

교육 데이터 활용 활성화 방안 연구1)

최인봉 (한국교육과정평가원 연구위원)*
 시기자 (한국교육과정평가원 선임연구위원)
 박상복 (한국교육과정평가원 부연구위원)

요약

본 연구는 교육 분야의 데이터 관련 전문가와 데이터플랫폼 운영 담당자를 대상으로 실시한 설문조사 결과에 근거하여 교육 데이터 활용 활성화 방안을 제안하는 데에 그 목적이 있다. 교육 분야 데이터 관련 전문가 설문은 교육 데이터 활용 현황 및 개선 방안, 교육 데이터플랫폼 활용 현황 및 활성화 방안과 관련된 내용으로 구성하였으며, 데이터플랫폼 운영 담당자 설문은 교육 데이터플랫폼 운영 현황 및 개선 방안과 관련된 내용으로 구성하였다. 설문 분석 결과, 전문가들은 데이터 활용의 장애 요인으로 개인정보보호로 인한 데이터 제공 범위의 한계, 사용자가 필요로 하는 데이터 소재 파악의 어려움, 데이터 요청 과정의 복잡함, 데이터 분석에 필요한 정보 부족, 데이터를 활용한 연구 공모의 제한적 운영으로 인한 데이터 활용의 한계를 지적하였으며, 교육 빅데이터플랫폼 운영의 장애 요인으로는 시스템 관리·고도화를 위한 인력 및 예산 부족으로 인한 최신 데이터 업데이트 및 사용자 편의성 제공의 한계 등을 제시하였다. 데이터 활용 활성화 방안으로는 데이터 수집 및 활용의 실효성을 확보하기 위한 법적·제도적 기반 구축, 데이터의 수집·분석·제공의 효율성을 제고하기 위한 통합플랫폼 구축·운영과 데이터 역량 강화를 위한 지원 등 인적·물적 인프라 확대의 필요성을 제안하였다. 이상의 연구 결과를 토대로 법적·제도적 기반 구축과 물적·인적 인프라 구축을 위한 실행 방안을 제안하였다.

주제어: 빅데이터, 교육 데이터플랫폼, 데이터기반행정

1) 본 논문은 '교육 빅데이터 구축 및 활용 방안 - 국가수준 학업성취도 평가 자료를 중심으로 -(한국교육과정평가원 연구보고 RRE 2021-3)' 내용 중 일부를 발췌하여 수정·보완하였음.

* 제1저자 및 교신저자, ibchoi@kice.re.kr

I. 서 론

4차 산업혁명 시대로의 변화된 환경 속에서 폭발적으로 증가하는 데이터를 활용하여 얼마나 많은 가치를 창출해 내느냐가 더욱 중요해지고 있으며, 세계 주요국은 이러한 빅데이터의 직·간접적인 효과에 주목하고 빅데이터 활용을 통한 비용 절감, 공공부문 서비스 향상, 위기 상황 대비, 맞춤형 서비스 제공 등을 목표로 양질의 빅데이터를 확보하기 위해 정부 주도로 빅데이터 활성화 정책을 추진하고 있다. 우리나라에서도 4차 산업혁명 시대 도래에 따른 데이터 생태계 조성 및 데이터를 활용한 행정의 과학화 추진을 강조한 바 있으며, 교육부는 이러한 데이터 환경 변화에 적극적으로 대응하기 위해 '데이터 기반 교육정책 및 추진'을 핵심 사업으로 포함하여 '데이터를 활용한 교육행정의 과학화'를 추진하고 있다.

국내 빅데이터 정책은 2012년에 발표한 '스마트 국가 구현을 위한 빅데이터 마스터 플랜'을 시작으로, 공공데이터 혁신전략(2018.2), 데이터 산업 활성화 방안(2018.6), 데이터·AI 경제 활성화 계획(2019.1)을 거쳐 2020년에 「데이터기반행정활성화에관한법률」을 제정·공표함으로써 공공데이터 활용 활성화를 위한 기틀을 마련하였다. 이에 교육 분야에서도 기관별로 독자적으로 데이터를 수집하여 제공하는 체제를 개선하여 교육 전반에 대한 종합적이고 거시적 관점에서 데이터 활용 계획을 수립하고, 중장기적 교육목표와 교육 패러다임의 변화를 반영하여 체계적으로 교육관련 데이터를 수집·분석·제공·관리할 수 있는 체계를 구축할 필요성이 대두되었다. 이러한 상황속에서 교육부는 정부의 데이터 정책을 교육부의 주요 사업으로 포함하여 지능형 교육정보통계시스템(EDS) 개통, K-에듀 통합 플랫폼 구축 계획 등을 발표하였다.

공공 분야에서는 정부 부처의 분야별 현안 대응을 위해 공공행정 500건, 재난·안전 56건, 교통·물류 45건, 보건·의료 23건 등 1,119건의 과제를 발굴·분석하였으며, 기관의 정책수립 및 의사결정에 활용한 사례는 697건이다(관계기관 합동, 2021, p.2, p.8). 이를 토대로 국내 빅데이터 활용 사례를 살펴보면, 공공 분야에서는 행정, 재난·안전, 교통, 보건·의료 등 다양한 분야에서 빅데이터에 대한 분석 결과를 정책에 반영하는 선순환적 가치를 창출하고 있는 반면, 교육 분야에서는 일부 지자체를 중심으로 교육 정책 모니터링 및 교육 행정의 효율성을 제고하기 위한 목적으로 활용하고 있어 적용 범위가 매우 제한적이었다. 따라서 교육 데이터의 개방·공유·활용이 다소 미흡한 수준임을 고려할 때, 향후 교육 분야 데이터의 가치 및 활용성 증대가 중요한 선결 과제로 남아 있다고 할 수 있다.

특히 교육 분야의 경우 NEIS, 학교정보공시, 에듀파인, 교육통계, 수능, 학업성취도 평가, 국가수준 또는 시도교육청 수준의 종단연구 자료 등 초·중등 교육과 관련한 다양하고 방대한 자료가 축적되어 있으며, 그린 스마트 미래학교, K-에듀 통합플랫폼, 컴퓨터 기반 학업성취도 평가 도입 등의 정책 추진으로 교육 환경이 빠르게 변화됨에 따라 온라인 로그 데이터뿐 아니라 사물인터넷, 웨어러블 디바이스나 가상현실 테크놀로지 등을 통해 수집되는 디지털 데이터의 종류와 양도 폭발적으로 증가할 것으로 예상된다. 따라서 데이터 기반의 과학적 교육정책 수립 및 교육 행정 추진을 가속화하기 위한 기반을 마련하기 위해서는 교육 데이터 분석 연구 및 활용 경험이 많은 전문가(교육정책담당자 포함)와 기관

의 플랫폼 운영 담당자를 대상으로 교육 분야 데이터 활용 현황을 파악하고 교육 데이터의 활용을 활성화하기 위한 실질적 방안에 대한 의견을 수렴할 필요가 있다.

본 연구는 교육 분야의 데이터 분석 연구 및 활용 경험이 많은 전문가(교육정책담당자 포함)와 기관의 플랫폼 운영 담당자를 대상으로 온라인 설문 조사를 실시하여 이를 토대로 교육 데이터 활용 활성화 방안을 제안하는 데 그 목적이 있다. 교육 데이터의 체계적인 수집·분석·활용을 바탕으로 한 교육정책 구현의 필요성에 대한 사회적 요구가 증대되고 있음을 고려할 때 본 연구의 결과는 교육 빅데이터의 구축 및 활용의 활성화를 추진하는 데 의미있는 시사점을 제공할 것으로 기대한다.

II. 교육 데이터 활용 방안에 대한 설문 조사

다양한 교육 데이터 활용 현황을 파악하여 교육 데이터 활용의 활성화를 위한 개선 방안을 모색하기 위해 교육 데이터 분석 연구 및 활용 경험이 많은 전문가(교수 및 교육 연구기관 연구자)와 정책 마련을 위해 활용하고 있는 교육부 및 시도교육청 정책 담당자를 대상으로 설문 조사를 실시하였고 그 결과를 제시하였다. 또한 교육 데이터의 활용을 위해서는 이를 제공하는 교육 데이터플랫폼의 역할이 중요하기 때문에 기관의 교육 데이터플랫폼 담당자를 대상으로 운영 현황 및 활성화 방안에 대한 조사를 실시하여 그 결과를 제시하였다. 설문 조사는 온라인으로 실시하였으며 각 설문 조사 대상은 다음과 같다.

〈표 1〉 설문 조사 대상

| 설문명 | 교육 데이터 활용 방안 | | | 교육 데이터플랫폼 운영 | 합계 |
|--------|--------------|---------|-------------|---------------|----|
| | 교육 데이터 전문가 | | | 교육 데이터플랫폼 담당자 | |
| 구분 | 교수 | 교육 유관기관 | 교육부 및 시도교육청 | | |
| 대상자(명) | 25 | 13 | 10 | 12 | 60 |

이상과 같이 교육 데이터의 활용자와 제공자 양측의 인식을 조사하여 실질적인 활성화 방안 마련을 위한 시사점을 도출하고자 하였다.

1. 교육 데이터 활용 경험 및 개선 방안

전문가의 교육 데이터 활용 현황을 파악하고 실제 활용 경험에 기초한 개선 요구사항을 수렴하기 위해 교육 데이터 활용 방안에 관한 설문 조사를 실시하였다. 구체적인 설문의 구성은 다음과 같다.

〈표 2〉 교육 데이터 활용 방안에 관한 설문 구성

| 영역 | 하위영역 | 설문 주요내용 |
|-----------|-------------|--|
| 교육 데이터 | 활용 경험 및 만족도 | 교육 데이터 활용 경험(빈도) 교육 데이터 활용 만족도 교육 데이터 활용도 평가 |
| | 활용 활성화 | 교육 데이터 활용 저해요인 교육 데이터 활용 활성화 방안 |
| 교육 데이터플랫폼 | 활용 경험 및 만족도 | 교육 데이터플랫폼 활용 경험 교육 데이터플랫폼 활용 만족도 |
| | 활용 활성화 | 교육 데이터플랫폼 활용 저해요인 교육 데이터플랫폼 활용 활성화 방안 |

설문 조사에서는 전반적인 교육 데이터 및 교육 데이터플랫폼(정보시스템)의 활용에 관한 의견을 수렴하였으며 분석 결과는 다음과 같다.

가. 교육 데이터 활용 경험 및 만족도

교육 데이터의 제공 유형별 최근 3년간 교육 데이터 활용 빈도를 조사한 결과는 다음과 같다.

〈표 3〉 교육 데이터 활용 빈도(N=47)

| 구분 | 주1회 이상 | 월1회 이상 | 분기별 1회 이상 | 연 1회 이상 | 전혀 없음 |
|-----------------------------------|-------------|------------|--------------|--------------|--------------|
| | 빈도(%) | 빈도(%) | 빈도(%) | 빈도(%) | 빈도(%) |
| 공공기관의 정보시스템을 통해 제공된 교육 데이터 | 7 (14.9) | 3 (6.4) | 14 (29.8) | 17 (36.2) | 6 (12.8) |
| 공공기관 정보시스템 이외의 방법으로 제공된 교육 데이터 | 0 (0.0) | 1 (2.1) | 13 (27.7) | 19 (40.4) | 14 (29.8) |

공공기관의 정보시스템(NEIS, EDS/EDSS, 정보공시 웹사이트 등)을 통해 제공된 교육 데이터 활용 빈도는 ‘연 1회 이상’이라는 응답이 36.2%로 가장 많았고, ‘분기별 1회 이상’(29.8%), ‘주 1회 이상’(14.9%), ‘전혀 없음’(12.8%), ‘월 1회 이상’(6.4%) 순으로 나타났다. 공공기관 정보시스템 이외의 방법으로 제공된 교육 데이터 활용은 ‘연 1회 이상’이라는 응답이 40.4%로 가장 많았고, ‘전혀 없음’(29.8%), ‘분기별 1회 이상’(27.7%), ‘월 1회 이상’(2.1%) 순으로 나타났다. ‘주 1회 이상’이라고 응답한 경우는 없었다. 전반적으로 공공기관의 정보시스템을 통해 교육 데이터를 활용하는 경우가 상대적으로 많았으며, ‘연 1회 이상’ 활용한다는 응답이 많았다.

공공기관의 정보시스템 이외의 데이터 획득 경로로는 종단연구 관련 학술대회 참여, 연구계획서 제출을 통한 공모 신청이 주를 이루었고, 교육부 등 관계 기관과의 협동연구 수행을 위해 데이터를 제공받는 경우도 있었다.

교육 데이터 전문가가 실제 교육 데이터를 활용한 경험에 비추어 만족도를 평가한 결과는 다음과 같다.

〈표 4〉 교육 데이터 활용 만족도(N=46)

| 구분 | 매우 만족 | 만족 | 보통 | 불만족 | 매우 불만족 | 평균 |
|----------------|-------------|--------------|--------------|-------------|------------|------|
| | 빈도(%) | 빈도(%) | 빈도(%) | 빈도(%) | 빈도(%) | |
| 전반적인 만족도 | 3 (6.5) | 26 (56.5) | 15 (32.6) | 2 (4.4) | 0 (0.0) | 3.65 |
| 데이터 요청 및 제공 절차 | 5 (10.9) | 23 (50.0) | 11 (23.9) | 7 (15.2) | 0 (0.0) | 3.57 |
| 데이터 품질 및 신뢰도 | 5 (10.9) | 30 (65.2) | 8 (17.4) | 3 (6.5) | 0 (0.0) | 3.80 |
| 데이터 활용 및 분석 지원 | 4 (8.7) | 10 (21.7) | 28 (60.9) | 4 (8.7) | 0 (0.0) | 3.30 |

교육 데이터 활용에 대한 전반적인 만족도는 ‘만족’(56.5%), ‘보통’(32.6%), ‘매우 만족’(6.5%) 순으로 나타났다. 구체적으로 보면 데이터 요청 및 제공 절차에 대한 만족도는 ‘매우 만족’과 ‘만족’이 각각 10.9%, 50.0%, ‘보통’이 23.9%, ‘불만족’은 15.2%였으며, ‘매우 불만족’인 경우는 없었다. 데이터 품질 및 신뢰도에 대한 만족도는 데이터 요청 및 제공 절차에 대한 만족도와 유사한 양상을 보였다. 반면 데이터 활용 및 분석 지원에 대한 만족도는 ‘보통’이 60.9%로 가장 많았고 ‘매우 만족’과 ‘만족’은 각각 8.7%와 21.7%였다. 상대적으로 데이터의 품질에 대한 만족도가 가장 높았고, 데이터 활용과 분석 지원에 대한 만족도가 가장 낮았다.

교육정책 결정이나 교육현안 해결 측면에서 현재 교육 데이터 활용의 활성화 정도에 대한 전문가의 평가 결과는 다음과 같다.

〈표 5〉 교육 데이터 활용 활성화 정도(N=46)

| 구분 | 매우 활성화되어 있음 | 활성화되어 있음 | 보통 | 활성화되어 있지 않음 | 전혀 활성화되어 있지 않음 |
|--------------------------------------|-------------|-------------|--------------|--------------|----------------|
| | 빈도(%) | 빈도(%) | 빈도(%) | 빈도(%) | 빈도(%) |
| 교육정책 결정이나 교육현안 해결 측면에서 교육 데이터의 활용 정도 | 1 (2.2) | 9 (19.6) | 21 (45.7) | 15 (32.6) | 0 (0.0) |

전문가들은 교육 데이터 활용도가 ‘보통’(45.7%) 이거나 ‘활성화되지 않았다’(32.6%)고 응답한 경우가 많았다.

나. 교육 데이터 활용 활성화

교육 데이터의 활용이 활성화되지 못한 이유로 전문가들은 다음과 같은 사항을 지적하였다. 첫째, 교육 데이터를 획득하는 과정이 복잡하거나 접근 경로가 명확히 공개되지 않아 수행하고자 하는 연구에 필요한 데이터에 대한 접근성이 낮음을 지적하였다. 특히 데이터 요청 공문 작성과정에 많은 시간과 노력이 요구되는데 불구하고 공모 기간이 짧거나 신청방법에 대한 상세한 공지가 이루어지지 않아

어려움을 겪고 있었다. 이러한 데이터 획득 절차상의 복잡성 외에도 공모를 통해 데이터를 최종 획득하는 과정에서 데이터 관리 부서 담당자가 수시로 변경되거나 담당자 자의에 따른 데이터 요청에의 승인으로 인해 요청한 데이터를 필요한 시기에 제공받지 못하거나, 데이터 활용 계획에 차질이 생기는 경우가 있음을 지적하였다. 둘째, 제한적인 경로를 통해 교육 데이터를 제공받은 경우에도 개인정보 보호법으로 인해 실제 연구자에게 공개되는 데이터의 종류나 내용의 공개범위가 제한적이어서 필요한 연구 결과를 도출해 내는 데 제대로 활용하지 못하는 경우를 언급하였다. 이뿐 아니라 제공된 데이터에 최신자료가 업데이트되어 있지 않거나 분류나 기록상의 오류로 인해 신뢰도가 떨어지는 경우가 발생하여 연구자가 계획한 연구 목적에 맞게 데이터를 활용하는 데 제한이 있었다. 셋째, 데이터가 원자료 형태로 제공되지 않아 직접 조작 및 활용이 어려우며 데이터 수집 과정에 대한 구체적인 정보가 제공되지 않거나 데이터 성격에 맞는 분석 방법에 대한 안내가 되어있지 않아 필요한 데이터를 직접 활용하는 과정에서 어려움을 겪을 수 있음을 지적하였다.

이상과 같이 전문가들은 교육 데이터 활용의 장애요인으로 기관별 데이터 산재, 연계 자료 활용 불가능, 데이터 요청 과정의 복잡성, 데이터 활용 관련 안내 및 정보 부족, 제한적인 데이터만 제공되는 점 등을 들었다. 이러한 문제를 해결하기 위한 교육 데이터 활용 활성화를 위한 방안별 중요도에 대한 전문가의 평가는 다음과 같다.

〈표 6〉 교육 데이터 활용 활성화 방안의 중요도(N=47)

| 교육 데이터 활용 활성화 방안 | 매우 중요함 | 중요함 | 보통 | 중요하지 않음 | 전혀 중요하지 않음 | 평균 |
|----------------------------------|--------------|--------------|--------------|------------|------------|------|
| | 빈도(%) | 빈도(%) | 빈도(%) | 빈도(%) | 빈도(%) | |
| 데이터 분류체계 및 형식 표준화 | 26 (55.3) | 18 (38.3) | 3 (6.4) | 0 (0.0) | 0 (0.0) | 4.49 |
| 데이터 품질관리 및 신뢰도 확보 | 37 (78.7) | 10 (21.3) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 4.79 |
| 데이터 활용 홍보 강화 | 10 (21.3) | 22 (46.8) | 15 (31.9) | 0 (0.0) | 0 (0.0) | 3.89 |
| 데이터 제공을 위한 통합플랫폼 구축 | 27 (57.5) | 16 (34.0) | 3 (6.4) | 1 (2.1) | 0 (0.0) | 4.47 |
| 데이터 제공 편리성 강화 (요청 절차 간소화) | 26 (55.3) | 14 (29.8) | 5 (10.6) | 2 (4.3) | 0 (0.0) | 4.36 |
| 데이터 이용 매뉴얼 제공 | 17 (36.2) | 25 (53.2) | 5 (10.6) | 0 (0.0) | 0 (0.0) | 4.26 |
| 데이터 통합관리 및 활용 지원 서비스 전문 인력 확보 | 23 (48.9) | 18 (38.3) | 6 (12.8) | 0 (0.0) | 0 (0.0) | 4.36 |
| 데이터 활용 규제 완화 | 24 (51.1) | 13 (27.7) | 9 (19.2) | 1 (2.1) | 0 (0.0) | 4.28 |
| 개인정보 인식 및 보호 강화 | 14 (29.8) | 24 (51.1) | 8 (17.0) | 1 (2.1) | 0 (0.0) | 4.09 |

교육 데이터 활용 활성화를 위한 9가지 방안에 대한 중요도를 평가한 결과, 전반적으로 해당 방안이 중요하다고 응답한 비율이 데이터 활용 홍보 강화를 제외하고 80% 이상이였다. 즉, 제시된 방안들이 교육 데이터 활용을 활성화시키기 위해 중요함을 전문가들이 공감하고 있음을 알 수 있다. 교육 데이터 활용을 활성화하기 위해 전문가들은 데이터의 품질관리 및 신뢰도의 확보가 가장 중요하다고 응답하였으며, 데이터의 분류체계와 형식을 표준화하고, 데이터를 제공하기 위한 통합플랫폼을 구축하는 것 역시 중요한 활성화 방안이라고 인식하고 있었다. 교육 데이터 활용 활성화 방안에 대한 기타 의견으로는 데이터 통합관리를 위한 전문담당부서 조직 및 전문 인력 확보, 데이터 제공을 위한 통합플랫폼 구축을 통한 사용자의 편리성 강화, 데이터 활용 규제 완화와 개인정보 인식 및 보호 강화를 통한 교육 데이터 제공 등이 제안되었다.

다. 교육 데이터플랫폼 활용 경험 및 만족도

최근 3년간 주요 교육 데이터플랫폼(정보시스템)을 활용한 빈도를 조사한 결과는 다음과 같다.

〈표 7〉 교육 데이터플랫폼 활용 빈도(N=46)

| 구분 | 주1회 이상 | 월1회 이상 | 분기별 1회 이상 | 연 1회 이상 | 전혀 없음 |
|----------------------------------|--------------|-------------|--------------|--------------|--------------|
| | 빈도(%) | 빈도(%) | 빈도(%) | 빈도(%) | 빈도(%) |
| NEIS | 12 (26.1) | 0 (0.0) | 2 (4.4) | 5 (10.9) | 27 (58.7) |
| 에듀파인(edufine) | 11 (24.4) | 1 (2.2) | 1 (2.2) | 1 (2.2) | 31 (68.9) |
| 교육정보통계(EDS) | 7 (15.2) | 1 (2.2) | 8 (17.4) | 15 (32.6) | 15 (32.6) |
| 에듀데이터서비스(EDSS) | 1 (2.3) | 1 (2.3) | 5 (11.6) | 12 (27.9) | 24 (55.8) |
| 정보공시(학교알리미, 대학알리미, 유치원통합정보공시) | 2 (4.3) | 5 (10.6) | 7 (14.9) | 24 (51.1) | 9 (19.2) |
| 기타(통계청 국가통계포털 등) | 0 (0.0) | 0 (0.0) | 4 (25.0) | 2 (12.5) | 10 (62.5) |

주요 교육 데이터 정보시스템의 활용도를 보면 전체적으로 활용도가 낮은 것으로 나타났으나, 교육 정보통계와 정보공시(학교알리미, 대학알리미, 유치원통합정보공시)는 다른 교육플랫폼에 비하여 상대적으로 활용도가 높은 것으로 나타났다. 구체적으로 보면 NEIS 활용은 ‘전혀 없음’이라는 응답이 58.7%로 가장 많았고, 그 다음으로 ‘주 1회 이상’(26.1%), ‘연 1회 이상’(10.9%), ‘분기별 1회 이상’(4.4%) 순으로 응답하였다. 에듀파인 활용의 경우 ‘전혀 없음’(68.9%), ‘주 1회 이상’(24.4%), ‘연 1회 이상’, ‘분기별 1회 이상’, ‘월 1회 이상’은 각각 2.2% 순으로 나타났다. 교육정보통계는 ‘연 1회 이상’과 ‘전혀 없음’이 각각 32.6%로 가장 많았고, ‘분기별 1회 이상’(17.4%), ‘주 1회 이상’(15.2%), ‘월 1회 이상’(2.2%) 순으로 응답하였다. 에듀데이터서비스 활용은 ‘전혀 없음’이라는 응답이 55.8%로 가장 많

왔고, 그 다음으로 ‘연 1회 이상’(27.9%), ‘분기별 1회 이상’(11.6%), ‘월 1회 이상’과 ‘주 1회 이상’은 각각 2.3% 순으로 응답하였다. 정보공시(학교알리미, 대학알리미, 유치원통합정보공시) 활용은 ‘연 1회 이상’(51.1%), ‘전혀 없음’(19.2%), ‘분기별 1회 이상’(14.9%), ‘월 1회 이상’(10.6%), ‘주 1회 이상’(4.3%) 순으로 나타났다. 통계청 국가통계포털 등 기타의 경우 ‘전혀 없음’(62.5%), ‘분기별 1회 이상’(25.0%), ‘연 1회 이상’(12.5%) 순으로 응답하였다.

이상에 제시된 주요 교육 데이터플랫폼 이외의 대표적인 데이터 획득 경로로는 데이터 접근을 위해 안내 이메일을 통한 각종 종단연구 기관 홈페이지에서 공모를 통해 개별 요청하거나, 공개된 패널 데이터를 활용하고 있었다. 그 외에도 OECD가 제공하는 PISA, TIMSS, ICILS 등의 공개 데이터를 주로 활용하고 있었다. 중·고등학교 홈페이지 활용도도 높았는데, 학교 알리미 자료를 개별 학교를 검색하여 얻거나 직업계 고등학교 포털 하이파이브(hifive), 청소년정책연구원 통계 데이터베이스, 기타 KOISS, KERIS, 시도 유치등 교육통계 서비스를 이용하고 있었다. 주요 교육 데이터플랫폼 활용의 만족도를 조사한 결과는 다음과 같다.

〈표 8〉 교육 데이터플랫폼 활용 만족도(N=46)

| 구분 | 매우 만족 | 만족 | 보통 | 불만족 | 매우 불만족 | 전체 | 평균 |
|-------------------------------|-------------|--------------|--------------|-------------|------------|---------------|------|
| | 빈도(%) | 빈도(%) | 빈도(%) | 빈도(%) | 빈도(%) | | |
| 나이스 | 1 (4.8) | 5 (23.8) | 10 (47.6) | 4 (19.0) | 1 (4.8) | 21 (100.0) | 3.05 |
| 에듀파인(edufine) | 1 (6.3) | 6 (37.5) | 6 (37.5) | 2 (12.5) | 1 (6.3) | 16 (100.0) | 3.25 |
| 교육정보통계(EDS) | 2 (6.7) | 14 (46.7) | 11 (36.7) | 3 (10.0) | 0 (0.0) | 30 (100.0) | 3.50 |
| 에듀데이터서비스(EDSS) | 2 (10.5) | 4 (21.1) | 9 (47.4) | 4 (21.1) | 0 (0.0) | 19 (100.0) | 3.21 |
| 정보공시(학교알리미, 대학알리미, 유치원통합정보공시) | 3 (7.9) | 14 (36.8) | 19 (50.0) | 1 (2.6) | 1 (2.6) | 38 (100.0) | 3.45 |
| 기타(통계청 국가통계포털 등) | 1 (14.3) | 3 (42.9) | 3 (42.9) | 0 (0.0) | 0 (0.0) | 7 (100.0) | 3.71 |

교육 데이터플랫폼에 대한 만족도는 전체적으로 보통 혹은 만족으로 나타났다. NEIS의 경우 ‘보통’(47.6%), ‘만족’(23.8%), ‘불만족’(19.0%) 순으로 응답하였고, ‘매우 만족’과 ‘매우 불만족’은 각각 4.8%로 응답하였다. 에듀파인에 대한 만족도는 ‘만족’과 ‘보통’이 각각 37.5%로 나타났고, ‘불만족’은 12.5%, ‘매우 불만족’, ‘매우 만족’은 각각 6.3%로 응답하였다. 교육정보통계는 ‘만족’(46.7%), ‘보통’(36.7%), ‘불만족’(10.0%), ‘매우 만족’(6.7%) 순으로 응답하였고 ‘매우 불만족’이라고 응답한 경우는 없었다. 에듀데이터서비스는 ‘보통’이라는 응답이 47.4%로 가장 많았고, ‘만족’과 ‘불만족’이 각각 21.1%, ‘매우 만족’이 10.5% 순으로 응답하였다. 정보공시(학교알리미, 대학알리미, 유치원통합정보공시)의 경우 ‘보통’(50.0%), ‘만족’(36.8%), ‘매우 만족’(7.9%) 순으로 응답하였고, ‘불만족’과 ‘매우 불만족’은 각각 2.6%로 응답하였다. 통계청 국가통계포털 등 기타의 경우 ‘만족’과 ‘보통’이 각각 42.9%

로 나타났고, '매우 만족'에 응답한 비율은 14.3%였다.

라. 교육 데이터플랫폼 활용 활성화

교육 데이터플랫폼 활용의 활성화 방안 도출을 위해 우선, 교육 데이터플랫폼 활용을 저해하는 요인 및 활용상의 어려움을 조사하였으며, 이를 해결하기 위한 활성화 방안에 대하여 전문가의 의견을 조사하였다. 교육 데이터플랫폼 활성화의 저해요인과 개선 방안에 대한 전문가 의견은 개방형 질문을 통해 수집되었으며, 그 내용을 정리하면 다음과 같다.

〈표 9〉 교육 데이터플랫폼 활용의 저해요인 및 활성화 방안

| 교육 데이터플랫폼 활용의 저해요인 | 교육 데이터플랫폼 활용 활성화 방안 |
|---|---|
| <ul style="list-style-type: none"> - 데이터에 대한 불명확한 안내 및 부족한 홍보 - 데이터 요청 절차의 복잡성 - 최신 정보부족 및 데이터 신뢰도 문제 - 활용 가능한 데이터 및 연계 데이터 부족 - 기타 운영 시스템상 문제 | <ul style="list-style-type: none"> - 다양한 플랫폼 활용 방안 안내 - 종합적인 관리 및 통합 운영 가능한 플랫폼 개발 - 데이터 접근성 확장 - 전문인력배치를 통한 메뉴얼 상시 업로드 - 안내와 홍보 |

교육 데이터 전문가들이 교육 데이터 정보시스템 활용 과정에서 공통적으로 불편함을 느낀 부분은 데이터 획득 과정에서의 어려움, 데이터 자체에 대한 만족도 부족, 그리고 데이터의 활용의 어려움으로 구분된다. 첫째로 데이터 신청방법 및 과정에 대한 안내와 설명이 부족하고, 데이터마다 신청방법이 상이하여 산재되어 있는 교육 데이터를 수집하기 위해 정보를 검색하고 연구 목적에 적합한 데이터를 획득하는 데 어려움을 겪고 있었다. 또한, 연구공모나 학술대회를 통해 데이터를 신청한 후에 사용 승인 과정에의 오랜 시간이 소요되어 연구가 진행되는 동안 데이터 확보에 차질이 생기기도 하며, 실제 연구 목적에 맞게 요청한 자료와 다른 내용의 데이터를 제공받기도 하는 등 필요한 데이터를 활용하기 전 단계에서 겪는 어려움이 매우 크게 나타났다.

둘째로는 이처럼 복잡한 과정을 통해 획득한 데이터 자체의 신뢰도의 문제로 인해 수집한 데이터가 실제 연구 자료로 활용되기에 부적합하거나 단순 추출 기능 외(예: NEIS의 경우) 필요한 교육통계 자료 추출이 어려운 관계로 계획한 연구에 해당 데이터를 활용하는 데 불편함을 겪고 있었다. 실제 데이터를 검색하고 수집하는 과정에서도 다양한 어려움이 나타났는데 제공되는 각 데이터에 대한 상세한 설명이 제공되지 않아 필요한 데이터를 활용하지 못하게 되거나 해당 데이터의 목적과 결과를 이해하기 어렵고, 연구 취지에 맞게 선별적으로 선택하는 것이 용이하지 않았다. 그 외 시스템적인 부분에서도 불편사항이 나타났는데, 컴퓨터 소프트웨어 상의 활용이 원활하지 않은 경우와 데이터 다운로드가 불가능한 경우가 있었다. 이처럼 필요한 데이터를 선별적으로 선택하는 과정에서도 많은 자료가 체계 없이 산재되어 있거나 일괄적으로 다운로드를 제공하는 시스템이 아닌 일일이 데이터를 검색해서 수집해야 하는 등 활용 과정에서의 번거로움을 데이터플랫폼 활용에 대한 불만족의 이유로 서술하였다.

교육 데이터플랫폼 활용을 활성화하기 위해 검색기능을 강화하여 정보를 시각화하고, 데이터별로 설명을 추가하여 데이터에 대한 이해를 넓히고 손쉽게 필요한 데이터를 활용할 수 있도록 할 것을 제

안하였다. 또한, 데이터 자체의 신뢰도를 향상시키고 이용자들에게 활용 가능한 변수를 다양하게 제공함으로써 데이터 활용도를 확대할 수 있을 것이라 응답하였다. 즉, 데이터 요청 과정에서 내용에 대한 상세한 안내를 제공하고 교육 데이터를 활용한 연구결과를 홍보하며, 교육 데이터를 종합적으로 제시할 수 있는 교육 데이터 포털 구축을 통하여 통합적인 데이터를 제공하는 방안 등을 제안하였다. 또한 데이터 활용 과정에서 불편사항을 개선하기 위해 응답자들은 국내외 타 교육정보시스템을 활용한 경험을 바탕으로 다음과 같은 몇 가지 벤치마킹 요소를 제안하였다. 국내 예로는 한국청소년정책연구원 데이터 아카이브 홈페이지처럼 별도의 신청 과정을 생략하여 누구나 원하는 데이터에 쉽게 접근 가능케 함으로써 데이터 활용도 향상을 기대할 수 있다. 또한, 많은 이용자들이 공통적인 데이터 활용 시 어려움으로 언급한 개인정보보호를 위한 제한적 데이터 제공으로 인한 불편함을 해결하기 위해 보건복지부 개인정보보호위원회의 '보건의료 데이터 활용 가이드라인'을 참고할 수 있다. 이 가이드라인에서는 개인정보를 보호하되, 연구에 필요한 정보는 이용자가 선별적으로 활용할 수 있는 방안을 제시함으로써 데이터 활용도를 높이고 있다. 응답자들은 국내 뿐 아니라 해외의 데이터 활용 경험과 참고할 만한 다양한 사례를 제시하였는데 예를 들면, 미국 NELS의 아동-청소년 패널 데이터의 경우 학생 관련 교육용 빅데이터를, 타당화 검사를 거친 변수 중심으로 구성하여 해당 홈페이지에서 최신자료를 다운로드할 수 있는 방법을 제공하고 있다. 영국이나 일본 등 국가에서 현재 일원화된 통합 시스템을 통해 자체적으로 데이터를 운영 및 관리하고 있는 사례가 있으며 그 외에도, 카네기멜론대학을 비롯한 MIT, 스탠포드 등 해외 유명대학에서 데이터와 분석 도구까지 제공하는 education data repository를 운영하여 LearnShpere 프로젝트 진행을 통해 교육 데이터 활용을 적극적으로 권장함으로써 우수한 연구가 활발히 이루어지게 하는 시스템 운영사례를 벤치마킹하여 국내 교육 데이터플랫폼의 질적 개선 방안에 접목할 것을 제안하였다. 그리고 이러한 장애요인을 해결하기 위해서는 데이터 수집 및 생성기관의 협력체계를 구축하여 단계적인 교육 데이터 분류체계 및 형식 표준화를 추진하는 것이 필요하다. 이를 위해 데이터 관리기관 및 전문인력이 필요할 뿐 아니라 기관의 사용자 교육이 체계적으로 이루어져야 할 것을 제안하고 있다. 또한 데이터 접근성을 높이기 위해 시도교육청 또는 국책연구기관의 경우 학술대회를 통해 최신 수집 자료를 참여 연구자들에게 우선적으로 공개하는 것이 필요하며, 데이터를 활용하는 주요 수요자에 맞게 제공함으로써 데이터 활용도를 높이는 방안의 필요성을 제시하였다. 이 외에도 데이터 활용에 대한 부정적 인식을 개선하고 데이터를 활용하여 진행한 연구 결과를 공유함으로써 데이터 활용의 긍정적인 측면을 강조하고 관련 데이터 간의 연관성을 안내하는 상세한 설명이 제공된다면 한국교육 발전을 위해 다양한 연구가 활발히 진행될 수 있을 것이라 응답하였다.

2. 교육 데이터플랫폼 운영 경험 및 개선 방안

교육 데이터플랫폼(정보시스템) 운영 현황 및 개선 방안 모색을 위해 교육유관 기관의 교육 데이터 플랫폼을 운영하는 담당자 12명을 대상으로 '교육 데이터플랫폼 운영에 관한 인식 조사(담당자용)' 설문문을 실시하였으며, 설문문의 구성은 다음과 같다.

〈표 10〉 교육 데이터플랫폼에 관한 설문 구성

| 영역 | 설문 내용 |
|-----------------|---|
| 교육 데이터플랫폼 운영 현황 | - 데이터 제공 빈도 - 교육 데이터 제공의 활성화 정도 - 교육 데이터플랫폼 운영 상 어려움 및 요구사항 |
| 교육 데이터플랫폼 활성화 | - 교육 데이터 제공 확대 방안의 중요도 - 교육 데이터플랫폼 활성화 방안 |

설문 조사에서는 교육 데이터플랫폼 운영 현황과 활성화 방안과 관련한 의견을 수렴하였으며 분석 결과는 다음과 같다.

가. 교육 데이터플랫폼 운영 현황

교육 데이터플랫폼을 통한 데이터 제공 빈도는 다음과 같다. 응답자의 소속기관에서 관리하는 데이터를 외부 기관 또는 일반인에게 제공한 횟수에 대한 질문은 ‘주 1회 이상’이라는 응답이 50.0%로 가장 많았고, 그 다음으로 ‘월 1회 이상’(25.0%), ‘연 1회 이상’(16.7%), ‘분기별 1회 이상’(8.3%) 순으로 나타났다.

〈표 11〉 교육 데이터플랫폼을 통한 데이터 제공 빈도

| 구분 | 주1회 이상 | 월1회 이상 | 분기별 1회 이상 | 연 1회 이상 | 전혀 없음 | 전체 |
|---|-------------|-------------|------------|-------------|------------|---------------|
| | 빈도(%) | 빈도(%) | 빈도(%) | 빈도(%) | 빈도(%) | |
| 소속 기관에서 관리하는 데이터를 외부 기관 또는 일반인에게 제공한 횟수 | 6 (50.0) | 3 (25.0) | 1 (8.3) | 2 (16.7) | 0 (0.0) | 12 (100.0) |

응답자의 소속 기관에서 보유한 교육 데이터 제공의 활성화 정도를 묻는 질문에 대한 응답 결과는 다음과 같다.

〈표 12〉 교육 데이터 제공 활성화 정도에 대한 인식

| 구분 | 매우 활성화 되어 있음 | 활성화 되어 있음 | 보통 | 활성화 되어 있지 않음 | 매우 활성화 되어 있지 않음 | 전체 |
|--------------------------------|--------------|-------------|-------------|--------------|-----------------|---------------|
| | 빈도(%) | 빈도(%) | 빈도(%) | 빈도(%) | 빈도(%) | |
| 소속 기관에서 보유한 교육 데이터의 제공의 활성화 정도 | 1 (8.3) | 7 (58.3) | 3 (25.0) | 1 (8.3) | 0 (0.0) | 12 (100.0) |

‘활성화되어 있음’이라는 응답이 58.3%로 가장 많았고, 그 다음으로 ‘보통’(25.0%), ‘매우 활성화 되

어 있음'(8.3%), '활성화되어 있지 않음'(8.3%) 순으로 응답하였다.

교육 데이터플랫폼 운영상의 어려움 및 데이터 제공 관련 장애요인에 대한 개방형 질문에 대한 응답 결과를 정리하여 제시하면 다음과 같다.

〈표 13〉 교육 데이터플랫폼 운영상 어려움 및 요구사항

| 주제 | 영역 | 항목 |
|------------------------------------|-----------------------|---|
| 교육 데이터플랫폼 운영 상 어려움 및 요구사항 | 전문인력 부족 | <ul style="list-style-type: none"> - 자료검색기능 작동 모니터링 - 데이터 검색을 위한 데이터 큐레이터 - 데이터 관리를 위한 상시 모니터링 - 데이터 표준화 및 관리 - 데이터 매뉴얼 동영상화 - 융합사례 제공 |
| | 데이터 품질 및 신뢰도 문제 | <ul style="list-style-type: none"> - 최신 데이터 업데이트 - 담당자별로 다른 해석과 데이터 제공 - 원 데이터 제공 필요 - 데이터 간 이해상충 |
| | 데이터 관리를 위한 체계적 시스템 필요 | <ul style="list-style-type: none"> - 통합거버넌스 구축 필요 - 플랫폼 확장 - 전담조직 구성으로 데이터 관리 플랫폼 구축 필요 - 데이터 연계 기관 간 상설 협의체 구성 - 이용자별로 최적화된 UI/UX제공 - 데이터 정비와 표준화 - 데이터 추출 및 기본적인 통계처리 지원 서비스 |

데이터 제공 경험이 있는 경우, 데이터 제공 과정에서의 어려운 점이나 개선이 필요한 사항에 대한 의견을 묻는 개방형 질문에 응답자들은 교육 데이터를 제공하는 과정에서 공통적으로 경험하는 어려움으로 데이터 관리를 위한 전문 인력 부족과 데이터 제공 및 활용 관련 법적 근거가 모호한 점을 언급하였다. 이러한 법적 근거의 부족은 민원 대비 시 해당 담당자를 찾기 어렵거나 담당자별로 해석이 달라 이용자에게 데이터를 제공하는 과정에서 어려움을 겪고 있었다. 데이터 표준화를 위한 기준과 시스템이 필요하나, 현재 국내 여건에 적합한 방법론의 부재로 데이터 담당자가 데이터 제공 시 수작업으로 진행해야 하는 번거로움이 있으며, 데이터 개별 인용 명시에 어려움을 겪고 있다고 응답했다.

교육 데이터플랫폼 운영 시 개선이 필요한 사항에 대한 개방형 질문에서 응답자들은 다음과 같은 사항을 제안하였다. 첫째, 빅데이터에 대한 충분한 내부교육 시행을 통해 원하는 데이터를 담당자들이 컬럼 단위에서 찾을 수 있도록 목록화하는 것이 필요하다. 둘째, 데이터의 연계 수집 및 적재 주기를 명시함으로써 적시에 이용자들에게 제공하기 위해 데이터를 추출 선별, 가공하는 과정에서의 업무의 효율화가 필요하다. 셋째, 플랫폼 내의 기능 향상이 필요한데, 이를 위해 데이터 수집 및 분석 기능을 활용하여 데이터를 쉽게 추출하고 기본적인 통계처리를 지원하는 OLAP서비스를 운영할 필요성을 제안하였다. 또한, 비정형 데이터를 수집하고 분석 시 지원을 위해 빅 데이터 처리 플랫폼을 구축하고 플랫폼 확대를 통해 운영에 따른 플랫폼 운영 업무의 복잡성을 증대시킬 필요가 있음을 제안하였다. 넷째,

운영 중인 연구 데이터 리포지터리의 한계의 경우 다수의 접속 및 활용이 가능하도록 기능을 향상시키고, 데이터 다운로드 절차의 간소화를 통해 이용자들의 활용도를 높여야 할 필요가 있다. 현재 데이터 활용 건수의 파악이 어렵고 초기 효용성 파악이 용이하지 않은 한계가 있기 때문에 문헌 정보를 생산하고 reference의 지속적 확대를 통해 데이터 활용 건수를 확장할 필요가 있다. 다섯째, 이러한 플랫폼이 원활하게 운영되고 필요한 개선사항이 반영되기 위해서는 플랫폼 유지관리와 활용 지원을 위한 지속적 예산 배정이 필요하다. 데이터플랫폼 설계를 체계화하고 자료 검색 기능의 지속적 모니터링, 데이터 검색에의 효율성 향상을 위한 데이터 큐레이터의 활동, 윤리적 문제 해결을 위한 가이드라인 개발 및 홍보에 전문 인력이 적재적소에 배치되어야 한다. 여섯째, 전문 인력 배치는 무결성 테스트를 위한 프로세스를 설계하고 상시 데이터 출처와 유형에 대한 모니터링, 분류별로 제공 가능한 데이터베이스 필드명이나 메타 데이터 정보 제공을 위해 필요하다. 마지막으로 시스템의 안정성이 중요하므로 대용량 시스템을 구축하고 정보 시스템 내 저장된 데이터의 직접적인 분석환경 지원 방안을 마련할 필요성을 언급하였다. 예를 들면, 시스템 내의 빅데이터를 바로 분석할 수 있는 환경을 통해 이용자의 데이터 분석 및 활용도를 높일 수 있을 것으로 응답하였다. 그 외에도 응답자들은 다음과 같은 추가적인 고려사항을 제안하였다. 메타데이터 수집체계 구축을 통해 데이터의 위치, 속성 등 상세정보 파악을 가능케 하며, 플랫폼 운영 목적에 따른 거버넌스 체계 마련 및 운영이 필요하다. 그리고 개인정보 보호법으로 인한 데이터 제공에의 규제에 대한 대책과 법안을 체계적으로 마련하고 데이터 제공 및 플랫폼 운영을 위한 법·제도적 기반을 마련하고, 데이터 연계 기관 간 상설 협의체를 구성하여 운영할 필요성을 제안하였다.

나. 교육 데이터플랫폼 활성화 방안

교육 데이터플랫폼의 활성화 방안에 대한 응답자의 의견을 데이터 제공자와 기관 및 정부로 구분하여 요약·제시하면 다음과 같다.

〈표 14〉 교육 데이터플랫폼 활성화 방안

| 주제 | 영역 | 항목 |
|-------------------|--------------------------|--|
| 교육 데이터 플랫폼 활성화 방안 | 데이터 제공자 입장에서의 플랫폼 활성화 방안 | <ul style="list-style-type: none"> - 업무담당자의 인식 중요 - 데이터 제공 추진 체계 표준화를 위한 운영진 노력 - 데이터 활용 가이드 마련 및 배포 - 연구성과 공유 - 타 분야와 융합사례 제공 - 주기적 데이터 수요 파악 |
| | 기관 및 정부 입장에서의 활성화 방안 | <ul style="list-style-type: none"> - 데이터별 표준화 및 기관 차원의 기준 확립 - 업무 연계를 위한 분석 표준 필요 - 데이터 정비와 표준화 - 데이터 관리 플랫폼 구축 및 고도화 - 전담조직과 전문인력 확충 - 법제도적 여건 선행 |

먼저 데이터 제공자 입장에서는 이용자가 필요한 데이터를 검색하는 과정에서 데이터 제공 편리성을 강화하고 명확한 데이터 확보를 위해 데이터 형식을 표준화하여 검색어 입력을 통해 정보에 접근하게 하는 통합 시스템이 필요함을 제안하였다. 원천 데이터를 연구자들에게 제공하거나 연계 데이터를 결합하여 활용할 수 있게 하고 분석 표준을 제정하여 제공할 필요가 있으며, 데이터 간 결합을 위한 인프라 구축이 필요하다. 또한 데이터 수집에서 활용까지 가능한 통합 거버넌스 형태의 플랫폼 구축 환경 내에서 지속적인 데이터 관리와 시스템 정비를 통해, 통계성이 높은 데이터를 활용 가능한 상세 데이터로 이용자에게 제공하고 개인정보 등 민감한 정보는 세부 항목 수준에서 비식별처리를 하여 제공하는 등 다양한 활용 방안의 가능성을 제안하였다. 기관 차원의 노력으로는 다른 분야와의 융합 사례를 제공하고, 데이터 메뉴얼을 동영상화하여 제공함으로써 이용자들의 데이터 접근과 활용에 편의성을 높일 수 있으며, 데이터 접근 경로를 확대하고, 데이터 분류 및 관리 표준화를 통한 장기적인 품질 관리가 이루어져야 함을 강조하였다. 이뿐만 아니라, 데이터 관리를 위한 전담조직구성과 전문 인력 확충을 통해 주기적 데이터 수요 파악과 데이터 제공 확대를 위한 제도 정비가 이루어져야 하며, 지속적이고 안정적인 활용을 위한 예산 확보가 필요함을 지적하였다. 그 외에도 데이터 활용지원을 위한 자체 연구 및 타 연구부서와의 협업 연구 추진, 연구 성과 공유, 데이터 활용 가이드 마련 및 배포의 필요성을 강조하였다. 더불어 국가적 차원에서의 플랫폼 기능 향상과 전문 인력 배치를 위한 예산 지원 및 개인정보관련 법안을 정비할 필요가 있음을 제안하였다.

3. 시사점

교육 데이터 전문가와 교육 데이터플랫폼 운영 담당자는 데이터 활용의 장애 요인으로 개인정보보호로 인한 데이터 제공 범위의 한계, 사용자가 필요로 하는 데이터 소재 파악의 어려움, 데이터 요청 과정의 복잡함, 데이터 분석에 필요한 정보 부족, 데이터를 활용한 연구 공모의 제한적 운영으로 인한 데이터 활용의 한계 등을 지적하였다. 한편, 교육 데이터플랫폼 운영 측면에서는 최신 데이터의 업데이트와 수요자 맞춤형 데이터 제공을 위한 데이터 전문 인력 부족 및 데이터플랫폼 유지 관리 및 기능 개선을 위한 예산 부족으로 수요자 편의성을 고려하는 데 한계가 있는 것으로 나타났다. 데이터 활용 활성화를 위한 방안으로는 데이터 수집 및 활용의 실효성을 확보하기 위한 법·제도적 기반 구축, 데이터의 수집·분석·제공의 효율성을 제고하기 위한 교육 빅데이터 통합플랫폼 구축·운영, 데이터 역량 강화를 위한 지원 등 인적·물적 인프라 확대의 필요성을 강조하였다. 이를 토대로 교육 빅데이터 구축 및 활용 활성화를 위한 주요 쟁점과 과제를 제시하면 다음과 같다.

〈표 15〉 교육 데이터 활용 활성화를 위한 주요 쟁점 및 과제

| 구분 | 주요 쟁점 및 과제 |
|---------|--|
| 법·제도 관련 | <ul style="list-style-type: none"> ○ 데이터 제공 및 활용 관련 규정 및 법규 마련 - 개인정보보호 관련 규정 및 법규 - 유관기관 간 데이터 제공 협력을 위한 법적 근거 ○ 데이터 가공 및 활용 관련 연구윤리 마련 ○ 정보화 거버넌스 체계 구축 |

| 구분 | 주요 쟁점 및 과제 |
|-----------------------------|---|
| 인적·물적 인프라 관련 | <ul style="list-style-type: none"> ◦ 기관별 데이터 연계 ◦ 데이터 수집, 저장, 추출, 연계, 활용을 위한 표준 메타 데이터 체계 구축 ◦ 데이터 형태 표준화 <ul style="list-style-type: none"> - 비정형 비연속적 데이터의 처리 - 이미지 데이터 디지털화 ◦ 데이터 품질 관리 <ul style="list-style-type: none"> - 데이터 정제화 매뉴얼 제공 - 데이터 정확성, 신뢰도, 연속성 확보 ◦ 데이터 분석 관련 서비스 범위 및 절차 설정 <ul style="list-style-type: none"> - 데이터 분석 서비스 범위 및 절차 수립 - 데이터 분석 플랫폼 기획 및 설계 - 분석 프로토타입 및 예시 개발 ◦ 데이터 서비스 구조 설계 <ul style="list-style-type: none"> - 데이터 요청 절차 - 데이터 제공 기준 및 절차 - 서비스 방식에 따른 플랫폼 구조 설계 ◦ 데이터 분석 및 활용 매뉴얼 제공 및 사용자 교육 |
| 교육 빅데이터 플랫폼 구축· 운영 관련 | <ul style="list-style-type: none"> ◦ 데이터 수집 및 관리 전문 인력 확보 ◦ 주기적 데이터 수요 분석 및 서비스 개선을 위한 예산 확보 ◦ 데이터 활용 지원을 위한 연구 수행 ◦ 데이터 활용 연구 성과 공유 체계 구축 |

III. 결론 및 제언

설문 응답 결과를 토대로 분석한 결과는 다음과 같다. 먼저 전문가들은 데이터 활용의 장애 요인으로 개인정보보호로 인한 데이터 제공 범위의 한계, 사용자가 필요로 하는 데이터 소재 파악의 어려움, 데이터 요청 과정의 복잡함, 데이터 분석에 필요한 정보 부족 등을 지적하였으며, 교육 빅데이터 플랫폼 운영의 장애 요인으로는 시스템 관리·고도화를 위한 인력 및 예산 부족으로 인한 최신 데이터 제공 및 수요자의 편의를 고려한 데이터 제공의 한계 등을 지적하였다. 이를 통해 데이터 수집 및 활용의 실효성을 확보하기 위한 법적·제도적 기반 구축, 데이터의 수집·분석·제공의 효율성을 제고하기 위한 교육 빅데이터 통합플랫폼 구축·운영과 데이터 역량 강화를 위한 지원 등 인적·물적 인프라 구축 확대의 필요성을 확인하였다.

이러한 연구 결과를 바탕으로 교육 데이터 활용 활성화 방안에 대해 제안하면 다음과 같다.

첫째, 초·중등교육법 개정을 통한 교육 빅데이터 구축 환경을 마련할 필요가 있다. 「개인정보 보호법」이 개정되고 「가명 정보의 처리에 관한 특례법」이 신설되면서 학생 정보를 활용하여 데이터를 연계·활용할 수 있는 기반이 마련되었다. 그러나 「초·중등교육법」 30조의6에 의하면, NEIS의 학생 학교생활 활기록 및 건강검사기록은 관리 주체가 단위학교의 장으로서 당사자가 누구인지 식별할 수 없는 형태로 제공하는 경우에는 학생과 학생 보호자 동의 없이 학교장이 제공할 수 있도록 되어 있지만, 모든 학교에 활용 동의를 구해야 하는 현실적인 어려움이 있다. 따라서 개정된 「개인정보 보호법」과 상충되는

부분을 해소할 수 있도록 「초·중등교육법」을 개정할 필요가 있다.

둘째, 교육 분야 가명·익명정보 처리 가이드라인의 실효성을 제고할 필요가 있다. 교육부와 개인정보보호위원회는 「가명·익명정보 처리 가이드라인」(교육부, 개인정보보호위원회, 2020)을 배포하고, 법령에 따라 한국교육학술정보원을 ‘가명정보결합기관’으로 지정(한국교육학술정보원, 2021, p.3)하여 학생 개인 정보를 활용할 수 있는 근거가 마련되었다. 그러나 가명정보 결합 절차가 복잡하고 가명정보 결합키의 부재 등으로 활용에 많은 제약이 있을 것으로 예상된다. 따라서 중장기적으로 데이터 결합 및 활용의 활성화를 위해서는 데이터를 생산하는 기관에서 자체적으로 개인정보 비식별화 업무를 수행할 수 있도록 법적 근거와 조직 및 인력 지원이 필요하다.

셋째, 교육 유관기관 간 데이터를 공동으로 활용할 수 있는 법적인 근거를 마련할 필요가 있다. 교육 유관기관 간 데이터 공동 활용을 활성화하기 위해서는 행정안전부 ‘공공데이터 관리지침’(행정안전부 고시 제2019-71호, 2019. 9. 3. 개정, pp.19-22)과 같은 데이터 권리의 소재와 이용 범위에 대한 지침을 마련할 필요가 있다.

넷째, 데이터 통합관리 및 활용을 위한 추진 체계를 구축할 필요가 있다. 교육 데이터 활용 현황 및 개선 방안에 대한 설문 조사 결과에 따르면, 기관별로 데이터가 흩어져 있어 데이터 소재와 연계에 어려움이 있음을 지적하였다. 따라서 「2021년도 데이터기반 행정 활성화 시행계획」에 따른 ‘데이터 관리 및 공동활용 기반 마련’ 실현을 위한 추진 체계가 필요하다. 이를 위해 각 기관의 데이터 관련 조직이나 담당자들의 유기적 협력을 위한 공식적인 협력체를 구성하여 상호 자유롭게 활용할 수 있도록 허용하고, 외부 연구자도 해당 기관에 개별적으로 요청하지 않고 지정 기관을 통해서 협력체에 속한 모든 기관의 데이터를 요청할 수 있도록 제도를 정비하여 데이터에 대한 접근성을 높일 필요가 있다.

다섯째, 데이터 수집·연계, 활용을 위한 표준 메타데이터 체계를 구축할 필요가 있다. 본 연구의 설문 조사 결과에 따르면 교육 데이터 활성화를 위해 데이터 분류체계 및 형식 표준화가 중요하며 교육 데이터플랫폼 운영 시의 어려움 중 하나로 데이터 표준화 및 관리를 지적하였다. 따라서 교육 분야의 분류 체계는 통계청에서 고시한 ‘한국표준교육분류’를 활용하고 있으나 교육 데이터의 수집 범위는 매우 다양하기 때문에 각 기관에서 수행하고 있는 연구 사업의 특성과 이에 따라 수집되는 데이터의 특성에 대한 조사를 토대로 데이터의 통합관리를 위한 표준 분류 체계를 정비할 필요가 있다.

여섯째, 기관별 교육 데이터 수집, 분석, 제공 플랫폼 구축·운영 표준 체계 마련하고 이를 교육 데이터 통합관리시스템으로 발전시켜야 한다. 본 연구의 설문 조사 결과에 따르면 대 다수의 응답자가 통합 데이터플랫폼 구축이 중요하다고 응답하였으며, 유관기관 간 데이터 연계 제공 및 데이터 활용 모델 개발의 필요성을 강조하였다. 이에 공공데이터 및 교육 유관기관이 보유한 데이터들을 통합적으로 수집, 분석, 제공하기 위한 교육 데이터 통합플랫폼을 구축하여 향후 자동화된 분석 서비스 제공 등의 다각적인 데이터 활용 지원 방안을 모색할 필요가 있다. 더 나아가 중장기적으로 이를 통합하여 하나의 시스템에서 관리하고 서비스를 제공할 수 있는 시스템으로 구축하여 데이터의 접근성과 데이터 활용의 편의성을 높일 필요가 있다.

일곱째, 데이터 기반 교육정책 및 행정 추진의 저변 확대를 위하여 교육 데이터를 수집·관리·분석하고 교육 데이터플랫폼을 운영·관리할 수 있는 전문 인력을 확충할 필요가 있다. 최근 온라인 수업의 활성화로 에듀테크에 대한 관심이 높아졌고 미래의 교육은 빅데이터, 인공지능, 가상·증강현실, 클라우드

드, 사물인터넷 등의 최신 정보 기술이 교육과 융합된 형태로 발전해 나갈 것으로 전망됨에 따라 관련 분야의 전문 인력 수요는 지속적으로 증가하고 있는 추세이다. 따라서 플랫폼을 운영·관리를 위한 전문가를 양성할 수 있는 전문인력양성센터를 설립하여 정보 기술과 교육의 융합 연구를 활성화할 필요가 있다.

참고문헌

- 「공공데이터 관리지침」 행정안전부 고시 제2019-71호, 2019. 9. 3. 개정
- 과학기술정보통신부(2019). **과기정통부, 데이터·인공지능(AI) 경제 활성화의 이정표 제시**. 보도자료. 2019. 1.
- 관계기관 합동(2021). **2021년도 데이터기반행정 활성화 시행계획**. 2021.5.
- 관계부처 합동(2018). **데이터 산업 활성화 전략 - I-KOREA 2.0 데이터 분야 계획, I-DATA+ -**. 2018. 6.
- 교육부, 개인정보보호위원회(2020). **교육분야 가명·익명정보 처리 가이드라인**. 2020. 11.
- 교육과학기술부, 행정안전부, 지식경제부, 방송통신위원회, 국가과학기술위원회(2012). **스마트 국가 구현을 위한 빅데이터 마스터 플랜**. 2012.11.
- 「데이터기반행정 활성화에 관한 법률」(약칭: 데이터기반행정법)[시행 2020. 12. 10.][법률 제 17370호, 2020. 6. 9., 제정]
- 「초·중등교육법」[시행 2021.3.23.][법률 제17954호, 2021.3.23., 타법개정]
- 최인봉, 시기자, 박상복(2021). **교육 빅데이터 구축 및 활용 방안 - 국가수준 학업성취도 평가 자료를 중심으로** - 한국교육과정평가원 연구보고 RRE 2021-3.
- 한국교육학술정보원(2021). **가명정보 결합 안내서**. 2021. 7.
- 행정안전부(2018). **새 정부 공공데이터 혁신을 위한 밑그림 나왔다**. 보도자료. 2018. 2.

· 논문접수 : 2023.04.05. / 수정본접수 : 2023.04.28. / 게재승인 : 2023.05.10.

ABSTRACT**A study on how to activate the use of educational data****Inbong Choi**

Research Fellow, Korea Institute for Curriculum and Evaluation

Kija Si

Senior Research Fellow, Korea Institute for Curriculum and Evaluation

Sang Bok, Park

Research Fellow, Korea Institute for Curriculum and Evaluation

The purpose of this study is to suggest ways to activate the use of educational data through an online survey targeting data-related experts and data platform operators in the field of education. The questionnaire for data-related experts in the education field consisted of contents related to the educational data utilization status and improvement plan, the educational data platform usage status and activation plan. The questionnaire for data platform operation personnel consisted of contents related to the educational data platform operation status and improvement plans. As a result of the analysis, experts pointed out the following as difficulties in utilizing data: the limits of data provision due to personal information protection, the difficulties in identifying the data materials required by users, the complexity of the data request process, lack of information required for data analysis and the limitations of data utilization due to the limited operation of the research competition. As difficulties in operating the education big data platform, they answered that there is a limit to considering the convenience of consumers due to the current lack of data experts to update the latest data and provide data tailored to consumers, and lack of budget for data platform maintenance and function improvement. In order to promote the use of educational data, the followings were proposed: the establishment of a legal and institutional basis to secure the effectiveness of data collection and use; the establishment and operation of an integrated platform to enhance the efficiency of data collection, analysis, and provision; the necessity of expanding human and physical infrastructure, such as support for data capacity enhancement.

Based on the above research results, implementation plans were proposed in the policy suggestions for the establishment of legal and institutional foundations and the establishment of human and physical infrastructure. As a way to establish a legal and

institutional foundation the followings were proposed: the revision of the provisions related to the provision of student information in 「the Elementary and Secondary Education Act」, the improvement of the effectiveness of the guidelines for handling pseudonymized and anonymous information in the education field, and the need to prepare a legal basis for the joint use of data among educational institutions. As a way to build human and physical infrastructure, the followings were emphasized: Establishment of promotion system for integrated management and utilization of data, Establishment of standard metadata system for data collection, linkage, and utilization, Establish a platform for collecting, analyzing, and providing educational data by institution, and establish a standard system for operation, Expansion of professional manpower to expand base of data-based education policy and administration promotion.

Key Words: Big data, educational data platform, data-based administration