

성인 문해능력 수준 결정요인의 변화 양상 : 의사결정나무를 활용한 2014년과 2023년 비교 연구*

박소연(안양대학교, 교수)**

권예일(위치타주립대학교, 조교수)***

요약

본 연구는 성인의 문해능력 수준을 결정하는 요인이 무엇이며 최근 10년간 어떻게 변화하였는지 탐색하는 것을 목적으로 한다. 이를 위하여 2014년과 2023년 성인문해능력조사 결과를 분석하였으며 주요 결과는 다음과 같다. 비문해자 결정요인을 탐색하기 위하여 최근 10년간 비문해자 집단을 구분하는 변수의 변화를 분석한 결과, 연령과 성별의 중요도는 감소한 반면 가구소득의 중요도는 증가한 것으로 나타났다. 2014년 비문해자 집단의 전형이 70세 이상의 여성이라고 한다면, 2023년 비문해자 집단의 전형은 가구소득 100만원 미만의 80세 이상의 노인으로 변화되었다. 문해교육 대상자 결정요인을 탐색하기 위하여 최근 10년간 문해교육 대상자 집단을 구분하는 변수의 변화를 분석한 결과, 가구소득의 중요도는 감소한 반면, 연령과 정치관심도의 중요도는 증가한 것으로 나타났다. 2014년 문해교육 대상자 집단의 전형이 가구소득 200만원 미만의 40대 이상이라고 한다면, 2023년 문해교육 대상자의 전형은 가구소득 200만원 미만으로 정치에 관심을 갖지 않는 70대 이상의 노인이다. 향후 성인의 문해능력 개발 지원을 위한 시사점 및 향후 과제를 제안하였다.

주제어: 문해능력, 성인문해교육, 비문해자, 문해교육 대상자, 의사결정나무

I. 서론

전통적으로 문해능력은 문자를 이해하고 활용하는 능력으로 읽고 쓰고 셈하는 3R을 포함

* 이 논문은 국가평생교육진흥원의 「성인문해능력조사 결과 종합 분석 연구(2024, FR 2025-2)」 수행 과정에서 수집된 데이터를 활용하여 작성되었으며, 국가평생교육진흥원의 공식적인 견해와는 다를 수 있음.

** 제1저자: 박소연(14028, 경기도 안양시 만안구 삼덕로37번길 22, sypark@anyang.ac.kr)

*** 교신저자: 권예일(1845 Fairmount St, Wichita, KS 67260, United States, yeil.kwon@wichita.edu)

하는 기초문해(basic literacy)를 의미한다. 그러나 문해는 시대와 공간이 비문해를 무엇으로 규정하는가에 따라 달라질 수 있는 상대적이고 상황적인 개념이다(이지혜, 2003). 즉, 문해의 의미는 문해 실천이 이루어지는 사회적, 역사적, 문화적 상황에 따라 매우 다양할 수 있는데(허준, 이경민, 이진희, 2016), 현재 우리나라에서 성인의 문해능력은 단순한 문자해독 능력뿐만 아니라, 가정, 경제, 공공, 여가, 미디어 등 사회적, 문화적으로 요청되는 기초생활에 필요한 능력을 의미한다(평생교육법 2조 3항).

문해능력이 삶을 살아가기 위해 필요한 기초생활 능력이라는 점을 고려할 때, 국가는 국민 모두의 문해능력 함양을 위해 노력할 필요가 있다. 우리나라에서는 1999년 제정된 평생교육법에 따라 2000년대부터 국가수준의 성인문해교육 지원을 시작하였다(서영아, 2022). 이후 성인문해교육 관련 정책은 시대적 상황을 반영하여 끊임없이 변화되어 왔다(허준, 이경민, 이진희, 2016). 특히 최근 10년 동안에는 평생교육법 개정에 따라 성인문해 영역이 확장되고 추진체계가 정비되면서 성인문해교육 예산과 사업이 크게 확대되는 등 성인문해교육을 둘러싼 정책 환경이 매우 크게 변화하였다(교육부, 2025).

성인문해교육 지원 활성화에 맞추어 우리나라 성인의 문해능력 수준도 향상되었다. 성인 문해능력조사에 따르면, 최근 10년간 비문해자의 비율은 2014년 6.4%에서 2023년 3.3%로, 문해교육 대상자의 비율은 2014년 28.5%에서 2023년 16.6%로 감소하였다(허준 외, 2024). 그럼에도 불구하고 여전히 약 146만 명은 일상생활에 필요한 기본적인 읽기, 쓰기, 셈하기가 불가능한 비문해자, 약 736만 명은 의무교육에 해당하는 중학 과정을 마치지 못한 수준에 해당하는 문해교육 대상자로 남아 있다(국가평생교육진흥원, 2024). 2023년 기준 성인문해교육 프로그램에 참여한 학습자가 약 8만 명이라는 점을 고려할 때, 여전히 대부분이 문해능력 개발 정책의 사각지대에 놓여 있는 것으로 볼 수 있다.

정책적 지원이 효과적으로 이루어지기 위해서는 무엇보다 정책 대상자를 정확하게 포착하고 이해할 필요가 있다. 문해교육의 경우에도 학습자의 시간과 공간, 지역사회 등 학습자에 대한 이해가 반드시 필요하다(한상길, 2017). 그러나 최근 10년간 비문해자와 문해교육 대상자의 인구통계학적 구성과 특징이 변화하였음에도 불구하고, 이들은 ‘나이 든 여성’일 것이라는 전통적인 가정이 여전히 지배적이다. 물론 최근 조사에서도 남성보다는 여성이, 성인보다는 노인의 문해능력이 낮게 나타난 것은 사실이다(국가평생교육진흥원, 2024). 그러나 중요한 것은 향후 다각적인 정책 지원을 위해서는 보다 다양한 대상 집단을 발굴할 필요가 있으며, 이를 위해 그들을 바라보는 보다 정교하고 포괄적인 틀거리가 필요하다는 것이다.

본 연구는 성인의 문해능력 수준을 결정하는 요인의 변화 양상을 탐색하는 것을 목적으로 한다. 이를 위하여 의사결정나무를 활용하여 최근 10간 문해능력 수준이 높은 집단과 낮

은 집단을 구분하는 결정요인이 어떻게 변화되었는지를 비교 분석한다. 이를 통해 성인의 문해능력 개발을 위한 정책의 초점을 점검하고 재설정하기 위한 시사점을 발견하고 나아가 잠재된 학습자를 신규 발굴하기 위한 기초로 활용되기 기대한다.

II. 우리나라 성인문해교육 지형의 변화

1. 성인문해교육 정책의 변화

성인문해교육 지원정책은 사회요구와 변화에 따라 변화되어 왔다(서영아, 2022). 곽삼근과 박세영(2016)은 광복 이후부터 1950년대 말까지의 정부 주도 문맹퇴치교육시기, 1960년대부터 1970년대까지의 문해교육쇠퇴기, 1980년대부터 1990년대까지의 시민사회 주도의 문해교육 유지시기, 2000년대 이후부터 현재까지의 국가 정책차원 문해교육 재편성기로 구분하였다. 2000년대 성인문해교육 지원정책이 실현되기 이전까지 성인문해교육은 민간 주도로 자율적으로 운영되어 왔으나, 2000년대에 들어 평생교육법의 시행과 함께 문해교육 지원에 대한 인식과 관심도가 향상되었고 문해교육 지원체계를 구축하기 위한 정책과 사업들이 시행되어 왔다(김수지 외, 2024). 2001년부터 ‘소외계층 평생교육 지원사업’의 일부 대상으로 저학력 성인을 지원하던 것이 2006년부터는 독립사업인 ‘성인문해교육 지원사업’으로 전환하여 추진되는 등(김명숙, 어재영, 이성엽, 2017; 서영아, 2022), 국가평생교육진흥원을 중심으로 성인문해능력 조사, 광역거점기관의 육성, 프로그램 운영지원, 성인기 학습자 학력인정, 문해교육 활성화 사업 등의 지원이 이루어지게 되었다(곽삼근, 박세영, 2016).

특히 최근 10년 동안에는 평생교육법 개정과 추진체계 정비, 예산 및 사업 확대, 통계 및 조사 진행 등 성인문해교육의 정책 환경이 크게 변화되었다(<표 1> 참고). 첫째, 최근 10년 평생교육법 개정을 통해 성인문해교육 정책 범위가 확대되고 추진체계가 정비되었다. 1999년 평생교육법이 제정되면서 성인문해교육 지원을 위한 법적 근거가 마련되었으며, 이후 2008년 전부개정을 통해 국가 및 지자체의 문해교육 지원과 문해교육 프로그램을 통한 학력인정 제도 등 성인문해교육 지원을 위한 제도가 보다 견고하게 마련된 바 있다. 2014년에는 평생교육법이 일부개정되면서 문해능력의 개념이 문자해득능력에서 사회적, 문화적으로 요청되는 기초생활능력까지 확장되었는데, 이는 곧 성인문해교육 지원사업의 변화로 이어졌다. 즉, 성인문해교육 지원이 학력인정을 위한 성인문해교육뿐만 아니라 금융문해, 교통안전문해, 건강

문해 등 다양한 영역에서의 생활문해능력 향상을 위한 정책적 지원으로 확대되었다(서영아, 2022). 2016년 평생교육법 및 시행령이 개정되면서는 국가 수준의 국가문해교육센터와 시·도 문해교육센터 설치를 위한 법적 기반이 마련되었다. 이에 따라 2016년 2월에는 국가평생교육진흥원 내 국가문해교육센터가 출범하였으며, 2016년 경기도 문해교육센터 지정을 시작으로 2022년에는 17개 광역단위 지자체 모두에 시·도문해교육센터가 설립됨에 따라 광역추진체제가 완비되었다.

둘째, 최근 10년 성인문해교육 지원을 위한 자원 투입이 확대되었다. 성인문해교육이 2000년대 이전까지는 민간 주도로 자율적으로 운영되었으나, 2000년대 이후부터는 정책적으로 추진되면서 공적인 예산이 투입되기 시작하였다. 2014년 성인문해교육 프로그램에 지원된 국고예산은 약 20억원 규모였으나 이후 지속적으로 증가하여 2023년에는 약 50억원으로 약 2.5배 규모로 확대되었다(교육부, 2025). 단순히 예산이 확대된 것뿐만 아니라 기초자치단체의 대응투자 비율도 확대되었다. 대응투자는 사업 초기인 2006년에는 기초자치단체가 관할 구역 내 성인문해교육기관과 교육청 협의를 통해 신청서를 제출하고, 국고지원금의 30% 대응투자를 필수로 하여 사업에 참여하도록 하였다. 이후 대응투자 비율은 2016년 50%, 2018년 70%, 2020년 100%로 점차적으로 확대되었다(서영아, 2022). 성인문해교육 프로그램 예산의 증가에 따라 참여학습자 역시 2014년 약 2.4만명에서 2023년 약 8.3만명으로 약 3.5배 규모로 늘어났다(교육부, 2025).

셋째, 최근 10년 성인문해능력 조사를 통해 국민의 성인문해능력 수준을 파악하여 데이터 기반의 정책 추진의 기반이 마련되었다. 물론 1989년 한국교육개발원의 ‘한국인의 문해실태 조사 연구’, 2002년도 한국교육개발원의 ‘한국 성인의 비문해 실태조사 연구’, 2008년도 국립국어원의 ‘국민의 기초 문해력 조사’ 등 이전에도 문해력 조사가 이루어진 바 있으나(박소연, 이지혜, 허준, 2014), 조사의 체계성, 일관성, 연속성, 활용도 측면에서 한계를 가지고 있었다(허준 외, 2013). 이에 2013년 교육부와 국가평생교육진흥원은 문해교육 정책을 수립을 위한 정확하고 공식적인 통계 데이터를 확보하고 문해교육 대상자를 발굴하기 위하여 국민기초문해력 측정을 위한 도구를 개발하였다(허준 외, 2013). 2014년에는 이 도구를 활용하여 1차 ‘성인문해능력조사’가 시작되었으며 이후 3년 주기로 2017년 2차 조사, 2020년 3차 조사, 2023년 4차 조사가 이루어졌다. ‘성인문해능력조사’는 조사가 시작된 2014년 통계청으로부터 국가통계로 신규승인되었으며(통계청 고시 제2014-315호), 지난 10년간 조사 도구를 정비하고 결과를 분석하기 위한 정책연구도 꾸준히 수행되어 왔다(허준 외, 2019; 허준 외 2023; 허준 외 2024; 현영섭, 신은경, 2018).

성인 문해능력 수준 결정요인의 변화 양상

<표 1> 최근 10년 성인문해교육 정책환경 변화

연도	법 및 추진체계	예산 및 사업	조사 및 통계
2014	<ul style="list-style-type: none"> 3차 평생교육진흥 기본계획 ('13-'17년) 실행 평생교육법 일부개정 (문해교육영역 확대) 	<ul style="list-style-type: none"> 성인문해교육 프로그램 국고지원 예산 20.0억원, 참여학습자 2.4만명 기초자치단체 대응투자비율 국고지원금의 30% 이상 	<ul style="list-style-type: none"> 성인문해능력 조사 국가승인통계 승인 1차 성인문해능력 조사 (4,057명)
2015		<ul style="list-style-type: none"> 성인문해교육 프로그램 국고지원 예산 22.5억원, 참여학습자 3.6만명 	
2016	<ul style="list-style-type: none"> 평생교육법 일부개정 (국가 및 시·도문해교육센터 설치) 국가문해교육센터 출범 성인문해교육 활성화를 위한 범부처 협력방안 마련 	<ul style="list-style-type: none"> 성인문해교육 프로그램 국고지원 예산 24.4억원, 참여학습자 3.6만명 기초자치단체 대응투자비율 국고지원금의 50%로 확대 	
2017	<ul style="list-style-type: none"> 2017년 성인문해교육 활성화 지원 기본계획 수립 	<ul style="list-style-type: none"> 성인문해교육 프로그램 국고지원 예산 24.4억원, 참여학습자 4.0만명 	<ul style="list-style-type: none"> 2차 성인문해능력 조사 (4,004명)
2018	<ul style="list-style-type: none"> 4차 평생교육진흥 기본계획 ('18-'22년) 수립 2018년 성인문해교육 활성화 지원 기본계획 수립 	<ul style="list-style-type: none"> 성인문해교육 프로그램 국고지원 예산 29.6억원, 참여학습자 5.2만명 기초자치단체 대응투자비율 국고지원금의 70%로 확대 	
2019	<ul style="list-style-type: none"> 2019년 성인문해교육 활성화 지원 기본계획 수립 	<ul style="list-style-type: none"> 성인문해교육 프로그램 국고지원 예산 31.9억원, 참여학습자 6.3만명 	
2020	<ul style="list-style-type: none"> 2020년 성인문해교육 활성화 지원 기본계획 수립 	<ul style="list-style-type: none"> 성인문해교육 프로그램 국고지원 예산 35.7억원, 참여학습자 7.0만명 기초자치단체 대응투자비율 국고지원금의 100%로 확대 	<ul style="list-style-type: none"> 3차 성인문해능력 조사 (10,429명)
2021	<ul style="list-style-type: none"> 2021년 성인문해교육 활성화 지원 기본계획 수립 	<ul style="list-style-type: none"> 성인문해교육 프로그램 국고지원 예산 46.2억원, 참여학습자 7.7만명 	
2022	<ul style="list-style-type: none"> 17개 시·도문해교육센터 설치 완료 2022년 성인문해교육 활성화 지원 기본계획 수립 	<ul style="list-style-type: none"> 성인문해교육 프로그램 국고지원 예산 44.6억원, 참여학습자 7.9만명 	
2023	<ul style="list-style-type: none"> 5차 평생교육진흥 기본계획 ('23-'27년) 수립 2023년 성인문해교육 활성화 지원 기본계획 수립 	<ul style="list-style-type: none"> 성인문해교육 프로그램 국고지원 예산 50.0억원, 참여학습자 8.3만명 	<ul style="list-style-type: none"> 4차 성인문해능력 조사 (10,126명)

2. 성인문해능력 수준의 변화

성인문해교육의 발전을 위해서는 성인문해능력 수준이 얼마나 향상되었는지를 정확하게 파악할 필요가 있다. 성인문해능력조사에서는 성인들의 기초문해능력을 네 수준으로 정하고 있다. 수준 1은 일상생활에 필요한 기본적인 읽기, 쓰기, 셈하기가 불가능한 수준으로, 이 수준에 해당하는 경우 비문해자로 볼 수 있다. 수준 2는 기본적으로 읽고, 쓰고, 셈하는 것은 가능하지만 일상생활에 활용은 미흡한 수준으로, 학령기로 보면 초등 수준으로 볼 수 있다. 수준 3은 가정생활 및 여가생활 등에 단순한 일상생활의 문제 해결은 가능하지만 복잡한 일상생활 문제 해결에 미흡한 수준으로, 의무교육에 해당하는 중학 과정을 마치지 못한 수준에 해당하므로 문해교육 대상자로 볼 수 있다. 수준 4는 일상생활에 필요한 충분한 문해력을 갖춘 수준이다.

성인문해능력조사는 2014년에 시작되어 3년 주기로 현재까지 총 4차까지 이루어졌다. 1차 조사는 2014년 국내 17개 시도에 거주하는 18세 이상 성인 4,057명을 대상으로 실시되었는데, 조사 결과 수준 1의 비율은 6.4%, 수준 2의 비율은 6.0%, 수준 3의 비율은 16.2%, 수준 4 비율은 71.5%로 나타났다(국가평생교육진흥원, 2015). 2차 조사는 1차 조사와 동일한 방식으로 18세 이상 성인 4,004명 대상으로 실시되었는데, 조사 결과 수준 1은 7.2%, 수준 2는 5.1%, 수준 3은 10.1%, 수준 4는 77.6%로 나타났다(국가평생교육진흥원, 2019). 3차 조사는 2020년 만 18세 이상 성인 10,429명을 대상으로 실시되었는데, 수준 1은 4.5%, 수준 2는 4.2%, 수준 3은 11.4%, 수준 4는 79.8%로 나타났다(국가평생교육진흥원, 2021). 4차 조사는 2023년 만 18세 이상 성인 10,126명을 대상으로 실시되었는데, 수준 1은 3.3%, 수준 2는 5.2%, 수준 3은 8.1%, 수준 4는 83.4%로 나타났다(국가평생교육진흥원, 2024). 총 4차에 걸쳐 이루어진 성인문해능력조사 결과 수준별 구성비와 추정 인구를 분석한 결과는 다음 <표 2>와 같다.

성인문해능력조사에 따르면, 최근 10년간 비문해자와 문해교육 대상자는 감소하였다. 먼저 비문해자에 해당하는 수준 1의 구성비는 2014년 6.4%에서 2023년 3.3%로 감소하였으며, 문해교육 대상자에 해당하는 수준 3 이하의 구성비도 2014년 28.5%에서 2023년 16.6%로 감소하였다. 추정인구를 중심으로 살펴보면, 비문해자에 해당하는 수준 1은 2014년 약 260만 명에서 2023년에는 약 146만 명으로 감소하였으며, 문해교육 대상자에 해당하는 수준 3 이하도 2014년 약 1,180만 명에서 2023년에는 약 736만 명으로 감소하였다.

<표 2> 성인문해능력조사 결과에 따른 문해능력 수준별 구성

구분	1차 (2014년)		2차 (2017년)		3차 (2020년)		4차 (2023년)	
	구성비 (%)	추정인구 (천명)	구성비 (%)	추정인구 (천명)	구성비 (%)	추정인구 (천명)	구성비 (%)	추정인구 (천명)
수준 1	6.4	2,642	7.2	3,111	4.5	2,001	3.3	1,460
수준 2	6.0	2,475	5.1	2,173	4.2	1,856	5.2	2,313
수준 3	16.2	6,679	10.1	4,328	11.4	5,039	8.1	3,584
수준 4	71.5	29,555	77.6	33,366	79.8	35,185	83.4	36,887
계	100.0	41,352	100.0	42,979	100.0	44,081	100.0	44,245

※ 출처: 각 차수별 성인문해능력 조사 보고서.

성인의 문해능력은 연령 및 세대, 성별 등 개인 배경과 관련되어 있다. 일반적으로 성인의 문해능력은 연령이 증가함에 따라 감소하는 것으로 알려져 있는데(김경년, 2019; 이주호, 최슬기, 2015; Willms & Murray, 2007), 2012년 성인역량조사(PIAAC) 조사 결과 우리나라의 16-24세와 55-65세 간 문해능력의 차이가 매우 크게 나타나 이러한 경향이 실증적으로 확인되었다(황혜진, 2015). 연령에 따른 문해능력의 변화를 보다 정교하게 살펴보면, 문해능력은 일정 연령에서 최고점에 도달한 이후 감소하는 경향을 보인다(한승희, 2014; Reder, 2008; Wylie & Hodgen, 2011), 우리나라의 경우 우리나라의 국제학업성취도 평가(PISA)의 읽기 영역의 점수는 최고 수준인 데 반해 성인역량조사(PIAAC)의 문해능력 평가는 평균 이하라는 점을 고려할 때(황혜진, 2015), 이러한 양상이 극적으로 나타난다고 할 수 있다. 또한 성별에 따른 차이를 살펴보면 남성보다 여성의 비문해 비율이 높은 경향을 보이는데, 그 원인으로 남성으로부터의 억압, 여성의 학습에 대한 부정적인 사회적 인식, 재정 부족 등을 들 수 있다(Rockill, 1993). 문해 영역별로 살펴보면 여성은 산문문해가 상대적으로 강하고, 남성은 수리문해가 상대적으로 강한 것으로 나타났다(Satherley & Lawes, 2008).

III. 연구방법

1. 연구대상

본 연구에서는 성인의 문해능력 수준 결정요인을 탐색하기 위하여, ‘성인문해능력조사’의

데이터를 분석하였다. ‘성인문해능력조사’는 성인 기초 생활문해수준을 진단을 통해 실시하는 유일한 조사로(서영아, 2022), 2014년 국가승인통계로 지정된 후 우리나라 성인의 문해능력 수준을 나타내는 공식적인 통계 데이터로 활용되고 있다. 2014년 1차 조사를 시작으로 2017년 2차 조사, 2020년 3차 조사, 2023년 4차 조사가 이루어졌으며, 각 조사의 표본은 통계적 할당이 이루어졌으므로, 성인의 문해능력 수준을 공시적(synchronic) 및 통시적(diachronic)으로 비교 분석하기에 가장 적합한 데이터라고 판단되었다. 최근 10년간의 변화를 파악하기 위하여 2014년과 2023년 데이터를 분석에 활용하였는데, 해당 데이터 확보 및 활용을 위해 해당 기관의 승인 절차를 거쳤다. 분석 대상이 되는 2014년과 2023년 조사 표본의 인구통계학적 특성별 구성은 다음 <표 3>과 같다.

2. 분석범위

분석에는 총 9개의 변수가 활용되었으며, 구체적인 변수명과 라벨은 다음 <표 4>와 같다. 먼저 종속변수인 문해능력 수준은 상대적으로 문해능력 수준이 낮은 비문해자와 문해교육 대상자, 상대적으로 문해능력 수준이 높은 문해자와 문해교육 비대상자로 구분하였다. 비문해자는 일상생활에 필요한 기본적인 읽기, 쓰기, 셈하기가 불가능한 수준으로, 성인문해능력조사 결과 수준 1에 해당한다. 문해교육 대상자는 가정생활 및 여가생활 등에 단순한 일상생활의 문제 해결은 가능하지만 복잡한 일상생활 문제 해결에 미흡한 수준으로, 의무교육에 해당하는 중학 과정을 마치지 못한 수준에 해당하므로 문해교육 대상자로 볼 수 있다. 성인문해능력조사 결과에서는 수준 3 이하에 해당한다. 비문해와 문해, 대상자와 비대상자의 두 가지의 분류 기준을 적용하여 문해능력이 낮은 수준과 높은 수준을 구분하여 분석에 활용하였다.

결정요인 탐색을 위한 독립변수로는 성별, 연령, 가구소득, 경제활동상태, 지역유형과 같은 주요 인구인구통계학적 특성 변수와 생활만족도, 교류활동 빈도, 정치관심도와 같은 응답자 특성변수가 포함되었다. 본 연구의 주요 목적은 예측 정확도 자체보다는 문해능력에 영향을 미치는 요인들을 탐색하고 각 집단의 특성을 기술적으로 파악하는 것에 있다. 따라서, 문해능력에 가장 강력하고 직접적인 영향을 미치는 교육수준 변수는 분석모형에서 제외하였다. 이는 교육수준 변수에 의해 다른 유의미한 설명 변수들의 영향력이 가려지는 현상을 방지하고, 이들 변수의 상대적 중요성을 보다 명확히 파악하기 위함이다.

성인 문해능력 수준 결정요인의 변화 양상

<표 3> 분석대상 표본의 인구통계학적 특성별 구성

구분		2014년		2023년	
		인원(명)	비율(%)	인원(명)	비율(%)
전체		4,057	100.0	10,126	100.0
성별	남성	2,005	49.4	5,013	49.5
	여성	2,052	50.6	5,113	50.5
연령	18-29세	751	18.5	1,615	15.9
	30-39세	746	18.4	1,473	14.5
	40-49세	867	21.4	1,765	17.4
	50-59세	782	19.3	2,029	20.0
	60-69세	463	11.4	1,776	17.5
	70-79세	323	8.0	1,185	11.7
	80세 이상	125	3.1	283	2.8
학력	무학	233	5.7	357	3.5
	초졸	384	9.5	500	4.9
	중졸	400	9.9	790	7.8
	고졸	1,677	41.3	3,796	37.5
	대졸 이상	1,363	33.6	4,683	46.2
가구소득	100만원 미만	441	10.9	365	3.6
	100-200만원	671	16.5	606	6.0
	200-300만원	747	18.4	1,247	12.3
	300-400만원	1,042	25.7	1,703	16.8
	400-500만원	556	13.7	1,824	18.0
	500만원 이상	600	14.8	4,381	43.3
경제활동 상태	취업	2,566	63.2	7,159	70.7
	실업	133	3.3	795	7.9
	비경제활동	1,358	33.5	2,172	21.4
지역유형	서울 및 광역시	1,785	44.0	4,395	43.4
	중소도시	1,917	47.3	4,784	47.2
	농산어촌	355	8.8	947	9.4

<표 4> 분석 활용 변수

변수명	라벨
문해능력 수준	비문해자(수준 1), 문해자(수준 2-4), 문해교육 대상자(수준 1-3), 문해교육 비대상자(수준 4)
성별	남성, 여성
연령	18-29세, 30-39세, 40-49세, 50-59세, 60-69세, 70-79세, 80세 이상
가구소득	100만원 미만, 100-199만원, 200-299만원, 300-399만원, 400-499만원, 500만원 이상
경제활동상태	취업, 실업, 비경제활동
지역유형	서울 및 광역시, 중소도시, 농산어촌
생활만족도	만족, 불만족
교류활동빈도	일주일에 2-3번, 한달에 2-3번, 한달에 1번 이하
정치관심도	관심 있음, 관심 없음

3. 분석방법

먼저 최근 10년간 비문해자 집단과 문해교육 대상자 집단의 성별, 연령, 가구소득, 경제활동상태, 지역유형, 생활만족도, 교류활동 빈도, 정치관심도의 변화를 살펴보았다. 이를 위하여 비문해자 집단과 문해교육 대상자 집단에 대한 기술통계분석을 실시하였으며, 교차검증 (Chi-square Test)을 통해 각 변수들의 범주별 비문해자 수, 문해교육 대상자 수의 분포에 차이가 있는지를 분석하였다.

비문해자와 문해자, 문해교육 대상자와 비대상자를 구분하는 요인을 탐색하기 위하여 의사결정나무(decision tree) 분석을 실시하였다. 의사결정나무 분석은 지도학습(supervised machine learning) 알고리즘의 한 유형으로, 분류 및 회귀나무(Classification and Regression Trees, 이하 CART)의 두 종류가 있다(Murphy, 2012). CART는 구조가 단순하여 모형에 대한 해석이 어렵지 않기 때문에 복잡한 수학적모형을 기반으로 한 로지스틱 회귀나 판별분석과 같은 다른 분류 모델에 비해 접근성이 높고 직관적으로 이해할 수 있다(James et al., 2021). 또한, 결측치 처리 및 변수 스케일링에 민감하지 않아 데이터의 전처리 과정이 비교적 간단하고 특정한 확률분포 가정이 필요하지 않기 때문에 실무자들이 모형을 비교적 쉽게 구축할 수 있다는 장점이 있어 지난 40여 년간 통계학을 비롯한 다양한 영역에서 가장 중요한 예측 모형 중 하나로 사용되어 왔다(Loh, 2014). CART는 연속형 반응변수의 값을 예측하는 회귀나무(regression tree)와 이산형 반응변수의 범주를 분류하는 분류나무(classification tree)의 두 가지 모형이 있는데(Breiman et al., 2017), 본 연구의 경우 반응변수가 ‘비문해와 문해’ 또

는 ‘문해교육 대상자와 문해교육 비대상자’에 대한 이진 예측문제를 다루므로 분류나무를 사용하였다.

최적 집단 분할을 위한 함수로 지니지수(Gini index)를 사용하였다. 분류나무는 모든 설명 변수들 중 표본을 최적으로 분할할 수 있는 변수와 분할지점을 찾아 완전히 구분되는 두 개의 하위표본으로 구분하는 과정을 반복하는데, 최적 분할의 정도를 측정하기 위해서 지니지수, 엔트로피(Entropy) 등의 함수가 사용된다(Murphy, 2012). 본 연구에서는 가장 보편적으로 이용되고 있는 지니지수를 사용하였는데, 이는 G 개의 집단 분할에 대한 분산으로 해석될 수 있으며, 각 노드에서 다음과 같이 정의된다.

$$Gini(N_j) = \sum_{g=1}^G \hat{p}_{jg}(1 - \hat{p}_{jg}) = 1 - \sum_{g=1}^G \hat{p}_{jg}^2$$

여기서 \hat{p}_{jg} 는 j 번째 노드에 포함된 학습 데이터 중 g 번째 집단에 속하는 관측치의 비율을 나타낸다. 지니지수는 모든 \hat{p}_{jg} 의 값이 0 또는 1에 가까울수록 0에 가까워지는데, 이는 해당 노드의 관측치가 어느 한 집단에 집중되어 있음을 의미한다. 극단적인 경우로 지니지수가 0이라면, 이는 모든 관측치가 하나의 집단에만 포함되어 있음을 의미한다. 한편, 지니지수는 모든 \hat{p}_{jg} 값이 $1/G$ 일 경우 즉, 각 집단에 같은 수의 관측치가 있는 경우, 최대값 $(G-1)/G$ 를 가지게 된다. 예를 들어, 본 연구와 같이 두 개의 집단에 대한 모형인 경우 ($G=2$), 지니지수는 0과 1/2 사이의 값을 갖는다. 최적의 분할은 두 하위 노드에서 지니지수를 최소화하는, 즉 불순도(impurity)를 최소화하는 분할이다.

모형의 적합 시, 그 일반화 성능을 평가하기 위한 대표적인 방법으로 k-폴드 교차검증(k-fold cross-validation)이 활용된다. 그러나 집단간 표본수의 불균형이 심한 경우, 이러한 불균형으로 인한 편향을 완화하고 보다 신뢰성 있는 평가를 도모하기 위해 층화교차검증(stratified cross-validation) 기법이 적용되는데, 이 방법은 각 폴드 내에서 집단 비율이 전체 데이터의 분포와 유사하게 유지되도록 구성되어, 전통적인 교차검증보다 더 일관된 추정결과를 제공할 수 있다. 본 연구에서는 전체 표본을 학습데이터(training set)와 검증데이터(test set)로 80:20의 비율로 분할하였다. 2014년 및 2023년의 표본크기를 고려하여 교차검증의 폴드수는 각각 $k=5$ 과 $k=10$ 으로 설정하였다. 특히, 비문해집단의 표본수가 문해집단에 비해 현저히 적은 비대칭적인 분포 특성을 감안하여, 모형의 안정성을 제고하고 소수 집단의 특성이 왜곡되지 않도록 층화교차검증(stratified cross-validation) 기법을 적용하였다.

(Sammut & Webb, 2017). 모형 학습시에는 최대 트리 깊이는 4로 설정하였으며, 각 분할을 위한 최소 표본수는 100으로 제한하였다. 또한 리프노드(leaf node)에서의 최소 표본수는 해당 노드의 집단 비율 추정값의 표준오차가 과도하게 증가하지 않도록 50으로 설정하였다.

모형 성능의 평가는 [그림 1]과 같은 혼동행렬(confusion matrix)을 기반으로 수행되었다. 본 연구에서는 상대적으로 문해능력 수준이 낮은 비문해 집단(수준 1)과 문해교육 대상자 집단(수준 1-3)을 Positive, 상대적으로 문해능력 수준이 높은 문해 집단(수준 2-4)과 문해교육 비대상자 집단(수준 4)을 Negative로 정의하였다. 특이도(Specificity)는 실제 문해 집단 중 올바르게 분류된 비율(True Negative Rate)을, 민감도(Sensitivity)는 실제 비문해 집단 중 정확히 예측된 비율(True Positive Rate)을 의미한다. 한편, 정확도(Accuracy)는 전체 표본 중 실제 범주와 예측 범주가 일치하는 비율을 의미한다.

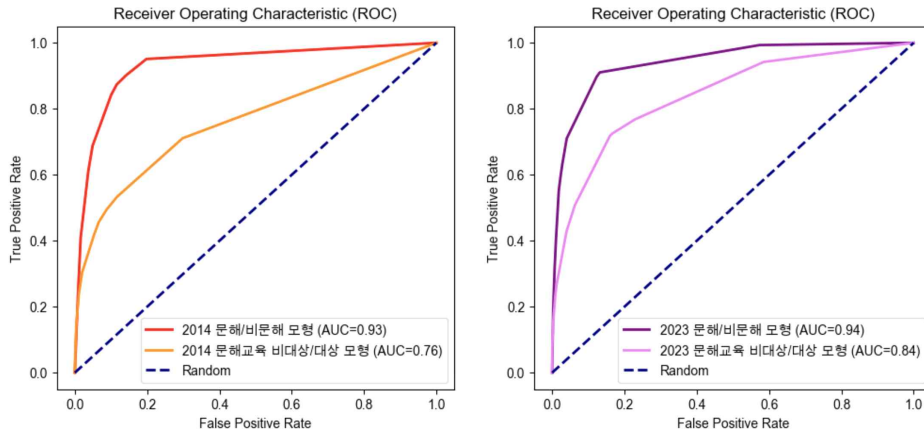
		Predicted		(Recall)
		Negative	Positive	
Actual	Negative	TN (True Negative)	FP (False Positive)	$\frac{TN}{TN+FP}$ (특이도: Specificity)
	Positive	FN (False Negative)	TP (True Positive)	$\frac{TP}{FN+TP}$ (민감도: Sensitivity)
		$\frac{TN}{TN+FN}$ (정밀도: Precision)	$\frac{TP}{FP+TP}$ (정밀도: Precision)	$\frac{TP+TN}{TP+TN+FP+FN}$ (정확도: Accuracy)

[그림 1] 혼동행렬

본 연구에서 활용된 자료의 문해 대 비문해 비율은 2014년의 경우 93% 대 7%, 2023년의 경우 97% 대 3%로 표본 분포가 비대칭적이며, 그 영향으로 인해 예측모형의 정확도가 높게 나타날 수 있다. 즉, 어떠한 모형도 사용하지 않고 모든 사례를 문해집단으로 분류하더라도 2014년과 2023년의 정확도는 각각 93%와 97%에 도달하므로, 단순한 정확도 수치는 모형의 성능을 제대로 평가하기 어렵다. 이에 본 연구에서는 비문해 집단의 프로파일링을 주요 분석 목표로 설정하고, 모형의 정확도 일부를 희생하더라도 실제 비문해를 문해로 잘못 예측하는 경우(False Negative Rate)를 최소화하는 것을 중점적으로 고려하였다. 혼동행렬은 기준 예측 확률(threshold)에 따라 달라지며, 일반적으로는 0.5를 사용하지만, 분석 목적에 따라 이보다

낮은 임계값을 설정하기도 한다. 본 연구에서는 문해/비문해 예측모형의 기준값을 0.1로, 문해교육대상/비대상 예측모형의 경우 0.3으로 설정하여 혼동행렬을 산출하였다.

모형 성능을 평가하기 위한 척도에는 AUC(Area Under the Curve), KS 통계량(Kolmogorov - Smirnov), Divergence 등이 있는데, 본 연구에서는 AUC를 사용하였다. AUC는 ROC 곡선(Receiver Operating Characteristic Curve) 하의 면적으로, 정확도에 비해 모형의 성능을 보다 종합적으로 파악할 수 있는 대표적인 척도이다. 2014년과 2023년의 ROC 곡선은 다음 [그림 2]과 같다.



[그림 2] ROC 곡선

이상의 각 모형별 및 혼동행렬 기반의 주요 성능 지표 및 AUC 값을 요약한 결과는 다음 <표 5>와 같다.

<표 5> 모형 성능 평가 결과

연도	모형	특이도	민감도	정확도	AUC		
					AUC	표준오차	95% 신뢰구간
2014년	문해/비문해	0.88	0.87	0.88	0.93	0.010	(0.911, 0.952)
	문해교육 비대상/대상	0.88	0.53	0.78	0.76	0.009	(0.743, 0.777)
2023년	문해/비문해	0.96	0.71	0.95	0.94	0.010	(0.921, 0.959)
	문해교육 비대상/대상	0.84	0.72	0.82	0.84	0.006	(0.829, 0.853)

마지막으로 의사결정나무 분석을 통해 탐색된 변수들의 기여도를 상대적으로 비교하기 위하여 변수 중요도(feature importance)를 산출하였다. 변수 중요도는 의사결정나무에서 한 변수가 분할에 사용될 때, 해당 분할이 전체 불순도(impurity)를 얼마나 감소시켰는지를 계산하여 그 변수가 트리 전체에서 분할에 사용된 모든 노드에서의 불순도 감소량을 모두 더한 값의 상대적 비율을 의미한다.

IV. 연구결과

1. 비문해자 집단의 변화

가. 비문해자 구성비 변화

2014년과 2023년 간에 각 변수들의 범주별 비문해자 수의 분포에 차이가 있는지를 분석하기 위하여 교차검증을 실시한 결과는 다음 <표 6>과 같다. 분석 결과, 생활만족도와 교류활동빈도를 제외한 모든 변수들에서 유의미한 차이가 있는 것으로 나타났다*. 특히 연령의 경우 60대의 비율이 2014년 20.4%에서 2023년 4.8%로 감소하였으며, 80세 이상의 비율은 2014년 28.4%에서 2023년 44.5%로 증가하였다. 가구소득의 경우 100만원 미만의 비율이 2014년 55.1%에서 2023년 41.0%로 감소한 것으로 나타났다.

나. 비문해자 집단의 특성 변화

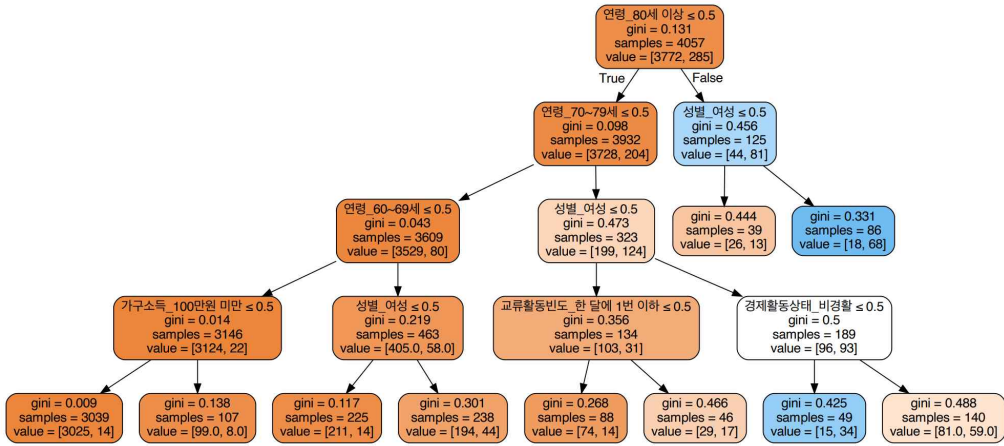
비문해자 집단을 구분하는 변수의 변화를 살펴보기 위하여 수준 1과 수준 2 이상을 구분하는 의사결정나무 분석을 실시하였다. 먼저 2014년 조사 데이터를 활용하여 비문해자 집단을 구분한 결과는 다음 [그림 3]과 같다. 80세 이상의 여성인 경우 비문해자 집단으로 구분되었으며, 이 집단의 비문해자 비율은 79.1%이며 지니지수는 .331로 나타났다. 70대 여성이면서 취업 혹은 실업 상태인 경우에도 비문해자 집단으로 구분되었으며, 이 집단의 비문해자 비율은 69.4%이며 지니지수는 .425로 나타났다. 종합하면 80세 이상의 여성, 70대이며 여전히 경제활동에 참여하는 여성이 비문해자로 구분되었으며, 비문해자와 문해자 집단을 구분하는 변수로는 연령, 성별, 경제활동 상태가 채택되었다.

* 검증의 타당성을 확보하기 위하여, 연령, 학력, 소득, 경제활동 변수 내에서 비문해자 수가 매우 적은 범주들은 연구의 맥락을 해치지 않는 범위 내에서 보다 넓은 범주로 통합하여 분석하였다. <표 6>에서 음영 처리된 부분은 교차검증 시 통합된 범주를 나타낸다.

<표 6> 비문해자의 인구통계학적 특성별 2014년과 2023년 구성비 차이

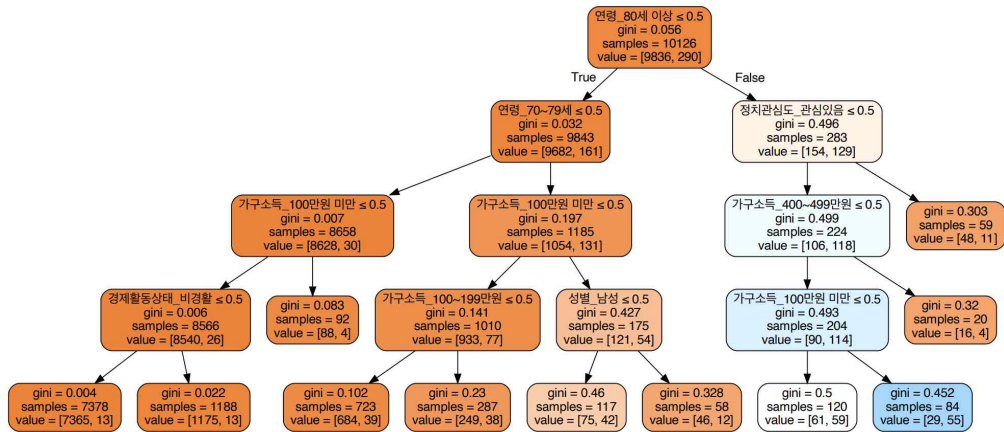
구분		2014년 (n=4,057)			2023년 (n=10,126)			χ^2
		해당 표본수	비문해자 수	비문해자 비율	해당 표본수	비문해자 수	비문해자 비율	
성별	남성	2,005	66	23.2	5,013	89	30.7	4.14* (d#1)
	여성	2,052	219	76.8	5,113	201	69.3	
연령	18-29세	751	1	0.4	1,615	3	1.0	38.96*** (d#3)
	30-39세	746	0	0.0	1,473	3	1.0	
	40-49세	867	3	1.1	1,765	5	1.7	
	50-59세	782	18	6.3	2,029	5	1.7	
	60-69세	463	58	20.4	1,776	14	4.8	
	70-79세	323	124	43.5	1,185	131	45.2	
	80세 이상	125	81	28.4	283	129	44.5	
가구 소득	100만원 미만	441	157	55.1	365	119	41.0	28.96*** (d#3)
	100-200만원	671	82	28.8	606	68	23.4	
	200-300만원	747	20	7.0	1,247	37	12.8	
	300-400만원	1,042	17	6.0	1,703	28	9.7	
	400-500만원	556	6	2.1	1,824	16	5.5	
	500만원 이상	600	3	1.1	4,381	22	7.6	
경제 활동 상태	취업	2,566	91	31.9	7,159	70	24.1	9.31** (d#1)
	실업	133	11	3.9	795	0	0.0	
	비경제활동	1,385	183	64.2	2,172	220	75.9	
지역 유형	서울 및 광역시	1,785	87	30.5	4,395	72	24.8	6.07* (d#2)
	중소도시	1,917	126	44.2	4,784	158	54.5	
	농산어촌	355	72	25.3	947	60	20.7	
생활 만족도	만족	3,351	199	69.8	8,874	210	72.4	.47 (d#1)
	불만족	706	86	30.2	1,252	80	27.6	
교류 활동 빈도	주 2-3번 이상	1,689	126	44.2	2,937	100	34.5	5.75 (d#2)
	달 2-3번	1,301	41	14.4	3,667	51	17.6	
	달 1번 이하	1,067	118	41.4	3,522	139	47.9	
정치 관심도	관심 있음	1,820	59	20.7	3,649	40	13.8	4.81* (d#1)
	관심 없음	2,237	226	79.3	6,477	250	86.2	

* p-value < .05, ** p-value < .01, *** p-value < .001



[그림 3] 2014년 비문해자 집단 구분 의사결정나무 분석

2023년 조사 데이터를 활용하여 비문해자 집단을 구분한 결과는 다음 [그림 4]와 같다. 80세 이상으로 정치에 관심이 없으며 가구소득이 100만원 미만인 경우 비문해자 집단으로 구분되었으며, 이 집단의 비문해자 비율은 65.5%이며 지니계수는 .452로 나타났다. 80세 이상으로 정치에 관심이 없으며 가구소득이 100만원 이상 400만원 미만인 경우 비문해자 집단으로 구분되었는데, 이 집단의 비문해자 비율은 49.2%이며 지니계수는 .500으로 나타났다. 종합하면 80세 이상이고 정치에 무관심하며 가구소득 400만원 미만인 경우 비문해자로 구분되었으며, 비문해자와 문해자 집단을 구분하는 변수로는 연령, 정치관심도, 가구소득이 채택되었다.



[그림 4] 2023년 비문해자 집단 구분 의사결정나무 분석

문해/비문해 모형에서 설명변수들이 예측에 기여한 상대적 중요도의 변화는 다음 <표 7>과 같다. 2014년과 2023년 모두에서 연령 관련 변수가 가장 높은 중요도를 보이며 핵심적인 설명 변수로 작용하였으나, 2014년 88.4%에서 2023년 78.3%로 다소 감소한 것으로 나타났다. 성별 관련 변수의 경우에도 2014년 7.1%에서 2023년 1.1%로 상대적 기여도가 감소하였다. 반면 가구소득 관련 변수는 2014년 0.5%에서 2023년 12.6%로, 정치적 관심도 관련 변수는 2014년 1.3%에서 2023년 6.5%로 상대적 기여도가 증가한 것으로 나타났다.

<표 7> 2014년과 2023년 비문해/문해 설명 변수의 중요도 변화

구분		2014		2023년		증감
		중요도	계	중요도	계	
성별	남성/여성	.071	.071	.011	.011	감소
연령	60-69세	.057	.884		.783	감소
	70-79세	.392		.144		
	80세 이상	.435		.639		
가구소득	100만원 미만	.005	.005	.125	.126	증가
	500만원 이상			.001		
생활만족도	만족/불만족	.026	.026	.014	.014	감소
교류활동빈도	달 1번 이하		.000	.002	.002	증가
정치관심도	관심 있음/없음	.013	.013	.065	.065	감소

2. 문해교육 대상자 집단의 변화

가. 문해교육 대상자 집단의 구성비 변화

2014년과 2023년 간에 각 변수들의 범주별 문해교육 대상자 수 분포에 차이가 있는지를 분석하기 위하여 교차검증을 실시한 결과는 다음 <표 8>과 같다. 분석 결과, 성별, 지역유형, 생활만족도를 제외한 모든 변수들에서 유의미한 차이가 있는 것으로 나타났다. 특히 연령의 경우 70대의 비율이 2014년 21.5%에서 2023년 37.9%로 증가하였으며, 50대의 비율은 2014년 20.5%에서 2023년 11.7%로 감소하였다. 소득의 경우 100만원 미만의 비율이 2014년 28.0%에서 2023년 16.0%로 감소하였으며, 500만원 이상의 비율은 2014년 6.0%에서 2023년 14.2%로 증가하였다. 교류활동 빈도의 경우 주 2-3번 이상의 비율은 2014년 43.8%에서 2023년 27.5%로 감소하였으나, 달 1번 이하의 비율은 2014년 33.8%에서 2023년 44.8%로 증가하였다.

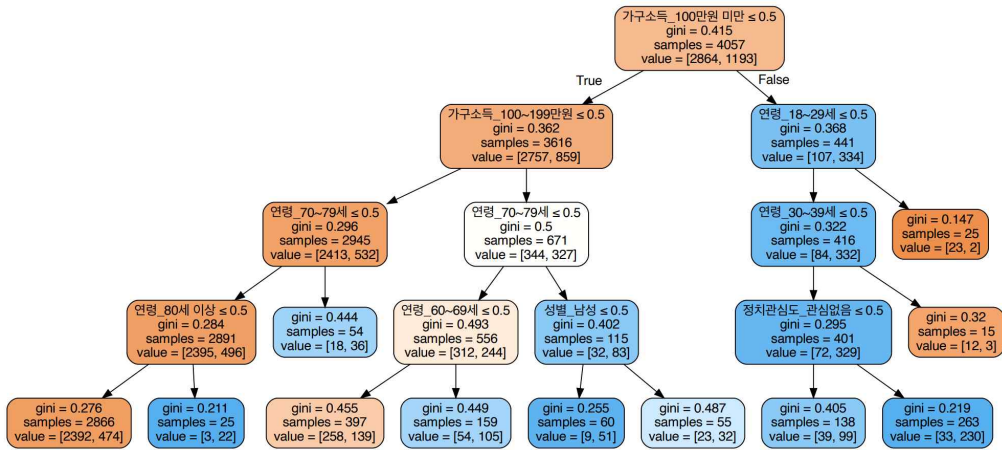
<표 8> 문해교육 대상자의 인구통계학적 특성별 2014년과 2023년 구성비 차이

구분		2014년 (n=4,057)			2023년 (n=10,126)			χ^2
		해당 표본수	대상자수	대상자 비율	해당 표본수	대상자수	대상자 비율	
성별	남성	2,005	480	40.2	5,013	702	41.2	.27 (d#1)
	여성	2,052	713	59.8	5,113	1002	58.8	
연령	18-29세	751	91	7.6	1,615	46	2.7	202.10*** (d#6)
	30-39세	746	80	6.7	1,473	60	3.5	
	40-49세	867	147	12.3	1,765	90	5.3	
	50-59세	782	244	20.5	2,029	199	11.7	
	60-69세	463	260	21.8	1,776	431	25.3	
	70-79세	323	257	21.5	1,185	645	37.9	
	80세 이상	125	114	9.6	283	233	13.7	
가구 소득	100만원 미만	441	334	28.0	365	272	16.0	153.10*** (d#5)
	100-200만원	671	327	27.4	606	324	19.0	
	200-300만원	747	200	16.8	1,247	369	21.7	
	300-400만원	1,042	177	14.8	1,703	260	15.3	
	400-500만원	556	83	7.0	1,824	237	13.9	
	500만원 이상	600	72	6.0	4,381	242	14.2	
경제 활동 상태	취업	2,566	619	51.9	7,159	851	49.9	16.60*** (d#2)
	실업	133	45	3.8	795	27	1.6	
	비경제활동	1,385	529	44.3	2,172	826	48.5	
지역 유형	서울 및 광역시	1,785	428	35.9	4,395	652	38.3	2.82 (d#2)
	중소도시	1,917	578	48.4	4,784	817	47.9	
	농산어촌	355	187	15.7	947	235	13.8	
생활 만족도	만족	3,351	910	76.3	8,874	1,299	76.2	.00 (d#1)
	불만족	706	283	23.7	1,252	405	23.8	
교류 활동 빈도	주 2-3번 이상	1,689	522	43.8	2,937	468	27.5	83.26*** (d#2)
	달 2-3번	1,301	268	22.5	3,667	473	27.8	
	달 1번 이하	1,067	403	33.8	3,522	763	44.8	
정치 관심도	관심 있음	1,820	441	37.0	3,649	479	28.1	25.39*** (d#1)
	관심 없음	2,237	752	63.0	6,477	1,225	71.9	

*** p-value < .001

나. 문해교육 대상자 집단의 특성 변화

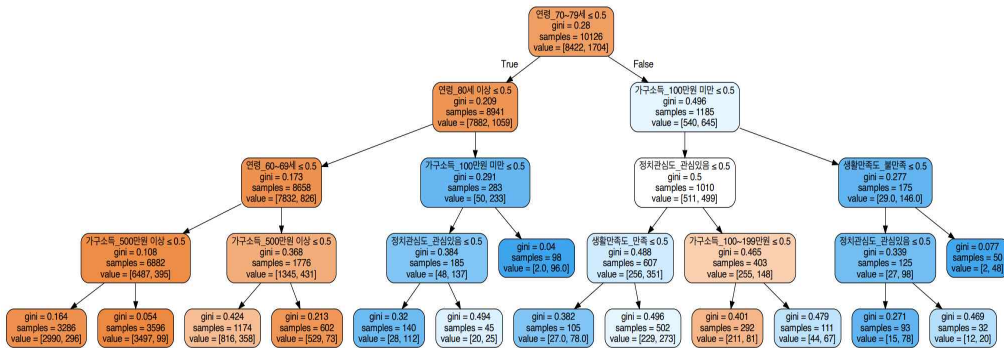
문해교육 대상자 집단을 구분하는 변수의 변화를 살펴보기 위하여 수준 3 이하와 수준 4를 구분하는 의사결정나무 분석을 실시하였다. 먼저 2014년 조사 데이터를 활용하여 문해교육 대상자를 구분한 결과는 다음 [그림 5]와 같다. 가구소득 100만원 미만의 40대 이상인 경우 문해교육 대상자 집단으로 구분되었는데, 정치에 관심 없는 집단의 경우 비문해자 비율은 86.5%, 지니계수는 .219, 관심 있는 집단의 경우 비문해자 비율은 71.7%, 지니계수는 .405로 나타났다. 가구소득 100만원 이상 200만원 미만의 70대인 경우에도 문해교육 대상자로 구분되었는데, 남성인 경우 비문해자 비율은 58.2%, 지니계수는 .487, 여성인 경우 비문해자 비율은 85.0%, 지니계수는 .255로 나타났다. 가구소득 100만원 이상 200만원 미만의 60대도 문해교육 대상자로 구분되었는데, 이때 비문해자 비율은 66.0%, 지니계수는 .449로 나타났다. 가구소득 100만원 이상 200만원 미만의 80대 역시 문해교육 대상자로 구분되었는데, 이때 비문해자 비율은 88.0%, 지니계수는 .211로 나타났다. 종합하면 가구소득 100만원 미만이면서 40대 이상, 가구소득 100만원 이상 200만원 미만이면서 60대 이상이 문해교육 대상자로 구분되었으며, 문해교육 대상자와 비대상자 집단을 구분하는 변수로는 가구소득과 연령이 채택되었다.



[그림 5] 2014년 문해교육 대상자 집단 구분 의사결정나무 분석

2023년 조사 데이터를 활용하여 문해교육 대상자 집단을 구분한 결과는 다음 [그림 6]과 같다. 70대이며 가구소득이 100만원 미만인 경우 문해교육 대상자 집단으로 구분되었는데, 생활에 불만족한 경우 비문해자 비율은 96.0%, 지니계수는 .077, 생활에 만족하고 정치에 관심이 없는 경우 비문해자 비율은 62.5%, 지니계수는 .469, 생활에 만족하고 정치에 관심이 있는

경우 비문해자 비율은 83.9%, 지니계수는 .217로 나타났다. 70대이며 정치에 관심이 있는 경우에는 가구소득이 100만원 이상 200만원 미만도 문해교육 대상자로 구분되었는데, 이때 비문해자 비율은 60.3%, 지니계수는 .479로 나타났다. 70대이며 가구소득이 100만원 이상이고 정치에 관심이 없는 경우 문해교육 대상자로 구분되었는데, 생활에 만족하는 경우 비문해자 비율은 54.4%, 지니계수는 .496, 생활에 만족하지 않는 경우 비문해자 비율은 74.2%, 지니계수는 .382로 나타났다. 80세 이상인 경우에는 문해교육 대상자로 구분되었는데, 가구소득이 100만원 미만인 경우 비문해자 비율은 98.0%, 지니계수는 .040, 가구소득이 100만원 이상이면 정치에 관심이 있는 경우 비문해자 비율은 55.5%, 지니계수는 .494, 가구소득이 100만원 이상이면 정치에 관심이 없는 경우 비문해자 비율은 80.0%, 지니계수는 .320으로 나타났다. 종합하면 80세 이상, 가구소득 100만원 미만의 70대, 가구소득 100만원 이상 200만원 미만 70대 중 정치에 무관심한 경우 문해교육 대상자로 구분되었으며, 문해교육 대상자와 비대상자 집단을 구분하는 변수로는 연령, 가구소득, 정치관심도가 채택되었다.



[그림 6] 2023년 문해교육대상자 집단 구분 의사결정나무 분석

문해교육 대상/비대상모형에서 설명변수들이 예측에 기여한 상대적 중요도는 다음 <표 9>와 같다. 2014년과 2023년 모두 연령 관련 변수의 상대적 기여도가 가장 높은 것으로 나타났다. 특히 2023년에는 86.1% 수준으로 크게 증가하여 지배적인 설명력을 갖게 되었다. 정치관심도는 2023년 새로운 설명변수로 등장하였으며, 중요도는 2.4% 수준인 것으로 나타났다. 반면 가구소득 관련 변수의 중요도 2014년 48.8%에서 2023년 10.0% 수준으로 크게 감소한 것으로 분석되었다.

<표 9> 2014년과 2023년 문해교육 대상자/비대상자 설명 변수의 중요도 변화

구분		2014		2023년		변화 (B-A)
		중요도	계 (A)	중요도	계 (B)	
성별	남성/여성	.019	.019		.000	감소
연령	50-59세	.053	.473		.861	증가
	60-69세	.185		.109		
	70-79세	.198		.426		
	80세 이상	.037		.326		
가구소득	100만원 미만	.488	.488	.047	.100	감소
	500만원 이상			.053		
생활만족도	만족/불만족		.000	.010	.010	증가
교류활동빈도	달 2-3번	.020	.020		.004	감소
	달 1번 이하			.004		
정치관심도	관심 있음/없음		.000	.024	.024	증가

V. 결론

성인의 문해능력 수준 결정요인은 어떻게 변화되었는가? 본 연구는 성인의 문해능력 수준을 결정하는 요인이 어떻게 변화하였는지를 탐색하는 것을 목적으로 한다. 이를 위하여 의사결정나무를 활용하여 2014년과 2023년의 결과를 비교 분석하였다. 비문해자 수준에 대한 분석과 문해교육 대상자 수준에 대한 분석의 주요 결과와 이에 대한 논의는 다음과 같다.

먼저 최근 10년간 어떤 변수의 범주에서 비문해자의 구성이 변화되었는지를 분석한 결과, 연령, 가구소득, 경제활동 상태, 지역유형, 정치관심도, 성별에 따라 2014년과 2023년 비문해자 수의 분포에 유의미한 차이가 있는 것으로 나타났다. 연령의 경우 특히 그 차이가 가장 뚜렷하게 나타났는데, 이는 10년간 세대 이동에 따른 자연스러운 변화로 추정된다. 즉, 2014년 기준 60대는 2023년 기준 70대로 이동하고, 2014년 기준 60세 미만이 2023년 기준 60대를 새롭게 구성함으로써 비문해율이 연쇄적으로 변동된 것으로 해석되는바, 해당 코호트의 문해능력 수준이 향상되었다고 보기는 어렵다.

비문해자 결정요인을 탐색하기 위하여 최근 10년간 비문해자 집단을 구분하는 변수의 변화를 분석한 결과, 연령과 성별의 중요도는 감소한 반면 가구소득의 중요도는 증가한 것으로 나타났다. 사회적 원인으로 인해 남성보다 여성의 비문해 비율이 높게 나타난다는 기존의 연구 결과는 여전히 유효하지만(Rockill, 1993), 적어도 기존만큼 성별의 결정력이 높지는 않다

는 것이다. 2014년 비문해자 집단의 전형이 70세 이상의 여성이라고 한다면, 2023년 비문해자 집단의 전형은 가구소득 100만원 미만의 80세 이상의 노인이다. 일반적으로 가구소득이 기준 중위소득의 50% 이하일 경우 빈곤층으로 분류되며(김진욱, 박창원, 2001), 기초생활을 위한 지원의 대상이 된다. 2025년 기준 1인 가구 중위소득이 2,392,013원이라는 점을 고려할 때(보건복지부 보도자료, 2023. 07. 25), 가구소득이 100만원 미만인 경우 기본적인 생계 유지에 어려움을 겪는 빈곤층이라 할 수 있다. 우리나라의 경우 노인의 빈곤율이 39.7%로 매우 높은 편이며(지표누리, 2023), 노인 인구가 급속하게 늘어나고 있다는 점을 고려할 때(통계청 보도자료, 2024. 09. 26) 노인 빈곤층을 위한 성인문해교육 지원 전략을 적극적으로 모색할 필요가 있다.

한편, 최근 10년간 어떤 변수의 범주에서 문해교육 대상자의 구성이 변화되었는지를 분석한 결과, 연령, 가구소득, 교류활동 빈도, 정치관심도, 경제활동 상태에 따라 2014년과 2023년 문해교육 대상자 수의 분포가 유의미하게 차이가 있는 것으로 나타났다. 2014년과 비교하여 2023년에는 문해교육 대상자 중 70세 이상의 고령자가 차지하는 비율이 현저하게 높아졌다. 가구소득의 경우 2014년에는 소득이 낮을수록 문해교육 대상자의 비율이 높아지는 경향이 뚜렷했으나, 2023년에는 그러한 경향이 둔화되었다. 특이할만한 것은 교류활동 빈도인데, 2014년에는 상대적으로 교류활동이 빈번한 집단이 문해교육 대상자의 비율이 높았으나, 2023년에는 반대로 교류활동이 적은 집단의 문해교육 대상자의 비율이 높은 것으로 나타났다. 이는 최근 10년간 사회 전반적으로 사람들의 교류활동 패턴 변화에 따른 것으로 추정되는데, 특히 팬데믹으로 인해 노인의 사교 모임과 외부 활동이 줄어든 경향이 반영된 것으로 보인다(김성희, 2025; 남궁은하, 2021). 이와 같이 사회적 관계자본이 부족하거나 결핍된 상태 즉, 가족이나 직장 이외에 타인과 유의미한 교류가 없으며 사회적 지지체계가 결핍된 상태를 사회적 고립이라 한다(김성아 외, 2023). 최근 사회적 고립이 노인뿐만 아니라 청년층에서도 빠른 속도로 증가하고 있다는 사실을 고려할 때(김성아 외, 2023), 향후 문해교육 정책의 관심 범위는 청년층까지 확대될 필요가 있다.

문해교육 대상자 결정요인을 탐색하기 위하여 최근 10년간 문해교육 대상자 집단을 구분하는 변수의 변화를 분석한 결과, 가구소득의 중요도는 감소한 반면, 연령과 정치관심도의 중요도는 증가한 것으로 나타났다. 2014년 문해교육 대상자 집단의 전형이 가구소득 200만원 미만의 40대 이상이라고 한다면, 2023년 문해교육 대상자의 전형은 가구소득 200만원 미만으로 정치에 관심을 갖지 않는 70대 이상의 노인이다. 정치관심도는 사회자본과 깊은 관계가 있는데, 김상원(2018)에 따르면 사회자본 중에서 네트워크와 사회응집력이 높은 노인일수록 정치참여도가 높다. 노인의 정치관심도는 곧 사회에 대한 관심과 변화 의지를 반영한다. 성인

문해교육을 통해 노인들이 고립을 깨고 사회와 소통할 수 있는 계기를 마련해주고 비판적으로 참여할 수 있도록 지원해 줄 필요가 있다.

본 연구는 우리나라 성인의 문해능력 수준을 결정하는 요인을 탐색하기 위해 국가승인통계로 지정된 성인문해능력조사의 데이터를 분석에 활용하였다는 점에서 의미가 있다. 또한 시대적 변화를 포착하기 위하여 10년 전과 최근의 성인의 문해능력 수준과 결정요인을 비교 분석하였다는 점에서 기존의 연구와 차별화될 수 있다. 특히 분석 방법으로 비교적 접근성이 높고 직관적으로 이해할 수 있는 의사결정나무 분석을 활용함으로써 문해 연구에 머신러닝의 적용가능성을 확인하였다는 것도 본 연구의 의의라 할 수 있다.

향후 보다 심층적인 분석을 위해 유관 데이터와의 통합 분석을 시도해 볼 필요가 있다. 먼저 본 연구의 분석에 활용된 전통적인 설명 변수들뿐만 아니라 우리 사회의 변화된 모습을 반영하는 디지털 역량, 시민의식 등의 변수를 추가하여 분석한다면 성인의 문해능력 수준을 결정하는 요인 변화를 입체적으로 파악할 수 있을 것이다. 또한 2006년 성인문해교육 지원사업이 추진된 이후 예산, 참여기관, 교사, 학습자 등 사업운영과 관련한 데이터가 누적 관리되고 있다는 점을 고려할 때, 이를 투입변수로 고려하여 성인문해능력 수준과 연계 분석한다면 성인문해정책의 효과를 평가하는 것도 가능할 것이다. OECD의 PIAAC(Programme for the International Assessment of Adult Competencies) 데이터를 활용하여 글로벌 수준에서의 성인 문해능력 수준 및 결정요인을 비교 분석하는 것도 가능한데, 이를 통해 본 연구 결과의 해석이 확장될 수 있을 것으로 기대된다.

이와 더불어 다각적 검증을 위해 보다 다양한 분석 모형을 적용할 필요도 있다. 관심을 가지고 있는 예측 문제 또는 그 예측에서 사용되는 변수들에 대한 탐색을 위해 하나의 모형에만 의존하는 것보다는 다양한 모형을 적용하는 것이 바람직하다. 본 연구의 경우 머신러닝 알고리즘 중 분류 예측을 위한 모형인 로지스틱 회귀(logistic regression), 판별 분석(discriminant analysis), 서포트 벡터 머신(support vector machine) 등의 수학적 기반 모형을 적용할 수도 있다. 또한 랜덤 포레스트(random forest), 배깅(bagging), 극한 변화도 부스팅(extreme gradient boosting, XGBoost)와 같은 강화된 트리 기반의 모형도 적용해 볼 수 있다. 향후 이와 같은 다양한 모형들을 적용하여 공통된 부분과 그렇지 않은 차이점을 함께 살펴본다면 현상을 이해하고 미래를 예측하는데 도움이 될 것으로 보인다.

참고문헌

- 곽삼근, 박세영(2016). 한국 성인문해교육의 전개과정과 리터러시 확장의 과제. **평생교육학연구**, 22(2), 85-108.
- 교육부(2025). **2025년 성인문해교육 지원사업 기본계획**. 교육부 평생학습지원과.
- 국가평생교육진흥원(2015). **2014년 성인문해능력조사**. 국가평생교육진흥원.
- 국가평생교육진흥원(2019). **2017년 성인문해능력조사**. 국가평생교육진흥원.
- 국가평생교육진흥원(2021). **2020년 성인문해능력조사**. 국가평생교육진흥원.
- 국가평생교육진흥원(2024). **2023년 제4차 성인문해능력조사**. 국가평생교육진흥원.
- 김경년(2019). 성인의 문해력 상실과 획득에 미치는 연령과 문해 사용의 영향 탐색. **Andragogy Today: Interdisciplinary Journal of Adult & Continuing Education**, 22(1), 27-52.
- 김명숙, 어재영, 이성엽(2017). 국내 성인문해 연구동향 분석: 학술지 게재 논문을 중심으로. **평생교육학연구**, 23(2), 89-114.
- 김상원(2018). 노인의 사회자본이 정치참여에 미치는 영향: 선거에서의 투표참여를 중심으로. **21세기정치학회보**, 28(3), 97-118.
- 김성아, 김문길, 이덕영, 노혜진, 노현주(2023). **고립·은둔청년 실태조사 연구**. 한국보건사회연구원.
- 김성희(2025). 팬데믹과 노인의 사회적 네트워크: 초기 코로나19 영향을 중심으로. **노인복지연구**, 80(1), 31-59.
- 김수지, 송민희, 노한나, 정제영(2024). 텍스트마이닝을 활용한 성인문해교육 정책 관련 이슈 분석: 정책, 학술지, 언론보도 간 차이를 중심으로. **교육행정학연구**, 42(2), 219-249.
- 김진욱, 박창원(2001). 소득계층의 실태와 특성에 관한 분석. **공공경제**, 6(2), 57-77.
- 남궁은하(2021). 코로나 19로 인한 노인생활의 변화와 정책 과제. **보건복지포럼**, 300, 71-84.
- 박소연, 이지혜, 허준(2014). 성인 기초 문해력 조사문항 개발을 위한 문해 영역 및 수준 설계. **평생교육학연구**, 20(3), 31-55.
- 보건복지부 보도자료(2023. 07. 25). 2025년도 기준 중위소득 6.42%로 역대 최대 인상.
https://www.mohw.go.kr/board.es?mid=a10503010100&bid=0027&act=view&list_no=1482430&tag=&nPage=1에서 2025. 04. 20 인출.
- 서영아(2022). **성인문해교육 지원정책의 변화 고찰**. 고려대학교 HRD 정책연구소.
- 이주호, 최슬기(2015). **한국인의 역량: 실증분석과 미래전략**. 서울: 한국개발연구원.

- 이지혜(2003). 학습사회에서 성인문해의 의미. **평생교육학연구**, 9(3), 57-74.
- 정선영(2024). 아내구타 정당성 부여수준의 연령, 기간, 코호트 효과 분석. **한국가족복지학**, 71(4), 77-105.
- 지표누리(2023). 노인빈곤율.
<https://www.index.go.kr/unity/potal/indicator/IndexInfo.do?popup=Y&clasCd=2&idxCd=5057>에서 2025. 04. 20 인출.
- 통계청 고시 제2014-315호
https://kostat.go.kr/board.es?mid=a10403040000&bid=107&act=view&list_no=330631에서 2025. 04. 20 인출.
- 통계청 보도자료(2024. 09. 26). 2024 고령자통계.
https://kostat.go.kr/board.es?mid=a10301060500&bid=10820&act=view&list_no=432917에서 2025. 04. 20 인출.
- 한상길(2017). 문해의 개념과 문해교육 실천 방향. **Andragogy Today: Interdisciplinary Journal of Adult & Continuing Education**, 20(2), 30-50.
- 한승희(2014). 평생학습과 문해력. **국어교육학회 제57회 전국학술대회 자료집**.
- 허준, 이지혜, 길혜지, 김지현(2023). 성인문해능력조사 측정도구 개선 및 제4차 성인문해능력조사 예비조사. 국가평생교육진흥원.
- 허준, 이지혜, 박소연, 길혜지(2013). 국민 기초 문해력 측정 조사 도구 개발. 국가평생교육진흥원.
- 허준, 이지혜, 박소연, 길혜지(2019). 성인문해능력조사 측정도구 개선 연구. 국가평생교육진흥원.
- 허준, 이지혜, 박소연, 조현주, 고선경(2024). 성인문해능력조사 결과 종합 분석 연구. 국가평생교육진흥원.
- 허준, 이경민, 이진희(2016). 성인 문해교육 연구의 담론 지형. **평생교육학연구**, 22(4), 1-26.
- 현영섭, 신은경(2018). **성인 문해능력 조사 분석 연구 요약본**. 교육부·국가평생교육진흥원.
- 황혜진(2015). OECD 성인역량조사결과에 나타난 세대 간 문해력의 차이. **통일인문학**, 61, 585-612.
- Breiman, L., Friedman, J., Olshen, R. A., & Stone, C. J. (2017). *Classification and regression trees*. Routledge.
- Loh, W.-Y. (2014). Fifty Years of Classification and regression trees. *International statistical review*, 82(3), 329-348.

- Sammut, C. & Webb, G. I. (2017). *Encyclopedia of machine learning and data mining*. Springer.
- Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. MIT Press.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning (2nd ed.)*. Springer.
- Reder, S. (2008). The development of literacy and numeracy in adult life. In *Tracking adult literacy and numeracy skills* (pp. 79-104). Routledge.
- Rockhill, K. (1993). Gender, language and the politics of literacy. In B. Street (Ed.) *Cross-cultural approaches to literacy*. Cambridge University Press.
- Satherley, P., & Lawes, E. (2008). *The adult literacy and life skills (ALL) survey: Gender, ethnicity and literacy*. Comparative Education Research Unit, Research Division, Ministry of Education.
- Willms, J. D., & Murray, T. S. (2007, July). *Gaining and losing literacy skills over the lifecourse*.
- Wylie, C., & Hodgen, E. (2011). *Forming adulthood: Past, present and future in the experiences and views of the Competent Learners@ 20*. Wellington: New Zealand Council for Educational Research.

- 논문 접수 2025. 5. 1. / 게재 승인 2025. 5. 29.
- 박소연: 고려대학교 교육학과를 졸업하고, 같은 대학에서 성인계속교육 전공으로 석사학위 및 박사학위를 취득하였다. 주 관심은 평가 설계, 평가자 교육, 정책평가, HR Analytics 등이다.
- 권예일: 고려대학교 통계학과를 졸업하고, 같은 대학에서 통계학 석사학위 취득 후, 미국 Columbia University에서 Operation Research 전공으로 석사학위를, Temple University에서 Statistics 전공으로 박사학위를 취득하였다. 주 관심은 empirical Bayesian estimation, Bayesian machine learning 모형 등이다.

Abstract

Changes in the determinants of adult literacy level: A decision tree-based comparative analysis of 2014 and 2023

Soyoun, Park(Anyang University)

Yeil, Kwon(Wichita State University)

This study aims to identify the determinants of adult literacy levels and examine how these factors have changed over the past decade. To this end, data from the 2014 and 2023 National Adult Literacy Surveys in Korea were analyzed. The findings reveal a shift in the key variables associated with illiteracy. While age and gender were prominent determinants in 2014, their significance declined by 2023, whereas household income emerged as a more influential factor. Specifically, the typical profile of an illiterate adult changed from a woman aged 70 or older in 2014 to an individual over 80 with a household income below 1 million KRW in 2023. Similarly, the determinants of literacy program participation have also evolved. Over the last ten years, the significance of household income has decreased, while age and political interest have gained importance. The prototypical participant in 2014 was someone over 40 with a household income below 2 million KRW, whereas in 2023, it is a politically disengaged person over 70 with the same income level. Based on these findings, this study discusses implications for supporting adult literacy development and offers future directions for policy and practice.

* Keywords: Literacy proficiency, Adult literacy education, Adults with low literacy, Target group for literacy education, Decision tree