

HLS-EU-Q16 기반 3문항 건강문해력 축약형 선별도구의 구성과 예비평가

이 영 선¹

¹ 조선대학교

초 록

본 연구는 한국의료패널 자료를 활용하여 HLS-EU-Q16 기반 3문항 건강문해력 선별 후보도구를 구성하고 분류 성능을 평가하는 것을 목적으로 수행하였다.

연구 대상은 완전응답 성인 8,358명이며, 문항분포, 문항-총점 상관, 내적 일관성, 요인분석 및 내용적 대표성을 검토하여 세 예비조합을 구성하였다. 원척도 12점 이하를 저건강문해력군으로 정의하고 표본분할 및 민감도 분석을 시행하였다.

연구 결과 예비조합1(HLIT1, HLIT8, HLIT12)은 Cronbach's $\alpha=.787$, 원척도와 의 상관 $r=.902$ 였다. 2점 이하 기준에서 민감도 98.2%, 특이도 78.8%, 정확도 88.3%, 음성예측도 97.8%, 균형정확도 88.5%, AUC 0.965였다. 560개 조합 중 AUC는 2위였고, 최고 조합보다 민감도와 음성예측도가 높았다. 검증표본에서도 민감도 98.3%, 음성예측도 98.0%였다.

예비조합1은 저건강문해력 위험군을 신속히 포착하기 위한 3문항 선별 후보도구로 활용 가능성이 있다.

주요 용어: 건강문해력, HLS-EU-Q16, 축약형 도구, 선별도구, 한국의료패널, 분류 성능

알기 쉬운 요약

이 연구는 왜 했을까? 건강문해력은 건강정보를 찾고 이해하여 건강관리와 의사결정에 활용하는 능력이다. 기존 HLS-EU-Q16은 16문항이어서 대규모 조사나 지역사회 현장에서 부담이 될 수 있다. 이 연구는 한국의료패널 자료로 3문항만으로 건강문해력이 낮은 가능성이 있는 사람을 빠르게 선별할 수 있는지 살펴보았다.

새롭게 밝혀진 내용은? 한국의료패널 성인 8,358명을 분석한 결과, 3문항 후보도구는 건강문해력이 낮은 사람의 98.2%를 찾아냈다. 표본을 나누어 다시 확인했을 때에도 건강문해력이 낮은 사람의 98.3%를 찾아내는 등 비슷한 성능이 나타났다.

앞으로 무엇을 해야 하나? 이 도구는 점수로 건강문해력 수준을 측정하는 용도보다 도움이 더 필요한 사람을 찾는 1차 선별에 활용할 수 있다. 앞으로 다른 대상과 실제 현장에서 같은 성능이 나타나는지 확인하고, 건강행동이나 의료이용과의 관련성도 추가로 검증할 필요가 있다.

■ 투 고 일: 2026. 02. 13.

■ 수 정 일: 2026. 07. 13.

■ 게재확정일: 2026. 08. 27.

I. 서론

건강문해력(health literacy)은 개인이 건강 관련 정보를 단순히 읽고 이해하는 수준을 넘어, 필요한 정보를 탐색하고 그 의미를 평가하며, 이를 실제 건강행위와 의사결정에 활용할 수 있는 역량을 의미한다(Nutbeam, 2000; Sørensen et al., 2012). 이러한 건강문해력은 건강행위 선택, 만성질환 관리, 의료서비스 이용, 예방적 건강행동 전반에 영향을 미치는 핵심 요인으로 보고되어 왔으며, 개인의 건강 수준과 삶의 질, 의료비 지출과도 밀접한 관련이 있다(Berkman et al., 2011). 또한 교육수준, 소득, 연령과 같은 사회경제적 요인과 결합하여 건강 불평등을 심화시키는 매개 요인으로 작용할 수 있다는 점에서 공중보건 및 보건정책 영역에서 중요한 연구 주제로 다루어지고 있다(Nutbeam, 2008). 세계보건기구(WHO)도 건강문해력을 21세기 보건정책의 핵심 요소로 제시하면서, 낮은 건강문해력이 예방서비스 이용 저하, 약물 오남용, 만성질환 관리 실패, 의료비 증가 등 다양한 부정적 건강결과와 연결될 수 있음을 지적하였다(Kickbusch et al., 2013).

이러한 문제의식 속에서 건강문해력을 체계적으로 측정하기 위한 다양한 도구가 개발되어 왔다. 대표적으로 HLS-EU Consortium은 건강문해력을 의료, 질병예방, 건강증진의 세 영역과 정보의 접근, 이해, 평가, 적용의 네 가지 인지 과정을 결합한 개념틀로 정의하고, 이를 바탕으로 European Health Literacy Survey Questionnaire(HLS-EU-Q47)를 개발하였다(Sørensen et al., 2013). 이후 응답 부담을 줄이기 위해 HLS-EU-Q16, Q6 등 다양한 축약형 도구가 제안되었으며(Pelikan & Ganahl, 2017), 이 가운데 HLS-EU-Q16은 원칙도의 개념 구조를 비교적 충실히 반영하면서도 현장 적용성을 높인 도구로 평가되어 왔다(Storms et al., 2017; Rouquette et al., 2018).

그러나 대규모 인구조사나 지역사회 선별 현장에서는 16문항조차도 응답 부담으로 작용할 수 있으며, 간결한 선별도구의 필요성이 지속적으로 제기되어 왔다. 실제로 Chew 외(2004)는 임상 현장에서 1분 이내로 시행 가능한 3문항 건강문해력 선별질문을 제안하고, 소수 문항만으로도 건강문해력이 낮은 대상자를 비교적 높은 정확도로 구분할 수 있음을 보고하였다. Haun 등(2014) 역시 건강문해력 측정도구 51종을 검토한 결과, 연구 목적과 활용 맥락에 따라 문항 수를 축소한 도구가 측정 효율성과 현장 적용성을 높일 수 있음을 제시하였다. 이는 건강문해력의 전 영역을 정밀하게 측정하는 목적과, 제한된 자원과 시간 안에서 위험군을 신속히 선별하는 목적이 서로 구분될 필요가 있음을 보여준다.

국내에서도 건강문해력에 대한 관심이 높아지면서 관련 연구가 꾸준히 축적되어 왔다. 선행연구에 따르면 한국 성인 인구에서도 교육수준, 소득, 연령에 따라 건강문해력 격차가 나타나며, 건강불평등과 연계된 중요한 요인으로 논의되고 있다(강수진, 이미숙, 2015; 배재용, 김혜윤, 2023; 천희란 외, 2022). 특히 한국의료패널 2021년 부가조사 분석 결과, 성인 인구의 상당수가 부족 또는 경계 수준의 건강문해력에 해당하는 것으로 보고된 바 있다(배재용, 김혜윤, 2023).

그럼에도 불구하고 국내에서는 대규모 대표자료를 활용하여 건강문해력 축약형 도구의 분류 성능과 현장 적용 가능성을 검토한 연구가 제한적이며, 본 연구는 이러한 공백에 주목하였다. 특히 한국의료패널은 전국 단위 대표 표본을 기반으로 건강상태, 의료이용, 사회경제적 특성을 포괄적으로 수집하는 자료이며, 2021년 조사에는 HLS-EU-Q16 전 문항이 포함되어 있다. 이는 실제 인구집단의 응답을 바탕으로 문항 특성과 요인구조를 검토하고, 축약형 건강문해력 선별 후보도구의 원칙도 기준 분류 성능을 평가할 수 있는 자료를 제공한다. 이에 본 연구는 한국의료패널 2021년 자료를 활용하여 HLS-EU-Q16의 문항 특성을 검토하고, 통계적 적합성과 내용적 대표성을 함께 고려한 3문항 건강문해력 축약형 선별 후보도구를 구성하였다. 이어 원칙도 기준 저건강문해력군 분류와의 일치성, 절단점별 성능, 표본분할 검증 및 민감도 분석, 구성 문항 제외 보조준거 분석을 통해 후보도구의 예비적 활용 가능성을 평가하였다. 이를 통해 응답 부담을 줄이면서도 저건강문해력 위험군을 신속히 식별할

수 있는 실용적 대안을 제시하고, 향후 지역사회 및 보건의료 현장에서의 활용 가능성에 대한 기초근거를 제공하고자 한다.

II. 이론적 배경

1. 건강문해력의 개념과 측정

건강문해력은 초기에는 의료정보의 읽기 능력이나 수리 능력과 같은 기능적 문해력의 관점에서 이해되었으나, 이후 건강정보의 비판적 해석과 실제 활용 역량까지 포괄하는 다차원적 개념으로 확장되어 왔다. 특히 Nutbeam(2000)은 건강문해력을 기능적, 상호작용적, 비판적 문해력으로 구분하였고, Sørensen 등(2012)은 이를 의료, 질병예방, 건강증진의 세 영역과 정보의 접근, 이해, 평가, 적용의 네 가지 인지 과정으로 통합하여 개념화하였다.

건강문해력 측정 도구는 이러한 개념 확장에 따라 다양하게 발전해 왔다. 초기의 REALM(Davis et al., 1993)과 TOFHLA(Parker et al., 1995)는 의료 관련 어휘 이해나 기능적 문해력을 중심으로 구성되어 임상 현장에서 널리 활용되었으나, 건강정보의 평가와 적용 역량을 충분히 반영하지 못한다는 한계가 지적되었다. 이에 따라 유럽연합은 건강문해력의 다차원적 구조를 반영한 HLS-EU-Q47을 개발하였으며(Sørensen et al., 2013), 이후 HLS-EU-Q16, Q6과 같은 축약형 도구가 제안되었다(Pelikan & Ganahl, 2017). 이 가운데 HLS-EU-Q16은 원척도의 핵심 구조를 비교적 충실히 유지하면서도 적용 가능성을 높인 도구로 평가되어 여러 국가에서 번역, 적용 및 타당도 검토가 이루어져 왔다(Storms et al., 2017; Rouquette et al., 2018; Gustafsdottir et al., 2020).

국내에서도 건강문해력에 대한 연구가 증가하면서 다양한 측정 도구의 적용과 검토가 이루어져 왔다. 강수진과 이미숙(2015)은 국내 건강문해력 연구 동향을 검토하며 측정도구의 타당도와 대상자 특성에 따른 검증의 필요성을 제기하였다. 천희란 외(2022)는 국내 건강문해력 측정도구 연구를 주제범위 문헌고찰로 분석하면서, 도구 개발과 타당화 연구가 증가하고 있으나 대표 표본, 조사시간, 적용 대상 등 도구 활용과 관련된 정보가 더 체계적으로 제시될 필요가 있다고 지적한 바 있다.

2. 건강문해력 축약형 선별 도구

건강문해력 측정도구의 축약은 단순히 문항 수를 줄이는 문제가 아니라, 측정 목적에 맞는 정보 손실의 범위를 관리하는 과정이라 할 수 있다. 전면적 측정을 목표로 하는 경우에는 건강문해력의 다차원성을 충분히 반영하는 문항 구성이 요구되지만, 선별을 목적으로 하는 경우에는 짧은 시간 안에 저건강문해력 위험군을 식별할 수 있는 간결성과 적용 용이성이 중요하다.

기존 연구에서 건강문해력 축약형 도구는 대체로 두 가지 방식으로 개발되어 왔다. 첫째는 기존 도구의 일부 문항을 선정하여 원척도의 개념 구조를 유지하면서 문항 수를 줄이는 방식이다. HLS-EU-Q16과 Q6이 대표적이며(Pelikan & Ganahl, 2017), 이후 여러 국가에서 번역과 문화적 적응을 거쳐 타당성이 검토되어 왔다(Rouquette et al., 2018; Gustafsdottir et al., 2020). 이 경우 원척도의 개념 구조를 어느 정도 유지하면서 응답 부담을 줄이는 데 초점을 두고 있다. 둘째는 기존 개념틀과 표준 도구를 참고하되, 특정 목적과 활용 맥락에 맞게 소수 문항을 재구성하는 방식이다. Chew 등(2004)의 3문항 건강문해력 선별질문은 이러한 접근의 대표적 사례로,

임상 환경에서 짧은 시간 내에 건강문해력 취약 환자를 구분하기 위한 실용적 목적에 초점을 두고 개발되었다.

이러한 두 접근 모두 축약형 도구를 구성할 때에는 문항의 통계적 특성과 내용적 대표성을 함께 고려할 필요가 있다. 선행연구에서는 문항반응이론이나 측정 동일성 분석을 통해 문항 적합성과 집단 간 측정 안정성을 검토하거나(Rouquette et al., 2018), 요인분석, 문항-총점 상관, 내적 일관성, 응답 분포, 분류 성능 등을 활용하여 축약형 도구의 문항 구성을 평가하였다(Storms et al., 2017; Gustafsdottir et al., 2020). 특정 통계지표가 높은 문항만을 중심으로 도구를 구성할 경우 일부 영역에 문항이 편중될 수 있으며, 반대로 내용적 대표성만을 강조할 경우 선별 목적에 필요한 분류 성능 검토가 부족할 수 있다.

측정도구 개발 방법론에서도 문항 개발, 문항 평가, 신뢰도, 타당도, 요인구조, 척도 길이의 적정성을 종합적으로 검토해야 한다고 강조한다(DeVellis & Thorpe, 2021; Streiner et al., 2015). 특히 소수 문항으로 구성된 선별 도구는 원척도의 전체 개념 구조를 충분히 대체하기 어렵기 때문에, 해당 도구의 사용 목적과 해석 범위를 명확히 규정할 필요가 있다. 또한 분류도구의 성능을 평가할 때에는 민감도, 특이도, 양성예측도, 음성예측도 등 주요 지표를 제시하고, 동일 자료에서 문항을 선정하고 성능을 평가하는 경우에는 성능이 낙관적으로 추정될 가능성이 있으므로, 표본분할 등 내적 검증 절차를 통해 결과의 안정성을 점검할 필요가 있다(Steyerberg, 2019).

단축형 선별도구에서 적절한 문항 수에 대한 일치된 기준은 존재하지 않는다. 다만 선별을 주된 목적으로 할 경우, 건강문해력의 모든 하위 차원을 포괄적으로 측정하기보다는 제한된 시간 안에서 저건강문해력 위험 가능성이 있는 대상을 파악하는 데 필요한 문항을 선택하는 방식이 고려될 수 있다(Chew et al., 2004; Haun et al., 2014).

국내에서는 건강문해력 측정 도구의 연구 동향과 개념적 검토는 이루어져 왔으나(강수진, 이미숙, 2015; 천희란, 김수현, 박은자, 2022), 한국적 문화, 언어적 맥락과 대규모 대표자료를 바탕으로 3문항 수준의 초단축 선별도구를 구성하고 분류 성능을 평가한 연구는 제한적이다.

III. 연구 방법

1. 연구 설계

본 연구는 제2기 한국의료패널 2021년 부가조사를 활용하여 HLS-EU-Q16 기반 3문항 건강문해력 축약형 선별도구를 구성하고, 그 내적 일관성과 원척도 대비 분류 성능 및 절단점에 따른 성능 차이를 예비적으로 평가한 이차자료 기반의 계량적 연구이다. 이를 위해 전체 HLS-EU-Q16 문항의 응답 특성과 요인 구조, 문항-총점 상관성, 내적 일관성, 응답 분포의 변별력, 요인적재량 및 공통성, HLS-EU 개념틀에 따른 이론적 대표성을 종합적으로 고려하여 3문항 3개의 예비 조합을 구성하였다. 이후 각 예비 축약형 도구의 분류 결과를 원척도의 분류 기준과 비교하여 민감도, 특이도, 예측도, 우도비, 균형정확도(Balanced accuracy) 및 ROC/AUC를 산출하였다.

세 가지 예비조합을 구성하여 평가한 후, 추가 확인분석으로 가능한 모든 3문항 조합 560개의 ROC/AUC를 산출하여 비교하였다. 또한 문항선정과 성능평가가 동일 자료에서 이루어지는 데 따른 과적합 가능성과, 원척도 문항을 내부 준거로 사용하는 데 따른 부분-전체 중복 가능성을 검토하기 위해 표본분할 검증과 13문항 보조준거 분석을 실시하였다. 가중치 적용과 ‘잘 모르겠다’ 응답 처리 방식에 따른 민감도 분석을 통해 결과의 안정성을 확인하였다.

2. 연구 대상 및 자료

본 연구는 제2기 한국의료패널 2021년 연간자료와 건강문해력 부가조사를 활용하였다. 건강문해력 부가조사에 응답한 만 19세 이상 성인 중 건강보험 가입자와 의료급여 수급자는 총 9,714명이었으며, 이 중 HLS-EU-Q16 16개 문항 중 1개 이상 '잘 모르겠다'로 응답한 1,356명을 제외하였다. 이에 따라 본 연구의 주분석은 16개 문항 모두에 1~4점의 유효 응답을 제공한 성인 8,358명을 대상으로 수행하였다.

3. 측정 도구

건강문해력은 한국의료패널 건강문해력 부가조사에 포함된 HLS-EU-Q16을 사용하여 측정하였다 (Pelikan & Ganahl, 2017; 배재용, 김혜운, 2023). HLS-EU-Q16은 의료, 질병예방, 건강증진 영역에서 건강정보를 탐색하고, 이해하며, 평가하고, 적용하는 역량을 포괄적으로 측정하는 16개 문항으로 구성되어 있다. 각 문항은 '매우 어렵다', '어려운 편이다', '쉬운 편이다', '매우 쉽다', '잘 모르겠다' 중 하나로 응답하도록 구성되어 있으며, 본 연구는 HLS-EU-Q16의 기존 점수화 방식에 따라 주분석에서 '잘 모르겠다' 응답이 하나라도 있는 사례를 제외하였다(HLS-EU Consortium, 2012). 다만 '잘 모르겠다'가 건강정보 이해의 어려움이나 판단의 불확실성을 반영할 가능성을 고려하여, 해당 응답을 0점으로 처리한 민감도 분석을 실시하였다.

원칙도 분류 기준을 적용하기 위해 각 문항 응답은 이진화하였다. 즉 '매우 어렵다'와 '어려운 편이다'는 0점, '쉬운 편이다'와 '매우 쉽다'는 1점으로 변환하였다. 이진화된 16개 문항의 합산점수는 0점에서 16점 범위이며, HLS-EU-Q16의 기존 분류 기준에 따라 0~8점은 부족, 9~12점은 문제, 13~16점은 충분 수준으로 구분된다 (HLS-EU Consortium, 2012; Sørensen et al., 2013). 국내에서도 한국의료패널 자료를 활용한 연구에서 이러한 분류 기준에 근거한 건강문해력 수준 해석이 적용된 바 있다(배재용, 김혜운, 2023). 본 연구는 원칙도의 3수준을 정밀하게 대체하는 축약형 도구 개발보다는, 대규모 조사 및 지역사회 현장에서 저건강문해력 위험군을 1차적으로 넓게 포착하기 위한 선별 후보도구 구성에 초점을 두었으므로, 부족 및 문제 수준을 통합하여 총점 12점 이하를 저건강문해력군으로 정의하였다.

한편, 축약형 도구 구성 단계에서는 분석 목적에 따라 원자료와 이진화 자료를 구분하여 사용하였다. 탐색적 요인분석에는 원자료의 4점 Likert 응답을 사용하였고, 문항별 분포, 문항-총점 상관, 내적 일관성, 축약형 합산점수 및 원칙도 대비 분류 성능 평가에는 이진화 자료를 사용하였다.

4. 문항선정 및 분석절차

본 연구는 HLS-EU-Q16 기반 3문항 축약형 건강문해력 선별 후보도구를 구성하기 위해 통계적 적합성과 내용적 대표성을 함께 고려하는 단계적 분석 절차를 적용하였다.

첫째, HLS-EU-Q16 각 문항의 응답분포를 확인하였다. 각 문항은 '매우 어렵다'와 '어려운 편이다'를 0점, '쉬운 편이다'와 '매우 쉽다'를 1점으로 이진화하였으며, '잘 모르겠다' 응답이 하나라도 있는 사례는 주분석에서 제외하였다. 이진화된 문항별 평균, 표준편차, 응답 비율을 산출하여 문항의 난이도와 변별 가능성을 검토하였다. 특히 '쉬움' 응답 비율이 90% 이상인 문항은 저건강문해력 선별에서 변별력이 제한될 가능성이 있는 문항으로 검토하였다.

둘째, 전체 16문항의 내적 일관성을 확인하기 위해 Cronbach's α 를 산출하고, corrected item-total correlation과 문항 제거 시 Cronbach's α 변화를 검토하였다. 이는 축약형 도구 구성을 위한 문항 후보군을 선별하기 위한

보조 기준으로 활용하였다. 다만 본 연구의 목적은 전체 척도의 신뢰도를 극대화하는 것이 아니라 선별 목적에 적합한 소수 문항 조합을 구성하는 것이므로, 내적 일관성 지표만으로 최종 문항을 결정하지는 않았다.

셋째, 문항 간 경험적 구조를 확인하기 위해 탐색적 요인분석을 실시하였다. 분석에는 HLS-EU-Q16의 원자료인 4점 Likert 응답을 사용하였으며, 요인 추출은 주축요인추출법(principal axis factoring), 회전은 Oblimin 사각회전을 적용하였다. 요인분석에 앞서 KMO와 Bartlett 구형성 검정을 통해 요인분석 적합성을 확인하였다.

넷째, 문항선정은 문항분포, 문항-총점 상관, 내적 일관성 기여도, 요인적재량 및 공통성, 그리고 HLS-EU 개념틀에 따른 내용적 대표성을 종합하여 수행하였다. 수정 문항-총점 상관 .30 이상을 기본 기준으로 삼되, 선별 후보 문항에서는 .60 이상 문항을 우선적으로 검토하였다. 요인분석 결과는 축약형 도구의 최종 요인구조를 확정하기 위한 목적이 아니라, 문항들이 반영하는 경험적 차원과 문항 간 중복 가능성을 파악하기 위한 보조 근거로 활용하였으며, 요인적재량 .40 이상 및 공통성 .40 이상을 참고 기준으로 삼았다. 다만 최종 문항선정은 특정 통계지표만으로 결정하지 않고, 건강정보 접근, 이해, 평가, 적용이라는 HLS-EU 개념틀의 핵심 기능과 문항 간 내용 중복 가능성을 함께 고려하였다.

다섯째, 상기 기준을 바탕으로 세 가지 3문항 예비 조합을 구성하였다. 첫 번째 후보는 경험적 성능과 건강정보 탐색과 활용 기능을 중시한 주 후보조합, 두 번째 후보는 의료 의사소통 및 예방서비스 이해 중심의 비교조합, 세 번째 후보는 의료 접근과 판단 및 건강증진 정보 이해를 포함한 대안 조합으로 설정하였다. 각 예비 도구는 3개 문항의 이진화 점수를 합산하여 0~3점 범위의 총점을 산출하였다. 원칙도 기준 저건강문해력군은 HLS-EU-Q16 이진화 합산점수 12점 이하로 정의하였다. 각 축약형 도구의 분류 성능은 원칙도 분류 결과와 비교하여 진양성(true positive, TP), 위음성(false negative, FN), 위양성(false positive, FP), 진음성(true negative, TN)을 산출하고, 민감도, 특이도, 정확도, 양성예측도, 음성예측도, 균형정확도, 양성우도비, 음성우도비를 계산하였다. 또한 3문항 합산점수의 전체 절단점에 대한 판별력을 평가하기 위해 ROC 분석을 실시하고 AUC를 산출하였다. 주요 분류 성능 지표와 AUC에 대해서는 95% 신뢰구간을 함께 제시하였다. 이후 추가 확인분석으로 HLS-EU-Q16의 16개 문항에서 가능한 모든 3문항 조합 560개의 분류 성능을 검토하여 최종 후보조합 선택의 안정성을 확인하였다.

여섯째, 우선 후보안으로 도출된 예비조합1에 대해 절단점별 성능을 비교하였다. 예비조합1 총점 0점 이하, 1점 이하, 2점 이하, 3점 이하를 각각 저건강문해력군 분류 기준으로 적용하고, 각 기준의 민감도, 특이도, 정확도, 양성예측도, 음성예측도, 균형정확도, 양성우도비, 음성우도비를 비교하였다. ROC 분석에서는 점수가 높을수록 저건강문해력 위험이 높아지도록 3문항 합산점수를 역코딩한 위험점수를 사용하였다. 선별도구의 목적을 고려하여 저건강문해력군 누락을 줄이는 민감도와 음성예측도를 중점적으로 검토하되, 과도한 위양성 가능성을 확인하기 위해 특이도와 양성예측도도 함께 검토하였다.

일곱째, 동일 자료에서 문항을 선정하고 성능을 평가함으로써 발생할 수 있는 과적합 및 낙관적 성능 추정 가능성을 완화하기 위해 표본분할 검증을 실시하였다. 전체 분석대상자 8,358명을 원칙도 기준 저건강문해력 여부에 따라 층화한 뒤, 70% 개발표본과 30% 검증표본으로 무작위 분할하였다. 개발표본에서 산출된 예비조합1의 성능을 확인한 후, 동일한 문항 조합과 절단점을 검증표본에 적용하여 분류 성능이 유지되는지 평가하였다. 이는 외부 독립표본 검증을 대체하는 것은 아니지만, 동일 자료 내에서 도구 성능의 낙관적 추정 가능성을 점감하기 위한 내적 검증 절차로 활용하였다(Steyerberg, 2019).

여덟째, 예비조합1 구성 문항이 원칙도 점수에 포함됨으로써 발생할 수 있는 부분-전체 중복 가능성을 검토하기 위해 보조준거 분석을 수행하였다. 예비조합1의 구성 문항인 HLIT1, HLIT8, HLIT12를 제외한 나머지 13문항의 합산점수를 산출한 뒤, 이를 16점 만점으로 환산하였다. 환산점수 12점 이하를 저건강문해력군으로 정의하고, 예비조합1의 분류 성능을 재평가하였다.

아홉째, 한국의료패널 자료의 표본설계를 고려하여 가중치 적용 민감도 분석을 실시하였다. 본 연구의 주목적은 축약형 도구의 문항 수준 분류 성능을 평가하는 것이므로 주분석은 HLS-EU-Q16 완전응답자 기반 비가중 분석으로 수행하였다. 다만 대표성 추정과 관련된 해석의 안정성을 확인하기 위해 가구원 횡단 모집단 가중치(L_WGC)를 적용한 민감도 분석을 추가하였다. 가중치 분석 결과는 주분석 결과와 비교하여 성능 지표의 방향성이 유지되는지 확인하는 보조자료로 활용하였다.

마지막으로, 4점 Likert 응답을 이분화하는 과정에서 발생할 수 있는 정보손실 가능성을 검토하기 위해, 최종 후보조합의 이분화 점수 기반 합산점수와 4점 원점수 기반 합산점수의 ROC/AUC를 비교하였다. 아울러 ‘잘 모르겠다’ 응답 처리 방식에 따른 결과의 안정성을 확인하기 위해 해당 응답을 0점, 즉 ‘어려움’으로 처리한 민감도 분석을 추가로 실시하였다.

모든 통계분석은 SPSS 28.0을 사용하였다.

IV. 연구 결과

1. 연구대상자 및 HLS-EU-Q16 문항 특성

<표 1>은 HLS-EU-Q16 문항별 이진화 응답 분포를 제시한 것이다. 본 연구의 최종 분석 대상자는 HLS-EU-Q16 16개 문항에 모두 유효 응답한 성인 8,358명이었다. 성별은 남성 3,686명(44.1%), 여성 4,672명(55.9%)이었다. 연령대는 60대가 2,005명(24.0%)으로 가장 많았고, 다음으로 70대 1,736명(20.8%), 50대 1,405명(16.8%) 순이었다. 교육수준은 대학교 이상이 2,638명(31.6%), 고등학교가 2,520명(30.2%)이었다. 결혼 상태는 기혼이 5,865명(70.2%)으로 가장 많았고, 경제활동 상태는 임금근로자 3,397명(40.6%), 비경제활동 인구 2,859명(34.2%)이었다. 건강보험 유형은 지역가입자 3,263명(39.0%), 직장가입자 2,459명(29.4%) 순이었다.

문항별 평균은 0.529(HLIT5)에서 0.935(HLIT4)까지 분포하였다. ‘이해 쉬움’ 응답 비율은 HLIT5(52.9%), HLIT12(56.6%), HLIT8(57.8%), HLIT1(57.9%), HLIT11(58.2%)에서 60% 미만이었다. 반면 HLIT4(93.5%), HLIT7(92.3%), HLIT10(91.6%), HLIT14(90.3%)는 ‘이해 쉬움’ 응답 비율이 90% 이상이었다. 이러한 문항별 응답 분포는 축약형 도구 구성 시 문항의 변별 가능성을 검토하기 위한 기초자료로 활용하였다.

표 1. HLS-EU-Q16 문항별 기술통계(이진화 처리)

(N=8,358)					
문항	문항 내용(요약)	평균	표준편차	0(이해 어려움) 빈도(%)	1(이해 쉬움) 빈도(%)
HLIT1	1. 걱정되는 질병의 치료에 관한 정보를 찾는 것	0.579	0.494	3518 (42.1)	4840 (57.9)
HLIT2	2. 아플 때 전문적인 도움을 어디에서 받을 수 있는 지를 알아내는 것	0.630	0.483	3089 (37.0)	5269 (63.0)
HLIT3	3. 의사가 내게 말한 것을 이해하는 것	0.867	0.340	1114 (13.3)	7244 (86.7)
HLIT4	4. 처방된 약의 복용 방법에 대한 의사나 약사의 설명을 이해하는 것	0.935	0.246	542 (6.5)	7816 (93.5)
HLIT5	5. 의사에게 진료를 받은 후 추가로 다른 의사의 진료를 받을 필요가 있는지 판단하는 것	0.529	0.499	3938 (47.1)	4420 (52.9)
HLIT6	6. 내 질병치료에 관한 의사결정을 할 때 의사로부터 얻은 정보를 활용하는 것	0.621	0.485	3168 (37.9)	5190 (62.1)
HLIT7	7. 의사나 약사가 말한 지시를 따르는 것	0.923	0.267	643 (7.7)	7715 (92.3)

문항	문항 내용(요약)	평균	표준편차	0(이해 어려움) 빈도(%)	1(이해 쉬움) 빈도(%)
HLIT8	8. 스트레스나 우울과 같은 정신건강 문제를 관리하는 방법에 관한 정보를 찾는 것	0.578	0.494	3524 (42.2)	4834 (57.8)
HLIT9	9. 흡연, 운동부족, 과음과 같은 행동에 대한 건강 위험 경고를 이해하는 것	0.800	0.400	1670 (20.0)	6688 (80.0)
HLIT10	10. 나에게 왜 건강검진이 필요한 지를 이해하는 것	0.916	0.277	699 (8.4)	7659 (91.6)
HLIT11	11. 미디어에서 얻은 건강 위험에 대한 정보가 믿을만한지 판단하는 것	0.582	0.493	3490 (41.8)	4868 (58.2)
HLIT12	12. 미디어에서 얻은 정보에 따라 질병으로부터 나를 보호하는 방법을 결정하는 것	0.566	0.496	3630 (43.4)	4728 (56.6)
HLIT13	13. 나의 정신건강에 도움이 되는 활동을 알아내는 것	0.650	0.477	2922 (35.0)	5436 (65.0)
HLIT14	14. 건강에 대한 가족이나 친구의 조언을 이해하는 것	0.903	0.296	811 (9.7)	7547 (90.3)
HLIT15	15. 어떻게 하면 더 건강할 수 있는지 미디어가 제공하는 정보를 이해하는 것	0.731	0.443	2247 (26.9)	6111 (73.1)
HLIT16	16. 나의 일상적 행동이 내 건강과 어떤 관련이 있는지 판단하는 것	0.684	0.465	2640 (31.6)	5718 (68.4)

주: 0=매우 어렵다/어려운 편이다, 1=쉬운 편이다/매우 쉽다.

2. HLS-EU-Q16의 내적 일관성과 요인구조

<표 2>는 HLS-EU-Q16의 내적 일관성 분석 결과를 제시한 것이다. HLS-EU-Q16 전체 16문항의 Cronbach's α 는 0.906이었다. 개별 문항-총점 상관계수는 0.282~0.707 범위였으며, 대부분의 문항이 0.3 이상을 나타내어 각 문항이 전체 척도와 양호한 상관성을 갖는 것으로 확인되었다. HLIT7은 0.282로 상대적으로 낮았으나 문항 제거 시 α 의 변화 폭이 미미하여 전체 도구의 신뢰도에 크게 영향을 주지는 않은 것으로 보인다.

표 2. HLS-EU-Q16 내적 일관성 분석 결과

문항	문항-총점 상관	문항 제거 시 Cronbach's α
HLIT1	0.687	0.896
HLIT2	0.627	0.898
HLIT3	0.421	0.905
HLIT4	0.321	0.907
HLIT5	0.642	0.898
HLIT6	0.700	0.896
HLIT7	0.282	0.907
HLIT8	0.707	0.895
HLIT9	0.581	0.900
HLIT10	0.417	0.905
HLIT11	0.622	0.899
HLIT12	0.705	0.895
HLIT13	0.701	0.896
HLIT14	0.413	0.905
HLIT15	0.663	0.897
HLIT16	0.642	0.898

주: 전체 16문항의 Cronbach's α =.906. 문항-총점 상관은 해당 문항을 제외한 합산점수와의 corrected item-total correlation이다.

<표 3>에서 탐색적 요인분석에 앞서 자료의 요인분석 적합성을 확인한 결과, KMO 값은 0.952로 매우 높았으며, Bartlett 구형성 검정도 유의하였다($\chi^2(120)=86,391.28$, $p<.001$). 요인 수 결정은 고유값(eigenvalue) 1.0 이상 기준과 스크리플롯을 종합적으로 고려하여 이루어졌다. 분석 결과, 두 개의 요인이 고유값 1 이상을 충족하였으며, 각 요인의 고유값은 요인 1이 8.46, 요인 2가 1.51로 나타났다. 스크리플롯에서도 두 번째 요인 이후 기울기의 완만한 감소가 확인되어 2요인 구조가 적절한 것으로 판단하였다.

회전 후 분산 설명력은 요인 1이 41.0%, 요인 2가 18.9%로 나타났으며, 두 요인의 누적 설명력은 59.9%였다. 두 요인 간 상관관계수는 0.593으로 나타나 중등도 이상의 상관을 보였으며, 이는 두 요인이 완전히 독립적인 차원이라기보다는 건강문해력이라는 공통된 잠재 개념 하에서 서로 밀접하게 연결되어 있음을 의미한다.

각 문항의 요인 적재량을 살펴보면, 요인 1에는 건강정보의 탐색, 평가 및 활용과 관련된 문항들이, 요인 2에는 의료인의 설명 이해, 약물 복용 방법 이해, 건강검진 필요성 이해 등 의료·대면 의사소통 상황의 이해와 관련된 문항들이 주로 적재되었다. 문항별 공통성은 대체로 0.50 이상이었으나, HLIT9(0.352)와 HLIT14(0.420)는 상대적으로 낮았다. HLIT9의 경우 요인 1과 요인 2에 유사한 수준으로 적재되었으며(0.446, 0.392), 공통성은 0.352로 상대적으로 낮게 나타났다. 이는 해당 문항이 특정 요인에 명확히 귀속되지 않고 교차적재 경향을 보일 가능성을 시사한다.

표 3. HLS-EU-Q16 탐색적 요인분석 결과

문항코드	요인1 적재량	요인2 적재량	공통성
HLIT1	0.843	0.112	0.724
HLIT2	0.792	0.185	0.663
HLIT3	0.214	0.692	0.525
HLIT4	0.087	0.853	0.741
HLIT5	0.810	0.161	0.682
HLIT6	0.819	0.193	0.701
HLIT7	0.126	0.791	0.646
HLIT8	0.836	0.148	0.719
HLIT9	0.446	0.392	0.352
HLIT10	0.201	0.761	0.621
HLIT11	0.788	0.164	0.647
HLIT12	0.847	0.109	0.732
HLIT13	0.828	0.177	0.708
HLIT14	0.231	0.633	0.420
HLIT15	0.668	0.187	0.511
HLIT16	0.701	0.214	0.556

주: 추출방법은 주축요인추출법, 회전방법은 Oblimin 사각회전이다. KMO=0.952, Bartlett 구형성 검정 $\chi^2(120)=86,391.28$, $p<.001$. 요인 간 상관 $\rho=.593$. 고유값은 요인1=8.46, 요인2=1.51이며, 누적 설명력은 약 59.9%이다.

3. 3문항 예비 축약형 도구 구성 및 분류 성능

<표 4>는 세 가지 3문항 예비 축약형 도구의 구성 문항, 신뢰도, 원척도와의 상관을 제시한 것이다. 세 예비조합은 단일 통계지표에 따른 성능 상위 조합이 아니라, 문항분포, 문항-총점 상관, 내적 일관성, 탐색적 요인분석 결과와 HLS-EU 개념틀에 따른 내용적 대표성을 종합적으로 고려하여 구성하였다.

세부 문항 검토 결과는 <부표 1>에 제시하였다. 예비조합1은 HLIT1, HLIT8, HLIT12로 구성하였고, 예비조합2는 HLIT3, HLIT4, HLIT10으로 구성하였다. 예비조합3은 HLIT2, HLIT5, HLIT15로 구성하였다.

신뢰도 분석 결과, 예비조합1의 Cronbach's α 는 0.787로 나타나 3문항 도구임을 고려할 때 수용 가능한 수준의 내적 일관성을 보였다. 예비조합3은 0.694로 비교적 양호한 수준이었으며, 예비조합2는 0.575로 상대적으로 낮은 수준을 보였다. 원척도와의 상관 분석에서는 예비조합1(0.902)과 예비조합3(0.894)이 매우 높은 상관을 보였고, 예비조합2는 0.602로 중등도 수준의 상관을 나타냈다(표 4).

표 4. 3문항 예비조합의 구성 문항 및 신뢰도

도구명	구성 문항	Cronbach's α	원척도와 상관(r)
예비조합1	HLIT1,8,12	0.787	0.902
예비조합2	HLIT3,4,10	0.575	0.602
예비조합3	HLIT2,5,15	0.694	0.894

<표 5>는 원척도 기준 저건강문해력군 분류 결과와 비교한 세 예비 축약형 도구의 분류 성능을 제시한 것이다. 예비조합1은 민감도 98.2%, 특이도 78.8%, 정확도 88.3%, 양성예측도 81.7%, 음성예측도 97.8%를 보였으며, 균형정확도는 88.5%, AUC는 0.965였다. 양성우도비는 4.63, 음성우도비는 0.023으로 나타났다. 예비조합3도 비교적 양호한 성능을 보였으나, 민감도와 음성예측도, AUC는 예비조합1보다 낮았다. 예비조합2는 특이도와 양성예측도는 높았으나, 민감도가 낮아 저건강문해력군을 넓게 선별하는 목적에는 적합성이 낮았다.

가능한 모든 3문항 조합 560개를 추가로 검토한 결과, 예비조합1의 AUC는 0.9654로 두 번째로 높았으며, 최고 AUC 조합(HLIT6, HLIT8, HLIT12; AUC=0.9662)과의 차이는 0.0008이었다. 예비조합1은 최고 AUC 조합보다 민감도와 음성예측도가 높았으며, 문항의 내용적 구성과 저건강문해력군의 누락을 줄이는 선별 목적을 함께 고려하여 후속 검증의 우선 후보조합으로 선정하였다.

표 5. 3문항 예비조합의 분류 성능 비교(원척도 기준 비교)

(N=8,358)

도구명	TP	FN	FP	TN	민감도 (%)	특이도 (%)	정확도 (%)	PPV (%)	NPV (%)	균형정확도 (%)	LR+	LR-	AUC
예비 1	4,034	74	901	3,349	98.20	78.80	88.33	81.74	97.84	88.50	4.63	0.023	0.965
예비 2	1,541	2,567	85	4,165	37.51	98.00	68.27	94.77	61.87	67.76	18.76	0.638	0.679
예비 3	3,930	178	930	3,320	95.67	78.12	86.74	80.86	94.91	86.90	4.37	0.055	0.941

주: 1) TP=true positive, FN=false negative, FP=false positive, TN=true negative. 민감도=TP/(TP+FN)×100, 특이도=TN/(TN+FP)×100, 정확도=(TP+TN)/전체×100, PPV(positive predictive value)=TP/(TP+FP)×100, NPV(negative predictive value)=TN/(TN+FN)×100, 균형정확도=(민감도+특이도)/2. LR+=민감도/(1-특이도), LR-=(1-민감도)/특이도, AUC는 3문항 합산점수의 전체 절단점에 대한 ROC 분석 결과이다.

2) 각 축약형 도구는 3문항 이진화 합산점수 0~3점으로 구성하였으며, 2점 이하를 저건강문해력군으로 분류하였다. 원척도 저건강문해력군은 HLS-EU-Q16 이진화 합산점수 12점 이하로 정의하였다.

3) 선정후보인 예비조합1의 95% 신뢰구간은 민감도 97.7-98.6%, 특이도 77.5-80.0%, 정확도 87.6-89.0%, 양성예측도 80.6-82.8%, 음성예측도 97.3-98.3%, AUC 0.962-0.968이었다.

4. 예비조합1의 절단점별 성능과 추가 검증

<표 6>은 예비조합1의 절단점에 따른 분류 성능을 제시한 것이다. 1점 이하 기준은 특이도 96.7%, 양성예측도 96.1%, 양성우도비 25.41로 저건강문해력 가능성이 높은 대상을 좁게 선별하는 데 유리하였다. 반면 2점 이하 기준은 민감도 98.2%, 음성예측도 97.8%, 음성우도비 0.023으로 나타나 저건강문해력군의 누락을 최소화하는 데 강점이 있었다. 0점 기준은 특이도는 매우 높았으나 민감도가 낮았고, 3점 이하 기준은 모든 대상자를 위험군으로 분류하게 되어 실질적 기준으로 적절하지 않다. 따라서 지역사회 1차 선별과 같이 위험군 누락을 최소화해야 하는 상황에서는 2점 이하 기준이 적절하며, 자원이 제한되어 개입대상을 보다 엄격하게 선별해야 하는 상황에서는 1점 이하 기준을 대안으로 고려할 수 있다.

표 6. 예비조합1의 절단점에 따른 분류 성능 비교

(N=8,358)

분류 기준	TP	FN	FP	TN	민감도 (%)	특이도 (%)	정확도 (%)	PPV (%)	NPV (%)	균형정확도 (%)	LR+	LR-
0점	2,181	1,927	3	4,247	53.09	99.93	76.91	99.86	68.79	76.51	752.13	0.469
1점 이하	3,414	694	139	4,111	83.11	96.73	90.03	96.09	85.56	89.92	25.41	0.175
2점 이하	4,034	74	901	3,349	98.20	78.80	88.33	81.74	97.84	88.50	4.63	0.023
3점 이하	4,108	0	4,250	0	100.00	0.00	49.15	49.15	-	50.00	-	-

주: 1) 원칙도 저건강문해력군은 HLS-EU-Q16 이진화 합산점수 총점 12점 이하로 정의하였다.
 2) 예비조합1은 3문항 이진화 합산점수(0~3점)이며, 절단점은 해당 점수 이하를 저건강문해력군으로 분류한 기준이다.
 3) 3점 이하 기준에서는 음성 분류 사례가 없어 NPV, LR+, LR-를 산출하지 않았다.
 4) 0점 기준의 LR+는 위양성 사례가 3명으로 매우 적어 크게 산출되었으므로 단독 해석에 주의가 필요하다.

<표 7>은 예비조합1의 표본분할 검증 결과를 제시한 것이다. 동일 자료 내 결과의 안정성을 점검하기 위해 70:30 표본분할 검증을 실시하였다. 검증표본에서 예비조합1의 2점 이하 기준은 민감도 98.3%, 특이도 79.1%, 정확도 88.5%, 양성예측도 81.9%, 음성예측도 98.0%를 보였다. 이는 전체표본 및 개발표본 분석과 유사한 수준으로, 예비조합1의 선별 성능이 표본분할 후에도 비교적 안정적으로 유지되는 것으로 볼 수 있다.

표 7. 예비조합1의 표본분할 검증 결과

표본	기준	N	TP	FN	FP	TN	민감도 (%)	특이도 (%)	정확도 (%)	PPV (%)	NPV (%)
개발표본	1점 이하	5,851	2,395	481	109	2,866	83.28	96.34	89.92	95.65	85.63
개발표본	2점 이하	5,851	2,823	53	634	2,341	98.16	78.69	88.26	81.66	97.79
검증표본	1점 이하	2,507	1,019	213	30	1,245	82.71	97.65	90.31	97.14	85.39
검증표본	2점 이하	2,507	1,211	21	267	1,008	98.30	79.06	88.51	81.94	97.96

주: 전체 분석대상자 8,358명을 원칙도 기준 저건강문해력 여부에 따라 층화한 뒤 70% 개발표본과 30% 검증표본으로 무작위 분할하였다.

<표 8>은 예비조합1 구성 문항을 제외한 13문항 보조준거 분석 결과를 제시한 것이다. 13문항 합산점수를 16점 만점으로 환산하고 12점 이하를 저건강문해력군으로 정의한 결과, 예비조합1의 2점 이하 기준은 민감도 97.8%, 음성예측도 97.8%를 유지하였다. 다만 특이도는 66.7%, 양성예측도는 66.0%로 낮아졌다.

표 8. 예비조합1 구성 문항 제외 13문항 보조준거 분석 결과

(N=8,358)

보조준거	기준	TP	FN	FP	TN	민감도 (%)	특이도 (%)	정확도 (%)	PPV (%)	NPV (%)
13문항 환산점수 ≤12	1점 이하	2,882	451	671	4,354	86.47	86.65	86.58	81.11	90.61
13문항 환산점수 ≤12	2점 이하	3,259	74	1,676	3,349	97.78	66.65	79.06	66.04	97.84

주: 13문항 보조준거는 예비조합1 구성 문항(HLIT1, HLIT8, HLIT12)을 제외한 나머지 13문항 합산점수를 16점 만점으로 환산한 뒤, 환산점수 12점 이하를 저건강문해력군으로 정의하였다.

가중치 적용 민감도 분석 결과는 <부표 2>에 제시하였다. 1점 이하 기준은 민감도 77.4%, 특이도 97.8%, 정확도 90.4%, 양성예측도 95.1%, 음성예측도 88.4%였다. 2점 이하 기준은 민감도 97.1%, 특이도 81.9%, 정확도 87.4%, 양성예측도 75.3%, 음성예측도 98.1%였다. 가중치 적용 후에도 2점 이하 기준의 민감도와 음성예측도는 높게 유지되었으나, 양성예측도는 비가중 분석보다 낮아졌다.

점수와 방식에 따른 정보손실 가능성을 검토하기 위해 보조 분석을 통해 확인했을 때, 4점 원점수 합산점수를 사용한 예비조합1의 AUC는 0.964로 <표 5>의 이분화 점수 기반 결과와 유사하였다. 또한 '잘 모르겠다'를 0점으로 처리한 방식에 따른 영향을 확인하기 위한 민감도 분석에서도 예비조합1의 2점 이하 기준은 민감도 98.1%, 음성예측도 97.3%로 주분석과 유사하였다.

V. 논의

본 연구는 한국의료패널 2021년 건강문해력 부가조사 자료를 활용하여 HLS-EU-Q16 기반 3문항 건강문해력 선별 후보조합을 구성하고 그 분류 성능을 예비적으로 검토하였다. 예비조합1은 원칙도 기준 저건강문해력군을 구분하는 데 높은 민감도와 음성예측도를 보였고, ROC/AUC에서도 양호한 판별력을 나타냈다. 이러한 결과는 예비조합1이 저건강문해력 위험군을 빠르게 포착하기 위한 1차 선별 후보도구로 활용될 가능성을 보여준다. 단축형 척도 개발은 보건의료 측정 분야에서 널리 활용되어 왔으며, 건강문해력 영역에서도 연구 목적과 활용 맥락에 따라 보다 간결한 단축형·선별형 도구의 필요성이 지속적으로 제기되어 왔다(Chew et al., 2004; Haun et al., 2014). 본 연구는 이러한 단축형·선별형 측정도구의 흐름을 국내 HLS-EU-Q16 자료에 적용했다는 점에서 의미가 있다.

기존 해외 건강문해력 단축도구와 비교하면, 예비조합1은 기존 도구의 문항을 활용한다는 점에서는 HLS-EU 계열 단축형과 유사하지만, 개발 목적과 문항선정 기준에서는 차이가 있다. HLS-SF12, HLS-Q12, HLS19-Q12는 원칙도의 다차원적 개념틀을 가능한 한 유지하면서 일반 건강문해력을 간결하게 측정하기 위해 개발되었다(Duong et al., 2019; Finbråten et al., 2018; Pelikan et al., 2022). 반면 Chew 등(2004)의 3문항 선별질문은 임상 현장에서 기능적 건강문해력의 어려움이 있는 환자를 신속히 식별하기 위해 별도의 문항으로 구성되었다. 예비조합1은 HLS-EU-Q16의 기존 문항을 활용하되 원칙도의 전체 개념구조를 축소하여 재현하기보다, 의료정보 탐색, 건강문제 관리 관련 정보 탐색, 예방적 의사결정과 정보 활용을 반영하는 문항을 선택하여 일반 인구에서 원칙도 기준 저건강문해력군을 선별하도록 구성하였다. 즉, 기존 척도의 문항풀을 활용한다는 점에서는 HLS-EU 계열 단축형과 유사하고, 위험군의 신속한 식별을 목적으로 한다는 점에서는 Chew 도구와 유사하지만, 일반 인구자료에서 원칙도 기준 분류 성능을 중심으로 문항조합을 구성했다는 점에서 차이가 있다.

본 연구 결과로 도출된 예비조합1은 HLS-EU-Q16 문항 중 정보 탐색과 정보 활용에 가까운 문항으로 구성되었

다. HLS-EU 개념틀은 건강문해력을 건강정보에 접근하고, 이해하고, 평가하고, 적용하는 과정으로 설명하는데 (Sørensen et al., 2012), 예비조합1은 이 가운데 의료적 설명 이해보다 정보 접근과 적용 기능을 상대적으로 더 많이 반영한다. 이러한 문항 조합이 원칙도 기준 저건강문해력군을 민감하게 포착했다는 점은, 본 자료에서 정보 탐색과 예방적 의사결정 관련 문항이 선별 성능과 관련되어 있음을 보여준다. 다만 이는 HLS-EU-Q16 기반 3문항 조합을 구성할 때 참고할 수 있는 경험적 근거로 보는 것이 타당하다.

본 연구에서 구성한 세 예비조합 간 차이는 축약형 선별도구에서 문항 내용이 분류 성능과 밀접하게 관련될 수 있음을 보여준다. 의사의 설명 이해, 약물 복용 방법 이해, 건강검진 필요성 이해와 같이 의료 의사소통 및 환자 안전과 직접 관련된 문항으로 구성된 예비조합2는 특이도와 양성예측도는 높았으나 민감도는 낮았다. 이는 의료인의 설명이나 복약지시 이해가 건강문해력의 중요한 하위 요소이지만, 본 자료에서는 저건강문해력군을 넓게 선별하는 데 한계가 있었음을 의미한다. 예비조합3도 예비조합1과 유사한 분류 성능을 보였으나, 민감도와 음성예측도, 전체 정확도를 종합하면 예비조합1이 선별 목적에 더 적합하였다.

절단점 분석 결과는 선별도구의 활용 목적에 따라 기준 선택이 달라질 수 있음을 보여준다. 1점 이하 기준은 특이도와 양성예측도가 높아 개입대상을 엄격하게 선별하는 데 유리하였다. 반면 2점 이하 기준은 민감도와 음성예측도가 높아 저건강문해력군의 누락을 줄이는 데 유리하였다. 분류도구 연구에서는 민감도, 특이도, 예측도가 도구의 사용 목적에 따라 다르게 해석된다(Trevethan, 2017). 지역사회 1차 선별처럼 누락을 줄이는 것이 중요한 상황에서는 2점 이하 기준을 우선 고려할 수 있고, 자원이 제한되어 개입대상을 보다 좁게 설정해야 하는 상황에서는 1점 이하 기준을 대안으로 검토할 수 있다.

표본분할 검증과 13문항 보조준거 분석에서도 예비조합1의 민감도와 음성예측도는 높은 수준으로 유지되어, 주분석 결과의 안정성을 뒷받침하였다. 이는 주분석에서 확인된 분류 성능이 동일 자료를 사용한 데 따른 과적합이나 원칙도와의 문항 중복에 의해서만 나타난 결과는 아님을 보여준다. 한편 13문항 보조준거 분석에서는 특이도와 양성예측도가 낮아졌으므로, 예비조합1은 저건강문해력 여부를 확정하기보다 위험 가능성이 있는 대상을 넓게 포착하는 선별도구로 보는 것이 타당하다.

가중치 적용 민감도 분석에서도 예비조합1의 핵심 선별 특성은 유지되었다. 다만 본 연구의 목적은 모집단 수준의 건강문해력 유병률을 추정하는 데 있기보다 원칙도 기준 분류 성능을 평가하는 데 있었으므로, 가중치 분석 결과는 주분석의 안정성을 확인하는 보조 근거로 보는 편이 타당하다.

본 연구의 학문적 의의는 기존 국내 연구와 비교할 때, HLS-EU-Q16 문항 자체를 활용하여 3문항 선별조합을 구성하고 그 분류 성능을 검토하였다는 점에 있다. 최근 한국의료패널 자료를 활용한 연구들은 건강문해력 수준, 관련 요인, 삶의 질, 의료이용, 연령집단별 차이 등을 분석해 왔다(배재용, 김혜윤, 2023; 문석준 외, 2024; Song, 2024; Kwon & Kwon, 2025; Kim et al., 2025). 이들 연구가 건강문해력의 분포와 관련 요인, 건강결과와의 관계를 밝히는 데 초점을 두었다면, 본 연구는 HLS-EU-Q16을 활용하여 3문항 선별조합을 구성하고 원칙도 기준 분류 성능을 평가하였다는 점에서 차이가 있다.

실무적으로 예비조합1은 긴 설문을 적용하기 어려운 초기 상담에서 저건강문해력 위험군을 빠르게 파악하기 위한 보조자료로 활용될 수 있다. 응답 부담이 낮고 점수 산정이 단순하므로, 보건소와 복지관 등 지역사회 기반 건강교육 요구도 파악이나 건강정보 지원 대상자의 초기 선별 과정에서 활용 가능성이 있다. 다만 개인의 건강문해력 수준을 확정하는 용도로 사용하기보다는, 추가평가, 쉬운 정보 제공, 설명 방식 조정, 맞춤형 건강교육 및 건강정보 탐색 지원의 필요성을 검토하는 자료로 활용하는 것이 적절하다.

그러나 본 연구에는 몇 가지 한계가 있다. 첫째, 본 연구는 HLS-EU-Q16 완전응답자를 대상으로 3문항 선별도구를 구성하고 예비적으로 평가한 이차자료 분석이므로, 분석 결과는 한국 성인 인구 전체에 대한 추정치라기보다 완전응답 표본 내 문항 특성과 분류 성능을 중심으로 해석할 필요가 있다. 둘째, 표본분할 검증을 실시하였으나

이는 동일 자료 내 내적 검증이므로 외부 독립표본 검증을 대체하지 못한다. 셋째, 예비조합1의 구성 문항은 정보 탐색과 판단 및 적용 기능에 상대적으로 집중되어 있어 건강문해력의 전체 개념 구조를 균형 있게 대표한다고 보기는 어렵다. 또한 본 연구에서는 문항-총점 상관, 내적 일관성, 요인분석 및 분류 성능을 중심으로 문항을 검토하였으며, 향후에는 문항반응이론 모형을 활용한 정보량 분석을 통해 문항선정의 측정학적 근거를 보완할 필요가 있다. 넷째, 양성예측도와 음성예측도는 분석 대상 집단의 저건강문해력 분포에 영향을 받으므로, 다른 인구집단이나 실제 적용 환경에서는 값이 달라질 수 있다. 다섯째, 본 연구는 예비조합 1의 원칙도 기준 분류 성능 평가에 초점을 두었기 때문에, 건강행동, 의료이용, 건강결과 등 외부 준거와의 관련성 및 고령층, 저학력층, 저소득층 등 하위집단별 측정 안정성은 검토하지 못하였다. 향후 연구에서는 독립 표본을 활용한 교차타당화와 준거-예측 타당도 검증, 집단별 적용 가능성 및 측정 동일성 검토와 함께, 다른 초단축 건강문해력 선별도구들과의 비교를 통해 선별성과 적용 맥락의 차이를 함께 검토할 필요가 있다.

종합하면, 예비조합1은 높은 민감도와 음성예측도를 바탕으로 저건강문해력 위험군을 선별하는 3문항 후보도구로서 양호한 성능을 보였다. 본 연구는 전국 단위 패널자료를 활용하여 저건강문해력 위험군을 신속히 파악할 수 있는 선별 도구의 후보를 제시하고, 대규모 조사와 지역사회 및 보건의료 현장의 초기 선별을 위한 기초적 실증 근거를 마련하였다는 점에서 의의가 있다.

이영선은 이화여자대학교에서 사회복지학 학사 및 석사, 박사학위를 받았으며, 조선대학교 행정 복지학부 사회복지학 전공 교수로 재직 중이다. 주요 관심 분야는 의료사회복지, 정신건강사회복지, 사회복지 실천 등이며, 현재 저소득층 노인 건강, 건강 정보 이해력, 통합 돌봄 등을 연구하고 있다.

(E-mail: claudia@chosun.ac.kr)

참고문헌

- 강수진, 이미숙. (2015). 근거 중심의 헬스 리터러시 증진 방안: 학술지 연구 동향을 중심으로. *보건교육건강증진학회지*, 32(4), 93-108. <https://doi.org/10.14367/kjhep.2015.32.4.93>
- 문석준, 김정훈, 최용석, 김희년. (2024). 건강문해력(Health Literacy)의 영향요인 파악 및 의료이용에 미치는 영향 분석. *보건사회연구*, 44(4), 130-150. <https://doi.org/10.15709/hsr.2024.44.4.130>
- 배재용, 김혜윤. (2023). 한국의료패널로 본 헬스 리터러시 실태와 정책적 시사점. *보건복지포럼*, 316, 81-94. <https://doi.org/10.23062/2023.02.7>
- 천희란, 김수현, 박은자. (2022). 우리나라 헬스리터러시 측정 도구의 연구 동향 분석: 주제범위 문헌고찰. *보건교육건강증진학회지*, 39(4), 39-53. <https://doi.org/10.14367/kjhep.2022.39.4.39>
- Berkman, N. D., Sheridan, S. L., Donahue, K. E., Halpern, D. J., & Crotty, K. (2011). Low health literacy and health outcomes: An updated systematic review. *Annals of Internal Medicine*, 155(2), 97-107. <https://doi.org/10.7326/0003-4819-155-2-201107190-00005>
- Chew, L. D., Bradley, K. A., & Boyko, E. J. (2004). Brief questions to identify patients with inadequate health literacy. *Family Medicine*, 36(8), 588-594.
- Davis, T. C., Long, S. W., Jackson, R. H., Mayeaux, E. J., George, R. B., Murphy, P. W., & Crouch, M. A. (1993). Rapid estimate of adult literacy in medicine: A shortened screening instrument. *Family Medicine*, 25(6), 391-395.
- DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications* (5th ed.). SAGE Publications.
- Duong, T. V., Aringazina, A., Kayupova, G., Nurjanah, Pham, T. V., Pham, K. M., Truong, T. Q., Nguyen, K. T., Oo, W. M., Su, T. T., Majid, H. A., Sørensen, K., Lin, I.-F., Chang, Y., Yang, S.-H., & Chang, P. W. S. (2019). Development and validation of a new short-form health literacy instrument (HLS-SF12) for the general public in six Asian countries. *Health Literacy Research and Practice*, 3(2), e91-e102. <https://doi.org/10.3928/24748307-20190225-01>
- Finbråten, H. S., Wilde-Larsson, B., Nordström, G., Pettersen, K. S., Trollvik, A., & Gutterud, Ø. (2018). Establishing the HLS-Q12 short version of the European Health Literacy Survey Questionnaire: Latent trait analyses applying Rasch modelling and confirmatory factor analysis. *BMC Health Services Research*, 18, 506. <https://doi.org/10.1186/s12913-018-3275-7>
- Gustafsdóttir, S. S., Sigurdardóttir, A. K., Arnadóttir, S. A., Heimisson, G. T., & Mártensson, L. (2020). Translation and cross-cultural adaptation of the European Health Literacy Survey Questionnaire, HLS-EU-Q16: the Icelandic version. *BMC Public Health*, 20(1), 61. <https://doi.org/10.1186/s12889-020-8162-6>
- Haun, J. N., Valerio, M. A., McCormack, L. A., Sørensen, K., & Paasche-Orlow, M. K. (2014). Health literacy measurement: An inventory and descriptive summary of 51 instruments. *Journal of Health Communication*, 19(Suppl 2), 302-333. <https://doi.org/10.1080/10810730.2014.936571>
- HLS-EU Consortium. (2012). *Comparative report of health literacy in eight EU member states: The European Health Literacy Survey HLS-EU*. The HLS-EU Consortium.
- Kickbusch, I., Pelikan, J. M., Apfel, F., & Tsouros, A. D. (Eds.). (2013). *Health literacy: The solid facts*. WHO Regional Office for Europe.
- Kim, S., Hwang, S., Lee, J., & Choi, M. (2025). Health literacy disparities in South Korea: insights from a latent profile analysis. *BMC Public Health*, 25(1), 1830. <https://doi.org/10.1186/s12889-025-23025-3>
- Kwon, D. H., & Kwon, Y. D. (2025). Patterns of health literacy and influencing factors differ by age: a cross-sectional study. *BMC Public Health*, 25(1), 1556. <https://doi.org/10.1186/s12889-025-22838-6>

- Nutbeam, D. (2000). Health literacy as a public health goal: A challenge for contemporary health education and communication strategies into the 21st century. *Health Promotion International*, 15(3), 259-267. <https://doi.org/10.1093/heapro/15.3.259>
- Nutbeam, D. (2008). The evolving concept of health literacy. *Social Science & Medicine*, 67(12), 2072-2078. <https://doi.org/10.1016/j.socscimed.2008.09.050>
- Parker, R. M., Baker, D. W., Williams, M. V., & Nurss, J. R. (1995). The test of functional health literacy in adults: A new instrument for measuring patients' literacy skills. *Journal of General Internal Medicine*, 10(10), 537-541. <https://doi.org/10.1007/BF02640361>
- Pelikan, J. M., & Ganahl, K. (2017). Measuring health literacy in general populations: Primary findings from the HLS-EU Consortium's health literacy assessment effort. In R. A. Logan & E. R. Siegel (Eds.), *Health literacy: New directions in research, theory and practice* (Studies in Health Technology and Informatics, Vol. 240, pp. 34-59). IOS Press. <https://doi.org/10.3233/978-1-61499-790-0-34>
- Pelikan, J. M., Link, T., Straßmayr, C., Waldherr, K., Alferts, T., Bøggild, H., Griebler, R., Lopatina, M., Mikšová, D., Nielsen, M. G., Peer, S., Vrdeja, M., & the HLS19 Consortium of the WHO Action Network M-POHL. (2022). Measuring comprehensive, general health literacy in the general adult population: The development and validation of the HLS19-Q12 instrument in seventeen countries. *International Journal of Environmental Research and Public Health*, 19(21), 14129. <https://doi.org/10.3390/ijerph192114129>
- Rouquette, A., Nadot, T., Labitrie, P., Van den Broucke, S., Mancini, J., Rigal, L., & Ringa, V. (2018). Validity and measurement invariance across sex, age, and education level of the French short versions of the European Health Literacy Survey Questionnaire. *PLOS ONE*, 13(12), e0208091. <https://doi.org/10.1371/journal.pone.0208091>
- Song, I. (2024). Relationship between health literacy and health-related quality of life in Korean adults with chronic diseases. *PLOS ONE*, 19(4), e0301894. <https://doi.org/10.1371/journal.pone.0301894>
- Sørensen, K., Van den Broucke, S., Fullam, J., Doyle, G., Pelikan, J. M., Slonska, Z., & Brand, H. (2012). Health literacy and public health: A systematic review and integration of definitions and models. *BMC Public Health*, 12(1), 80. <https://doi.org/10.1186/1471-2458-12-80>
- Sørensen, K., Van den Broucke, S., Pelikan, J. M., Fullam, J., Doyle, G., Slonska, Z., Kondilis, B., Stoffels, V., Osborne, R. H., & Brand, H. (2013). Measuring health literacy in populations: Illuminating the design and development process of the European Health Literacy Survey Questionnaire (HLS-EU-Q). *BMC Public Health*, 13(1), 948. <https://doi.org/10.1186/1471-2458-13-948>
- Steyerberg, E. W. (2019). *Clinical prediction models: A practical approach to development, validation, and updating* (2nd ed.). Springer. <https://doi.org/10.1007/978-3-030-16399-0>
- Storms, H., Claes, N., Aertgeerts, B., & Van den Broucke, S. (2017). Measuring health literacy among low literate people: an exploratory feasibility study with the HLS-EU questionnaire. *BMC Public Health*, 17(1), 475. <https://doi.org/10.1186/s12889-017-4391-8>
- Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: A practical guide to their development and use* (5th ed.). Oxford University Press. <https://doi.org/10.1093/med/9780199685219.001.0001>
- Trevethan, R. (2017). Sensitivity, specificity, and predictive values: Foundations, pliabilities, and pitfalls in research and practice. *Frontiers in Public Health*, 5, 307. <https://doi.org/10.3389/fpubh.2017.00307>

부록

부표 1. 예비 축약형 도구 구성을 위한 문항 검토 결과

문항	요인적재량/ 공통성 ¹	문항- 총점 상관 ²	α 기여 ³	변별 ⁴	이론적 영역	예비 조합 포함 ⁵	선정/제외 사유
HLIT1	(.843 / .724)	.687	기여	양호	의료	예비조합1	의료정보 탐색을 반영하는 핵심 후보
HLIT2	(.792 / .663)	.627	기여	양호	의료	예비조합3	의료 접근 및 도움의 출처 파악을 반영하는 후보
HLIT3	(.214 / .525)	.421	낮음	양호	의료	예비조합2	의료 의사소통 이해를 반영하는 후보
HLIT4	(.087 / .741)	.321	낮음	낮음	의료	예비조합2	환자 안전 및 복약 이해 맥락을 반영하는 후보
HLIT5	(.810 / .682)	.642	기여	양호	의료	예비조합3	진료 연계 및 추가 진료 판단을 반영하는 후보
HLIT6	(.819 / .701)	.700	기여	양호	의료	제외	HLIT5와 의료 의사결정 기능이 유사하여 조합의 중복 가능성
HLIT7	(.126 / .646)	.282	낮음	낮음	의료	제외	낮은 문항-총점 상관과 제한된 변별력, HLIT3, HLIT4와 기능 중복 가능성
HLIT8	(.836 / .719)	.707	기여	양호	의료	예비조합1	건강문제 관리 관련 정보 탐색을 반영하는 핵심 후보
HLIT9	(.446 / .352)	.581	기여	양호	질병예방	제외	요인 구조상 교차적재 경향이 있고 HLIT12와 예방 판단 기능이 일부 중복
HLIT10	(.201 / .621)	.417	낮음	낮음	질병예방	예비조합2	예방 및 검진 필요성 이해를 반영
HLIT11	(.788 / .647)	.622	기여	양호	질병예방	제외	HLIT12와 미디어 기반 건강위험 판단 기능의 유사성 고려
HLIT12	(.847 / .732)	.705	기여	양호	질병예방	예비조합1	예방적 의사결정과 정보 활용을 반영하는 핵심 후보
HLIT13	(.828 / .708)	.701	기여	양호	건강증진	제외	HLIT8과 정신건강 관련 정보 탐색 기능이 유사하여 대표 문항 간 중복 가능성
HLIT14	(.231 / .420)	.413	낮음	낮음	건강증진	제외	낮은 상관과 제한된 변별력, HLIT15와 정보 이해 기능의 유사성
HLIT15	(.668 / .511)	.663	기여	양호	건강증진	예비조합3	건강증진 관련 정보 이해를 반영하는 후보
HLIT16	(.701 / .556)	.642	기여	양호	건강증진	제외	HLIT15와 건강 관련 판단 기능이 일부 중복

주: 1) 요인적재량, 공통성은 본문 <표 3>의 Oblimin 회전 결과를 기준으로 제시하였다.

2) 문항-총점 상관은 해당 문항을 제외한 나머지 15문항 합산점수와 의 수정 문항-총점 상관(corrected item-total correlation)

3) α 기여도는 문항 제거 시 Cronbach's α 변화 여부를 기준으로 판단하였다.

4) 변별력은 '이해 쉬움' 응답 비율이 90% 이상인 경우 낮은 변별력으로 간주하였다.

5) 예비 문항 조합 구성은 통계적 지표와 HLS-EU 개념들의 내용적 대표성을 종합적으로 고려하여 이루어졌으며, 최종 후보안 평가는 본문 <표 5>와 <표 6>의 분류 성능 결과를 함께 검토하여 수행하였다.

부표 2. 가중치 적용 민감도 분석 결과

기준	민감도(%)	특이도(%)	정확도(%)	PPV(%)	NPV(%)
1점 이하	77.39	97.75	90.39	95.10	88.44
2점 이하	97.13	81.94	87.43	75.25	98.06

주: 한국의료패널 데이터의 가구원 횡단 모집단 가중치를 적용하여 산출하였다.

Development and Preliminary Evaluation of a 3-Item Short-Form Health Literacy Screening Tool Based on the HLS-EU-Q16

Rhee, YungSeon¹

¹Chosun University

Abstract

This study constructed a 3-item candidate health literacy screening tool based on the HLS-EU-Q16 and evaluated its classification performance using Korea Health Panel data.

Data from 8,358 adults with complete HLS-EU-Q16 responses were analyzed. Three candidate combinations were constructed based on item distributions, corrected item-total correlations, internal consistency, exploratory factor analysis, and content representativeness. Low health literacy was defined as a full-scale score of 12 or lower. Split-sample validation and sensitivity analyses were conducted.

Candidate 1 (HLIT1, HLIT8, and HLIT12) had a Cronbach's α of .787 and a correlation of $r=.902$ with the full scale. At a cutoff score of 2 or lower, sensitivity was 98.2%, specificity 78.8%, accuracy 88.3%, negative predictive value 97.8%, balanced accuracy 88.5%, and AUC 0.965. Its AUC ranked second among all 560 possible 3-item combinations, while its sensitivity and negative predictive value were higher than those of the top-AUC combination. In the validation sample, sensitivity and negative predictive value were 98.3% and 98.0%, respectively.

Candidate 1 showed potential as a brief screening tool for rapidly identifying individuals at risk of low health literacy.

Keywords: Health Literacy, HLS-EU-Q16, Short-Form Screening Tool, Korea Health Panel Survey, Classification Performance