>과 [그림 3]에서 나타난 결과는 중학교에서는 전체 평균 방식이 미미하지만 정확성을 높이는데 기여함을 알 수 있고, 고등학교에서는 외층에 따라 학급 당 학생 수가 차이로 인해 외층별 평균을 활용하는 방식이 정확성을 높이는데 기여하는 것을 알 수 있다.

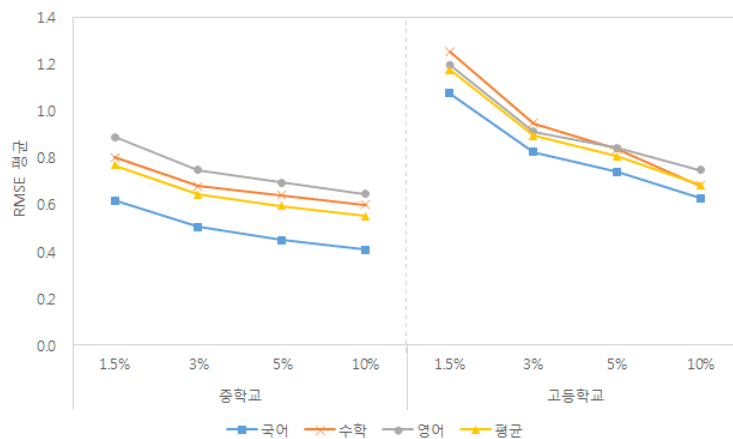
4. 표집 규모에 따른 추정의 정확성

현행 학업성취도 평가에서는 추이 분석의 안정성을 위해 각 검사유형별로 2,000명을 표집하고 있으며, 총 8,000명을 표집하고 있다. 이는 모집단의 약 1.5%에 해당하는 비율로 이 연구에서는 국가수준의 추정치의 정확성을 높이기 위해 표집 규모의 설정에 대한 기초자료를 확보하기 위해 표집 규모를 점차 늘려가며 정확성을 확인하고자 하였다. 표집 규모에 따른 추정치의 정확성은 <표 12>와 [그림 4]에 제시하였다.

<표 12> 표집 규모에 따른 과목별 평가항목별 추정의 정확성

항목	표집 규모	중학교				고등학교			
		국어	수학	영어	평균	국어	수학	영어	평균
성취수준 비율의 RMSE 평균	1.5%	0.6177	0.8034	0.8905	0.7705	1.0774	1.2561	1.2003	1.1779
	3%	0.5042	0.6791	0.7489	0.6441	0.8245	0.9474	0.9138	0.8952
	5%	0.4484	0.6402	0.6950	0.5946	0.7414	0.8390	0.8445	0.8083
	10%	0.4089	0.6001	0.6473	0.5521	0.6280	0.6810	0.7485	0.6858
ICC의 RMSE 평균	1.5%	0.0132	0.0256	0.0183	0.0190	0.0349	0.0967	0.0355	0.0557
	3%	0.0120	0.0260	0.0169	0.0183	0.0339	0.1013	0.0368	0.0573
	5%	0.0123	0.0264	0.0164	0.0184	0.0350	0.0992	0.0362	0.0568
	10%	0.0128	0.0263	0.0160	0.0184	0.0445	0.0978	0.0443	0.0622

항목	표집 규모	중학교				고등학교			
		국어	수학	영어	평균	국어	수학	영어	평균
학교별 성취수준 비율의 RMMSE 평균	1.5%	5.4069	5.7128	5.8273	5.6490	8.2517	11.1618	8.4602	9.2912
	3%	4.4224	4.6733	4.7780	4.6246	6.8217	9.2681	7.0017	7.6972
	5%	3.9564	4.1863	4.2722	4.1383	6.1653	8.4004	6.3297	6.9652
	10%	3.2690	3.4532	3.5335	3.4185	5.1986	7.0030	5.3302	5.8440



[그림 4] 표집 규모에 따른 성취수준 비율의 RMSE 평균

국가수준의 성취수준 비율을 기준으로 표집 규모가 증가할수록 RMSE 평균값이 점차 감소하고, 그 감소폭은 표집규모가 1.5%(현행)에서 3% 증가 시 가장 크게 나타나는 것을 볼 수 있다. 또한, 중학교에 비해 고등학교에서 RMSE 평균값이 크게 나타났으며, 과목 간 차이의 경우 중학교에서는 RMSE 평균값이 국어가 가장 낮고 영어가 가장 높았다. 고등학교에는 국어가 가장 낮고, 수학과 영어에서의 RMSE 평균값이 비교적 유사하였다. 학교별 성취수준 비율 기준의 RMMSE 비교에서도 유사한 결과를 보이고 있었다. 이러한 결과는 표집 규모가 증가할수록 일반적으로 모수에 대한 추정의 정확성이 높아지는데, 그 정도는 대체로 1.5%에서 3%로 증가할 경우 크게 증가하는 것을 의미하고 있다.

IV. 결론 및 제언

이 연구에서는 학업성취도 평가가 전수체제에서 표집체제로 전환되면서, 국가수준 및 학교 수준에서의 통계치에 대한 모수 추정의 정확성을 제고하기 위한 방안을 도출하고자 하였다.

구체적으로 군집 방법, 외층 설정, 학급 당 학생 수를 의미하는 기준 학급 수 설정, 표집 규모에 따라 모집단에서 학생을 반복 표집하였으며, 각 표집 방법에 따라 산출되는 표본 통계량과 모집단의 모수 간 차이를 통해 정확성을 점검하였다. 주요 연구 결과를 요약하면 다음과 같다.

첫째, 군집 방법 설정의 경우 학교 당 1개 학급 표집, 학교 당 N개 학급 표집, 학교 군집 표집 가운데 학교 군집 방안에서 국가수준에서의 성취수준 비율 산출 시 가장 큰 오차가 나타났으며, 학교 당 1개 학급 표집 방안에서 오차가 가장 작게 나타났다.

둘째, 표집 설계에서의 기준 학급 크기와 외층 설정 방안의 경우, 기준 학급의 설정을 기존 35명, 전체평균, 외층평균으로 구분하여 정확성을 분석한 결과, 기존의 35명 기준 학급 시 가장 오차가 크게 나타났으며, 학교급에 따라 약간 차이의 차이는 있으나, 전체평균(약 30명)과 외층평균 방법 간 결과가 유사하였다. 이에 기준 학급을 우리나라 전체 학급 당 학생 수를 반영하는 전체 평균을 활용하는 것이 적절할 수 있음을 확인하였다. 다만, 이 경우 기존의 35명인 경우와 비교할 때 학교 수가 늘어나는 효과가 있어 시행 상의 어려움을 감안해야 할 것이다. 이런 경우 외층(시·도교육청)별 1학급 당 학생 수의 평균으로 기준 학급을 제시하는 것도 하나의 방안이 될 수 있을 것이다. 두 가지 외층 설정 방법을 비교한 결과 시·도 교육청과 지역규모를 교차로 하여 외층을 설정하는 방법에서 보다 추정치의 정확성이 높아졌다. 이에 따라 추후 표집 설계에서 시·도교육청과 지역규모를 교차하는 방식으로 외층을 설정하는 것도 고려할 필요가 있다.

셋째, 기존 학업성취도 평가의 표집규모인 약 1.5% 표집 방법과 더불어 3%, 5%, 10% 표집 방안에 따른 표본통계의 정확성을 비교한 결과, 예상대로 표집 규모가 증가할수록 정확성이 높아졌다. 표집 규모 증가에 따른 정확성의 향상 정도를 분석한 결과, 1.5%에서 3%로 증가할 경우 정확성의 정도가 크게 높아지며, 3% 이후 표집 규모가 증가할수록 정확성이 높아지나, 향상의 정도가 1.5%→3% 증가 시 크기보다 작게 나타났다.

이상의 연구 결과를 종합하여 시사점을 도출하면 다음과 같다.

첫째, 평가 결과가 국가수준에서의 학력 수준 점검으로 제한될 경우, 표집 규모의 큰 변동 없이 표집 방법을 통해 정확성을 높이기 위한 방안으로 기준 학급 크기와 외층의 구분을 보다 현실화하는 것을 고려할 수 있다. 연구 결과 기준 학급 크기에 있어 기존 35명으로 설정할 경우보다 전체 평균, 외층별 평균으로 설정할 경우 표본 통계치의 정확성이 높아지며, 교차외층을 활용할 경우 표본 통계의 정확성이 시·도교육청 외층을 활용할 경우보다 높게 나타났다. 또한, 군집 방법에서 표집 학교의 모든 학생을 표집하는 학교 군집 방법의 경우 학생의 표집 규모는 크게 늘어나지만, 국가수준 통계의 정확성은 크게 증가하지 않았다. 이는 표집을 통한 통계치의 정확성이 담보되기 위해서는 현재와 같이 학교 수가 충분히 요구되어야 함을 의미하며, 외층에 따라 분포되어 있는 학교의 대표성을 확보할 필요가 있음을 보여준다. 따라서 표집 규모의 큰 변동없이, 즉 학생 수의 큰 변동없이 표본 통계의 정확성을 높이기 위해서는

먼저 우리나라 학교의 대표성 확보를 위해 표집 설계에서 시·도교육청과 지역규모를 교차하여 외층으로 설정하고, 기준 학급의 크기를 전체 학급 당 학생 수의 평균 또는 외층별 학급 당 학생 수의 평균으로 현실화하는 방안이 고려될 필요가 있다.

둘째, 실제 학업성취도 평가에 적용하기 위해서는 시행의 효율성, 표집을 통한 통계의 정확성을 함께 고려하여 표집 설계를 개선해야 할 필요가 있다. 연구 결과에 따르면 일반적으로 알려진 바와 같이 표집 규모가 증가할수록 정확성은 높아지는 결과를 보이고 있다. 또한, 학교 당 1개 학급 표집 방안이 적정함을 보이고 있다. 그러나 연구 결과와 같이 실제로 적용할 때 표집 규모를 늘리며, 학교 당 1개 학급을 표집할 경우 매우 많은 수의 학교를 표집하여 시행해야 하므로 시행 측면에서 부담이 있어, 현실적으로 표집 규모를 일방적으로 늘릴 수는 없는 실정이다. 이러한 시행 상의 측면을 고려할 때 시행 상의 효율성과 학교 현장의 평가 부담을 고려하고 정확성을 제고하기 위한 방안으로 연구 결과에 따라 정확성이 크게 증가하는 표집 규모인 3%를 표집하되 학교 수는 현행 1.5%에서의 학교 수로 유지하며, 표집 학생을 늘리는 방안으로 1개 학교에서 2개 학급을 표집하는 방안을 고민할 필요가 있다.

셋째, 학업성취도 평가의 2015 개정 교육과정의 개선 방안으로 표집으로 인한 오차의 허용 수준을 어느 정도로 할 것인지에 대한 논의가 필요하다. 2015 개정 교육과정의 적용에 따라 교육과정의 질 점검 기능이 강화되고 최근 역량중심의 평가 기조를 고려하여 후속 연구로 학생의 평가 부담을 줄이고 안정적인 점수를 산출하는 방안과 모집단 결과를 정확히 반영할 수 있는 표집가중치 산출 및 모수 추정 오차에 대한 연구가 수반되어야 할 필요가 있다.

끝으로 동일 표집 설계 조건 하에서 중학교 학교급에 비해 고등학교에서 추정의 정확성이 전반적으로 낮은 것을 확인할 수 있었다. 따라서 후속연구에서는 외층 변인을 보다 정교화 하는 방안을 탐색하고 추정의 정확성이 향상되는지를 살펴볼 필요가 있으며, 특히 중학교와는 달리 고등학교의 경우 학교목적 설립유형에 따라 학생 특성이 차이가 큰 점을 고려할 때 표집 설계 단계에서부터 학교목적유형을 포함하는 방안을 모색할 필요가 있다.

참 고 문 헌

- 구자옥, 김성숙, 임해미, 박혜영, 한정아(2015). **OECD 국제 학업성취도 평가 연구: PISA 2015 본검사 시행 보고서**. 한국교육과정평가원. 연구보고 RRE 2015-6-2.
- 교육부(2017. 06. 14). **국가수준 학업성취도 평가 시·도 교육청별 자율 시행**. 교육부 보도자료.
- 자료출처(2019. 1.10):
<https://www.moe.go.kr/boardCnts/view.do?boardID=294&lev=0&statusYN=C&s=moe&m=0204&opType=N&boardSeq=71431>
- 김명숙, 노국향, 박정, 부재율, 양길석, 이해영(1998). **국가수준 교육성취도 평가 방안 연구**. 한국교육과정평가원 연구보고 RRE 1998-8.
- 김수진, 김미영, 박지현, 전경희, 김민정, 서지희(2014). **수학·과학 성취도 추이변화 국제비교 연구: TIMSS 2015 예비검사 시행 및 결과 분석**. 한국교육과정평가원. 연구보고 RRE 2014-3-1.
- 김수진, 이재봉, 박지현, 이문수(2015). **수학·과학 성취도 추이변화 국제비교 연구: TIMSS 2015 본검사 시행**. 한국교육과정평가원. 연구보고 RRE 2015-11-2.
- 박민규, 강현철(2016). **SAS를 활용한 표본설계와 통계적 추론**. 자유아카데미.
- IEA. (2016). *Method and procedures in TIMSS 2015*. Retrieved from <https://timssandpirls.bc.edu/publications/timss/2015-methods.html> (2018. 1. 17. 검색).
- Lewis, T. H. (2017). *Complex survey data analysis with SAS[®]*. CRC Press: Taylor & Francis Group.
- Lohr, S. L. (2010). *Sampling: Design and analysis(2nd ed.)*. Boston, MA: Brooks/Cole.
- NCES. (2019a). “NAEP Assessment Sample Design”. Retrieved from https://nces.ed.gov/nationsreportcard/tdw/sample_design/default.aspx (2019. 1. 29. 검색)
- NCES. (2019b). “School Sample Selection for the 2013 State Assessment”. Retrieved from https://nces.ed.gov/nationsreportcard/tdw/sample_design/2013/school_sample_selection_for_the_2013_state_assessment.aspx (2019. 1. 29. 검색)

· 논문접수 : 2019.1.5. / 수정본접수 : 2019.2.1. / 게재승인 : 2019.2.18.

ABSTRACT

A Study on Improvement of Sampling Design of the National Assessment of Educational Achievement

In-Yong Park

Associate Research Fellow, Korea Institute for Curriculum and Evaluation

Hye-Kyung Jung

Associate Research Fellow, Korea Institute for Curriculum and Evaluation

Wan-Soo Kim

Research Fellow, Korea Institute for Curriculum and Evaluation

The purpose of the present study is to improve the sampling design of the National Assessment of Educational Achievement (NAEA) to enhance the representativeness of the sampled data and the accuracy of statistical estimates. Based on the census data of 2016 NAEA, we conducted a simulation study by varying sampling methods, sample size, explicit stratification variable, and the expected class size, and compared the results from sampled data with the census results. The study found the appropriateness of stratified sampling relative to cluster sampling and efficiency of 3% sampling, and suggests revision of the expected class size reflecting the real educational settings, and inclusion of area size as an explicit stratification to improve the representation of the sampled data. The study implications can be used for improving sampling design in large scale assessments.

Key Words: NAEA, Sampling design, Stratified sampling, Sample size