국민의 국어능력 조사 결과 분석*

김종철 · 김봉순 · 박재현 · 박현정 · 정희창 · 주재우
박은진 · 이관희 · 이지수 · 정재미 · 민병곤**

< 次 例 >

I. 서론
II. 연구 방법
III. 연구 결과
IV. 논의 및 제언

Ⅰ. 서론

작년 10월에 발표된 OECD 주관[1] 2013년 국제 성인 역량 조사(Program

---

* 이 논문은 2013년 국립국어원의 '2013년 국민의 국어능력 평가' 사업의 내용을 일부 수정하고 요약하여 재구성한 것이다.

** 김종철(서울대학교 국어교육과 교수, 제1저자), 박현정(서울대학교 교육학과 교수), 정희창(동덕여자대학교 국어국문학과 교수), 김봉순(공주교육대학교 국어교육과 교수), 박재현(상명대학교 국어교육과 교수), 주재우(서울대학교 BK21플러스 연수연구원), 이관희(서울대학교 국어교육연구소 연구원), 이지수(서울대학교 국어교육연구소 연구원), 박은진(서울대학교 국어교육과 박사과정), 정재미(서울대학교 국어교육과 석사과정), 민병곤(서울대학교 국어교육과 교수, 교신저자)

1) OECD 주관으로 시행된 국제 수준의 성인 문식력 비교 연구는 90년대 중반부터 이루어져 왔다. 먼저, 국제 성인 문식성 조사 연구(International Adult Literacy Survey, IALS)는 1994년, 1996년, 1998년의 세 차례에 걸쳐 시행되었으며, 성인 문식력 및 생애기능 조사 연구(Adult, Literacy and Lifeskills Survey, ALL)는 2003년과 2006년에 시행되었다. 이후 2013년에 1차 조사를 실시한 국제 성인 역량 조사(PIAAC)는 IALS와 ALL의 후속 연구에 해당하며 24개국(OECD 회원국 22개국 포함)이 참여하였다. 우리나라의 경우, IALS에는 직접적으로 참여하지 않았으나 해당 조사 도구를 자체적으

for the International Assessment of Adult Competencies, PIAAC) 결과[2]에 따르면, 우리 국민의 문식력 점수는 OECD 국가의 평균 정도에 불과한 것으로 나타났다. 이는 만 15세를 대상으로 한 국제 수준 학업 성취도 평가(Program for International Student Assessment, PISA)의 읽기 영역에서 우리나라 학생들이 보인 우수한 성취 결과[3]와 비교할 때 우려스러운 현상이다.

연령대가 높아질수록 문식력 점수가 낮아지는 현상은 PIAAC 결과에서도 드러나는데, 16세~24세 연령대의 평균 점수는 전체 4위였지만 35세~44세 연령대부터는 OECD 평균보다 점수가 낮아져 55세~65세 연령대의 평균 점수는 조사에 참여한 24개국 중 21위에 머물렀다. 특히, 최고 연령대(55세~65세)와 최저 연령대(16세~24세) 간의 점수 차이는 참여국 중에서 가장 큰 간극을 보였다.

이러한 결과는 우리 국민의 대다수가 학교 교육을 이수한 이후에는 문식력, 더 나아가 언어능력 개발에 소홀함을 여실히 보여 주는 것이라 할 수 있다. 그뿐만 아니라, 정책적 차원에서도 성인 대상의 국어교육은 주로 문자 해득(解得)[4]이나 기초 문해력 수준에 초점을 둔 경향(양명희, 2013)이 강하며, 국어 재교육을 전문적으로 시행하는 국립국어원 국어 문

로 번역하고 동일한 방식으로 조사를 시행하여 다른 국가와 비교한 결과를 이희수 외(2001)를 통해 보고하였다. 또한 ALL의 2차 주기 조사에 참여한 결과는 임언 외(2005)를 통해 보고된 바 있다.

2) 국내의 PIAAC 조사 참여는 한국직업능력개발원이 연구 총괄을, 통계청이 조사 실행을 담당하여 진행되었으며, 보다 자세한 내용은 국내 공식 웹페이지(piaackorea.krivet.re.kr)를 통해 확인할 수 있다.

3) 가장 최근에 이루어진 PISA 2012에서 우리나라 학생들의 읽기 평균 점수는 OECD 회원국 중에서 1~2위, 전체 참여국 중에서 3~5위를 차지하였다. 이에 대한 자세한 분석 결과는 송미영 외(2013)를 참조할 수 있다.

4) 교육과학기술부는 2009년에 문자 해득 교육과정을 고시하고, 이에 앞서 2007년에 성인 문해 교과서를 발간하였다. 대체적인 교육 내용은 초등학교 1학년에서 6학년 정도의 수준으로 마련되었다.

화 학교의 교육 과정도 주로 어문 규범을 익히는 차원(박미영, 2013)에만 한정되어 일반 국민을 대상으로 한 체계적인 국어교육은 이루어지지 않는 실정이다.

사정이 이러함에도 불구하고 성인 대상 국어교육을 설계하고 나아가 관련 국어 정책을 수립·시행하기 위해 현재의 상태를 체계적이고 정교하게 진단하고자 하는 노력 역시 미진한 편이다. 우리 국민의 국어능력이 현재 어느 정도의 수준인지, 영역별로는 어떤 수준 차이를 보이는지, 다양한 변인별로 어떠한 양상으로 국어능력을 보유하고 있는지 등에 대한 구체적이고 실증적인 기초 자료가 확보되어야 문제 상황을 타개하기 위한 국어능력 향상 방안을 마련할 수 있음에도 불구하고 지금까지 이에 대한 조사가 정합적으로 이루어지지 못해 온 것도 사실이다[5].

물론, 국어능력에 대한 실태 조사가 전혀 없었던 것은 아니다. 국어 기본법 제9조 및 동법 시행령 제2조에서 국민의 국어능력에 대한 실태 조사에 대해 명시하고 5년 주기로 국어 발전 기본 계획을 수립하는 데 실태 조사 결과를 반영할 것을 규정하면서 국어능력 실태 조사 역시 5년 주기로 시행할 필요가 있게 되었다. 이를 전후하여 국립국어원을 중심으로 국어능력 조사 방식을 논의한 일련의 연구들이 시행되었으며[6], 시범 수준의 조사나 특정 직업군을 대상으로 한 조사가 이루어졌다[7]. 특히, 김창원 외(2008)는 12,137명의 대규모 인원을 대상으로 조사를 실시하고 그 결과를 분석하여 실증적 자료로서의 가치를 확보하였다. 그렇지만, 이 연구는

5) 앞서 언급한 국제 수준의 연구 결과는 국가별로 상대적 비교를 위한 지표일 뿐, 2013년 현재 우리 국민의 국어능력에 대한 미세한 결을 읽어 내는 데는 한계가 있다. 또한 표준화된 조사 도구를 번역하여 사용한다는 점에서 국어 및 국어사용 맥락이 조사에 반영되기 어려운 면이 있다.

6) 김문오 외(2005), 김성규 외(2006), 김동환 외(2008), 장경희 외(2009) 등의 연구 성과를 들 수 있다.

7) 국군 장병을 대상으로 조사한 박재현 외(2008), 교사 집단을 대상으로 조사한 윤여탁 외(2009)의 성과가 이에 해당한다.

비문해율 및 기초 문해력 수준의 조사여서 고등 수준의 문식력을 정밀하게 조사하지 못하다 보니(이현희・박철우, 2013: 252), 일반 국민을 대상으로 한 국어교육 및 관련 정책의 기초 자료로 활용되기에는 한계가 있다.

본 연구는 이제까지의 국민 대상 국어능력 실태 조사 연구 성과물들이 지니는 의의와 한계를 참조하여 보다 타당하고 객관적인 방식으로 조사를 시행함으로써 결과 분석의 신뢰도를 제고하고, 향후 5년 주기의 안정적인 조사 체재를 구축하기 위한 방법론을 마련하고자 하였다. 이를 위해, 기존 성과들의 한계를 분석하여 다음과 같은 방향성을 가지고 타당도와 신뢰도를 갖춘 조사를 시행하였다.

첫째, 조사 영역은 듣기, 말하기, 읽기, 쓰기, 문법의 5개 영역으로[8] 구획하되 그간 시간상・예산상의 이유로 충분히 시행되지 못한 말하기, 쓰기 영역의 직접 평가를 시행한다. 이를 위해, 대규모의 채점단을 구성하고 채점 방식의 타당도를 확보한다.

둘째, 조사 대상의 대표성을 확보하기 위해 다양한 변인별로 실제 인구 분포 비율을 고려하여 대규모 표본 집단을 선정한다.

셋째, 조사 대상이 일반 국민들임을 고려하여 조사 문항을 연성화하여 설계하되, 조사 실행의 효율화를 위해 조사 도구를 이원적으로 설계하고 이를 해석할 수 있는 측정학적 장치를 마련한다.

넷째, 조사 결과를 통해 국민의 국어능력 수준을 상대적으로 비교하는 데 그치지 않고, 그것이 의미하는 바가 무엇인지를 절대적 등급 지표로 환산하여 해석할 수 있는 방법론을 구축한다.

다섯째, 5년 주기 조사 정례화를 통한 추이 비교가 가능하도록 필요한 방안을 제시한다.

---

8) 국어 기본법 시행령 제2조에서는 국어능력 실태 조사를 위한 영역으로 듣기, 말하기, 읽기, 쓰기를 설정하나, 본고에서는 규범적 국어 사용의 중요성을 감안하여 문법 영역을 추가하였다.

이상의 방향성을 가지고 국어능력에 대한 정합적 조사를 통해 본 연구에서는 다음과 같은 연구 문제를 밝히고자 한다.

첫째, 2013년 현재를 기준으로 우리 국민의 국어능력은 어느 수준이며, 수준 등급별로 어떠한 분포를 보이는가?

둘째, 듣기, 말하기, 읽기, 쓰기, 문법 영역에 따른 국민의 국어능력 수준 차이는 어떠한가?

셋째, 성, 연령대, 지역 규모, 학력군, 직업군의 배경 변인별로 국민의 국어능력 수준 차이는 어떠한가?

II. 연구 방법

1. 조사 도구

1) 조사 도구의 축소 및 내용 보정

본 연구의 조사 도구는 장소원 외(2012, 국립국어원)의 기초 조사에서 개발한 조사 도구를 토대로 설계되었다. 이에 듣기, 말하기, 읽기, 쓰기, 문법의 5개 영역 설정 및 영역별 평가 요소 구획 등은 기본적으로 기초 조사의 것을 유지하였으나, 시행상의 문제를 고려하여 조사 도구의 규모를 재설계하고 예비 조사 결과를 바탕으로 조사 도구의 내용을 보정(補正)하였다.

우선 본 조사는 조사 대상이 무작위로 추출된 20대~50대의 일반 국민이라는 점을 고려하여 문항의 난이도를 조정하고, 조사 시행 방식이 1:1 면대면 조사라는 점, 그리고 한정된 예산 속에서 피조사자의 참여도를 높

이고 피로도를 최소화해야 한다는 점 등을 고려하여 조사 도구의 규모를 축소하였다. 이때 장소원 외(2012)에서 실시한 예비 조사 결과를 참조하여 문항 양호도가 좋지 못한 문항을 배제하였으며 이를 통해 조사 도구의 타당성을 높이고자 했다. 그 결과 <표 1>과 같이 46개 문항, 110분, 1,000점 만점으로 조사 도구의 전체 틀을 확정하였다.

<표 1> 본 조사의 조사 도구 규모[9)]

영역	문항 수	시간	배점
읽기	20문항	40분	300점
문법	10문항	10분	150점
쓰기	3문항	30분	250점
듣기	10문항	15분	150점
말하기	3문항	15분	150점
누계	**46문항**	**110분**	**1,000점**

다음으로 장소원 외(2012)에서 실시한 예비 조사[10)] 결과에 대한 분석을 바탕으로 조사 도구의 내용을 보정할 필요가 제기되었다. 이 과정에서 기존에 제시된 문제점을 개선하는 논의와 함께, 새로 구성된 각 영역 전문가들의 검토 및 협의가 주기적으로 이루어졌다. 전문가 협의체에서는 다

9) 국민의 언어생활에서의 중요도와 사용 정도를 고려하여 각 영역별 요소를 세분화한 뒤 이를 균형 있게 반영할 수 있도록 평가 영역의 비중을 조정하였다. 읽기 영역은 구어인 말하기·듣기에 비해 다양한 텍스트로 양식화·형식화된다는 점을 고려하여 다른 영역에 비해 문항의 수가 많게 배정되었다. 특히 국어생활에서 읽기가 차지하는 비중을 고려하였다. 또한, 문법 영역은 말하기·듣기·읽기·쓰기의 기초 영역으로서 그러한 능력에 이미 작용하고 있다고 판단하여 배점을 낮추고, 쓰기 영역 역시 채점 요소의 일부가 문법 영역과 중복되는 면을 고려하여 배점을 하향 조정하였다.

10) 본 조사가 이루어지기 전 예비 조사는 크게 두 차례 이루어졌다. 장소원 외(2012) 기초 조사에서 문항 개발 시 1차 예비 조사가 이루어졌고, 이때 나타난 문제점들이 분석 결과로 제시된 것에 대해 2013년 연구에서 문항 내용 보정 과정을 거쳐 자체적으로 다시 2차 예비 조사를 실시하였다. 2013년에 이루어진 2차 예비 조사는 2013년 6월 24일부터 2013년 7월 16일까지 수도권에 거주하는 20대~50대 성인 남녀 400명을 대상으로 이루어졌으며, 이들 표본 추출은 성별, 연령대, 지역 규모, 직업군, 학력군에 따라 층화 표본 추출을 통해 이루어졌다.

음과 같은 기본 방향을 설정하고 각 영역의 세부 문항들을 보정하였다. 기본 방향은 다음과 같다.

첫째, 예비 조사를 통해 정답률이 지나치게 높게 나온 문항은 오답 매력도를 높이고 다양한 방식으로 발문을 구성하는 등 난이도를 조정한다.

둘째, 예비 조사를 통해 정답률이 지나치게 낮게 나온 문항은 담화의 분량 및 발문 방식 등을 간결하게 하고, 전문성이나 격식성 측면을 고려하여 문항을 수정한다.

셋째, 국민의 전반적인 국어능력을 조사할 수 있도록 특정 직업군이나 성별, 세대, 학력군에 유리하지 않은지 점검한다.

2) 영역별 세부 평가 요소

조사 도구의 세부 평가 요소는 기본적으로 기초 연구인 장소원 외(2012)에서 구안한 것을 유지하되, 규모 축소에 따라 일부 평가 요소를 통합하고 문항 수를 조정하였다. 각 영역별 세부 평가 요소는 <표 2>와 같다.

<표 2> 영역별 세부 평가 요소

영역	평가 능력 (또는 유형)	평가 요소	문항 수	
말하기 (S)	가. 설득하는 말하기	1. 설득(주장하는 말하기)	1	3
		2. 설득(요청하는 말하기)	1	
	나. 설명하는 말하기	1. 설명(설명하는 말하기)	1	
듣기 (L)	가. 사실적 듣기	1. 세부 정보의 확인	2	10
		2. 핵심 정보의 파악	1	
		3. 지시의 이행	1	
	나. 추론적 듣기	1. 내용의 추리	2	
		2. 심리 또는 태도의 추리	1	
		3. 적용	1	
	다. 비판적 듣기	1. 내용 또는 형식의 평가	2	

읽기 (R)	가. 사실적/축자적 읽기	1. 세부 내용 파악 능력	4	20
		2. 명시적인 중심 내용 파악 능력	2	
		3. 글의 구조 파악 능력	1	
		4. 내용 간 관계 파악 능력	2	
	나. 추론적 읽기	1. 비명시적 중심 내용 파악 능력	3	
		2. 전제나 가정의 파악 능력	2	
		3. 글 쓴 목적이나 동기 파악 능력	1	
		4. 생략된 내용 파악 능력	1	
		5. 글쓴이의 태도나 입장 파악 능력	1	
	다. 비판적 읽기	1. 구성의 적절성 파악 능력	0	
		2. 표현의 적절성 파악 능력	1	
		3. 내용의 타당성 파악 능력	1	
		4. 내용의 효용성 파악 능력	1	
쓰기 (W)	가. 친교적 쓰기	1. 3문장 정도의 글	1	3
	나. 설명적 쓰기	1. 300자 내외의 글	1	
	다. 논증적 쓰기	1. 400자 내외의 글	1	
문법 (G)	가. 규범	1. 맞춤법	1	10
		2. 띄어쓰기	1	
		3. 표준어	1	
	나. 어휘	1. 의미와 용법	2	
		2. 의미 관계	1	
		3. 관용 표현의 의미	1	
	다. 문장	1. 경어법	1	
		2. 호응 관계	1	
		3. 접속 관계	1	

또한 각 영역별로 요구하는 능력이 서로 다른 양식으로 실현된다는 점을 고려하여 각 영역별로 선택형, 서답형, 수행형 조사 방식을 구안하였다.

<표 3> 영역별 조사 방식

조사 영역	조사 방식	문항 수	배점
읽기	선택형(객관식)	20문항	300점
문법	선택형(객관식)	10문항	150점
듣기	선택형(객관식)	10문항	150점
말하기	수행형(주관식)	3문항	150점
쓰기	서답형(논술, 주관식)	3문항	250점
계		46문항	1,000점

3) 영역별 텍스트(담화) 제시 방식

조사 도구에서 각 영역별로 사용된 텍스트(담화) 유형은 <표 4>와 같다. 말하기 영역은 크게 설명 담화와 설득 담화로 구성되었으며, 듣기 영역은 이러한 담화를 듣고 답하는 문항으로 구성되었다. 읽기와 쓰기 텍스트는 크게 설명 텍스트, 설득 텍스트, 친교 텍스트로 구성되었으며, 문법 영역은 텍스트의 유형을 고려하기보다는 언어 단위를 고려하여 문항을 구성하였다.

<표 4> 영역별 담화 제시 유형

<table>
<tr><td rowspan="2">말하기</td><td>설명적</td><td colspan="2">설득적</td></tr>
<tr><td>경험이나 개념을 설명하는 말</td><td colspan="2">주장하는 말, 요청하는 말</td></tr>
<tr><td rowspan="2">듣기</td><td>설명적</td><td colspan="2">설득적</td></tr>
<tr><td>지시, 뉴스, 면담, 대화</td><td colspan="2">강연, 회의, 토론</td></tr>
<tr><td rowspan="2">읽기</td><td>설명적</td><td>설득적</td><td>친교적</td></tr>
<tr><td>도로명 안내, 가정 통신문, 기사문, 민원, 공문, 사용 설명서, 계약서</td><td>논설/사설, 만평, 공익 광고</td><td>편지, 문자메시지, 전자 우편</td></tr>
<tr><td rowspan="2">쓰기</td><td>설명적</td><td>설득적</td><td>친교적</td></tr>
<tr><td>설명하는 글</td><td>주장하는 글(논설문)</td><td>편지(초대하는 글)</td></tr>
<tr><td>문법</td><td colspan="3">담화 유형보다 언어 단위 고려(어휘, 문장, 담화/글)</td></tr>
</table>

4) 조사 도구의 이원화

앞서 언급한 바와 같이 시행상의 문제들을 고려하여 조사 도구 규모를 축소하였지만, 46개 문항을 일반 국민이 110분 동안 풀어야 하는 것, 수행형과 서답형의 문항을 준비하는 시간 등을 고려한다면 약 2시간 정도의 시간을 피조사자가 감수해야 한다는 점은 조사 시행의 커다란 난점으로

예상되었다.[11] 이에 조사 도구를 '가(A)' , '나(B)' 두 세트로 나누어 설계하고 두 세트의 동질성 확보를 위해 읽기 영역 6개 문항, 문법 영역 10개 문항을 공통 문항으로 설계하였다. 이는 '공통 문항 비동등 집단 설계' 방식에 해당한다.

<표 5> 조사 도구의 이원적 설계

《'가' 형_A set》				《'나' 형_B set》			
영역	문항 수	시간	배점	영역	문항 수	시간	배점
듣기	10	15	150	쓰기	3	30	250
말하기	3	15	150	읽기	13(6)	26	195
읽기	13(6)	26	195	문법	10	10	150
문법	10	10	150	누계	26문항	66분	595
누계	36문항	66분	645				

<표 5>와 같이 영역 배분을 한 이유는 첫째, '듣기'와 '말하기'의 구어적 속성을 고려하고, 둘째, 응답에 시간과 노력을 요하는 주관식 평가를 배분하며, 셋째, 평가 문항의 수가 많은 '읽기' 영역을 두 세트에 모두 포함되게 구성하고, 넷째, 동등화 검사의 기준이 될 수 있도록 하나의 영역(문법)을 두 세트에 모두 포함시키고자 하였기 때문이다.

이러한 설계 방식은 '공통 문항 비동등 집단 설계(common item non-equivalent design) 방식'이라고 한다. 이는 국가 수준 대단위 성취도 평가에서 일반적으로 활용되는 방식으로, 예를 들어, 국제 학업 성취도 평가인 TIMSS나 PISA에서 행렬 표집(matrix sampling) 방식으로 평가 도구를 10종 이상으로 분할하여 시행하는 것과 유사한 평가 방식이라 할 수 있다.

11) 조사 진행 시 피조사자의 적극적인 참여를 이끌어 내기 어려울 뿐만 아니라 긴 시간의 조사가 피조사자의 피로도를 누적시켜 국어능력을 제대로 측정하지 못할 것이라는 우려가 컸다.

이러한 조사 방식에서는 세트 간의 동등화가 보장되어야 한다. 일반적으로는 전체 문항의 20% 정도를 공통 문항으로 설정하지만, 공통 문항 수가 많으면 많을수록 동등화의 정확성은 보장된다. 이에 본 연구에서는 읽기 영역 6문항과 문법 영역 10문항을 동일하게 함으로써 조사 도구 간 동등화를 꾀하였다.[12)]

2. 조사 방법

1) 개별 가구 방문 조사

조사는 예비 조사와 본 조사 모두 표집된 개별 가구를 조사원들이 방문하여 1:1로 피조사자를 직접 만나 조사하는 방식으로 이루어졌다.

조사가 시작될 당시에는 '소수 집단 조사'를 실시한 후 '개별 방문 조사'를 실시하는 방안에 대한 고려가 있었다. 즉, 전국 17개의 국어 문화원과의 협조를 통해 조사 장소를 별도로 만들어, 표집된 피조사자를 해당 장소에 모아 조사를 실시하는 '소수 집단 조사'를 1차로 시행하는 것이다. 그런 다음 여기에서 배제된 변인들을 산출하여 해당하는 표본에 대해서만 개별적으로 1:1 가구 방문 조사를 실시하고자 하였다. 그러나 이렇게 할 경우 변인을 통제하기 어렵고, 조사 시행에 따른 비용과 시간의 부담이 크다는 문제가 제기되었다. 따라서 한정된 예산과 시간 속에서 조사 효율성을 담보할 수 있는 '1:1 개별 가구 방문 조사'를 조사 수행 방식으로 결정하였다.

---

12) 이때, 읽기 영역의 문항 중 어떤 문항을 공통 문항으로 설정할 것인가의 문제는 장소원 외(2012)의 기초 연구에서 시행된 예비 조사 결과를 바탕으로 문항의 변별도가 높고 난이도는 중간 정도인 문항을 설정하였다.

2) 조사원 교육

1:1 개별 가구 방문 조사의 경우 평가 결과에 영향을 미치는 중요한 요인이 바로 조사원의 숙달도이다. 이번 조사는 일반 국민들을 대상으로 하는 조사이면서, 어느 정도 '평가'의 속성을 띠고 있다. 또 피조사자가 1인당 1시간 이상을 조사에 집중적으로 참여해야 한다는 점 등을 고려했을 때 피조사자의 적극적인 참여를 이끌어 내기에 어려움이 따를 수 있다. 따라서 경험이 많은 능숙한 조사원들이 이 조사의 특수성을 이해하고, 국어능력 평가 조사를 위한 특정한 교육을 받은 후 조사에 임해야 할 필요가 있었다.

이에 본 연구에서는 몇 가지 방안을 강구하였다. 먼저, 이번 조사가 전국 단위의 대규모 조사라는 점과, 조사 시간이 최소 1시간 이상 소요된다는 점, 그리고 다른 설문 조사에 비해 피조사자들의 적극적인 참여를 유도하기 어렵다는 점을 고려하여 경험이 많은 능숙한 조사원들이 조사를 담당하기로 했다. 이에 따라 많은 설문 조사 경험을 보유한 30대~50대 여성을 위주로 조사원을 편성하였으며, 이들을 대상으로 1단계 이론 중심 교육, 2단계 실습 중심 교육으로 조사원 교육을 실시하였다.

다른 한편으로는, 조사원의 질 관리 방안으로 조사원 한 명당 수행할 수 있는 조사 부수에 제한을 두었다. 또한 조사원의 실질적인 질 관리를 위해 TNS Korea의 검증(Quality Control) 팀에서 조사원별로 조사 결과의 30%를 무작위로 추출해 조사원의 조사 수행 과정에 대해 점검하였다. 더하여 조사원 교육에서 표준화된 수행 절차와 상황 발생별 대응 매뉴얼을 교육함으로써 비표본 오차를 최소화하였다.

조사원들에게는 조사 도구에 대한 안내와 조사 진행 방법 외에도 조사원이 직접 조사 도구를 풀이해 보는 시간을 가져 피조사자들의 어려움과 예상되는 문제점들을 숙지하도록 하고, 이에 대처하는 방안도 교육하였다.

이 외에도 조사원을 운영하는 데에는 다음과 같은 사항들도 포함되었다. 첫째, 조사를 담당하는 조사원은 서울대학교 국어교육연구소와 TNS Korea의 직인이 포함된 조사원 명찰을 패용하고 이 조사가 정부 차원에서 이루어지는 것임을 명확히 고지하도록 하였다. 둘째, 국립국어원장의 인사말과 해당 조사에 대한 간단한 안내 및 조사 기관 소개 등이 포함된 안내 리플릿을 제공하여 피조사자들이 해당 조사에 대한 신뢰감을 갖도록 하였다. 셋째, 조사에 대한 감사의 표시로 간단한 보상품(문화 상품권)을 제공하였다.

3. 표본 추출 방법

1) 변인 설정

'국민의 국어능력 조사'라는 본 연구는 전체 국민을 모집단으로 하고 있다. 따라서 객관적인 결과를 얻기 위해 전체 국민을 대표할 수 있는 표본 집단의 선정이 무엇보다 중요한 연구이다. 이를 위해 다음과 같은 변인을 고려하여 표본 집단을 선정하였다.

① 성별 변인

성별 변인은 남성과 여성으로 구분하여 2개의 요소로 설정하였다.

② 연령 변인

연령 변인은 최소점을 20세로 설정하고 최대점을 59세로 설정하였다. 연령 변인을 이렇게 설정한 데에는 다음과 같은 이유가 있다. 먼저 20세 미만은 제도 교육을 받고 있는 대상으로 국가 수준 학업 성취도 평가와 같은 조사에서 국어능력을 측정하고 있어 대상에서 제외하였다. 다음으

로 59세를 초과한 일반 국민의 경우는 실제 조사에 응하기 어려운 측면이 있다는 점을 고려하였다.

연령 변인을 20대~50대로 설정하였지만 모든 연령을 변인으로 설정하기는 어렵기 때문에 이들을 일정한 연령대로 구획하여 집단화하였다. 이에 따라 연령 변인은 20대(만 20~29세), 30대(만 30~39세), 40대(만 40~49세), 50대(만 50~59세)의 4개 요소로 설정하였다.

③ 지역 변인

지역 변인은 다음과 같이 두 가지 방안을 고려하였다. 첫째, 제주도를 포함한 16개 시도의 전국적 분포를 고려하여 표본 집단을 추출하였다. 둘째, 지역 규모를 '대도시, 중소도시, 읍면 단위'의 세 요소로 설정하되, 각각 '서울특별시 및 5대 광역시의 동 지역, 이를 제외한 동 지역, 이를 제외한 읍면 단위'로 정의하여 지역 규모 변인을 상정하였다.

④ 학력 변인

학력 역시 국어능력을 결정하는 데 큰 영향을 미칠 것으로 예상되어 변인으로 설정하였다. 학력을 세분화하여 구획하면 '①무학, ②초등(국민)학교 졸업, ③중학교 중퇴, ④중학교 졸업, ⑤고등학교 중퇴, ⑥고등학교 졸업, ⑦대학교 재학, ⑧대학교 졸업, ⑨대학원 재학, ⑩대학원 졸업'이 가능하다. 그러나 이러한 열 가지 요소를 모두 통제하여 표본을 추출하는 것은 한정된 조사 단위에서 불가능하기 때문에, '중졸 이하(①~⑤), 고등학교 졸업(⑥), 대재 이상(⑦~⑩)' 세 요소로 구획하여 학력 변인을 통제하였다.

⑤ 직업 변인

직업 역시 국민의 국어능력을 결정짓는 중요한 변인이 된다. 본 연구에서는 한국 표준 직업 분류(KSCO)를 기준으로 직업군을 '정신노동(white

collar)', '육체노동(blue collar)', 학생, 주부, 무직자를 포함한 '기타'의 세 직업군으로 나누어 직업 변인을 통제하였다. 이때, 각 직업군에 속하는 세부 직업들은 <표 6>과 같다.

<표 6> 직업군별 세부 직업

직업군	세부 직업
정신노동자	관리자, 전문가 및 관련 종사자, 사무 종사자 등
육체노동자	서비스 종사자, 판매 종사자, 농림 어업 숙련 종사자, 기능원 및 관련 기능 종사자, 장치 기계 및 조립 종사자, 단순 노무 종사자, 군인 등
기타	주부, 학생, 무직 등

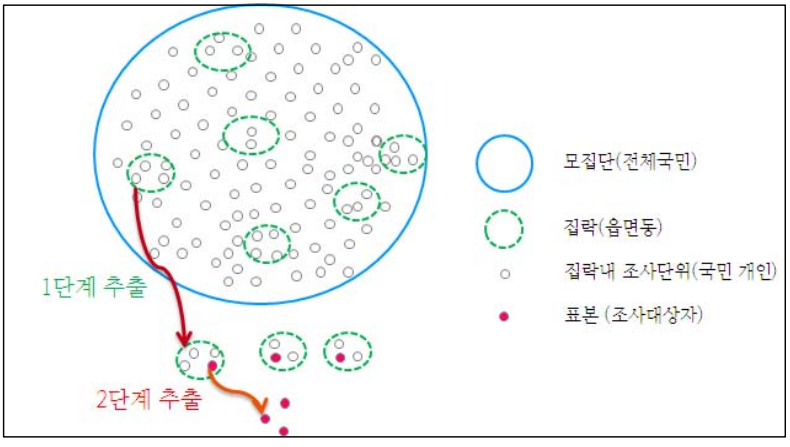
위에서 설정한 다섯 가지 변인들은 조사 도구의 가장 앞면에 기입하게 하였다. 그리고 피조사자들이 각 변인별로 어느 항목에 표시해야 하는지 판단하는 데 어려움을 느낄 수 있어 조사원이 구두로 물은 뒤 그 결과를 조사원이 직접 표시하는 방식으로 조사를 실시하였다.

2) 표본 추출 방식

표본 집단의 추출은 앞에서 설정한 성별, 연령, 지역/지역 규모, 직업, 학력 등 5개 변인들을 중심으로 이루어졌다. 먼저, 2단계 층화 표본 추출 방식과 집락 추출 방식을 혼용하였는데, 1차 층화 기준은 성별, 연령, 지역 변인이었으며, 2차 층화 기준은 직업과 학력 변인으로, 소프트 할당(soft quota) 방식으로 통제하였다.

이때, '집락(cluster)'은 서로 인접해 있는 개별 조사 단위들이 모여 있는 집단을 의미한다. 본 연구에서는 2단계 집락 추출 방식(cluster sampling)을 사용하여 표본 집단을 추출한 것으로 먼저 조사를 진행할 집락을 선정한 후 그 집락에서 표본을 추출하여 조사를 진행하는 방식으로 이루어진다. 이를 그림으로 표현하면 <그림 1>과 같다.

<그림 1> 2단계 집락 추출 방식



본 연구에서는 '집락'을 행정 구역상의 읍·면·동으로 삼고 조사 대상이 될 읍·면·동을 확률 비례 추출(Sampling with probabilities proportional to size)을 통해 먼저 추출한 후(1단계)[13], 해당 읍·면·동에서 조사에 참여할 개인을 추출(2단계)하였다.

본 연구에서는 조사 여건 및 조사 대상의 범위를 감안하여 하나의 집락(읍·면·동)에서 표집되는 표본 수를 10명으로 설계하였다. 1단계 표본 집락 추출 후, 추출된 집락(읍·면·동) 내에서 개별 표본을 추출한다. 이때에는 먼저 성별, 연령대의 분포를 고려하여 균등한 비율로 할당되도록 개별 표본(피조사자)을 추출한다.

다음으로, 학력과 직업 변인은 소프트 할당 방식으로 변인을 통제하는데, 이때 읍·면·동을 기준으로 학력군과 직업군을 할당하게 되면, 추출된 읍·면·동의 특성에 따라 특정 학력군이나 직업군이 상대적으로 많이 조사될 수도 있다. 이러한 문제를 보완하기 위해 학력군과 직업군은

13) 규모가 큰 집락일수록 표본으로 선정될 확률이 비례해서 높아지게 된다.

읍·면·동 단위가 아닌 시도별로 표본 수를 할당하고 그 비율에 맞춰 추출함으로써 대표성을 확보한다. 예컨대 특정 지역에서 고학력자가 상대적으로 많이 조사되었다면, 같은 시도 내 다른 읍·면·동에서는 저학력 응답자가 더 조사될 수 있도록 하여 해당 시도 전체의 학력 비율이 일정하게 유지될 수 있도록 하는 것이다. 이 과정에서 역시 통계청의 인구 통계 자료를 활용하여 비율을 유지한다.

4. 문항 채점 방법

1) 객관식 문항

본 연구에서 실시하는 조사에서 객관식 문항은 피조사자가 직접 조사 도구에 답을 기입하는 방식으로 진행되었다. 객관식 문항은 일반적으로 OMR 카드를 사용하는 경향이 있지만, 본 조사의 특성상 피조사자의 집단이 다양하여 자칫 OMR 카드의 미숙한 사용으로 조사 결과에 오류가 나타날 위험이 있기 때문이다. 피조사자들이 조사 도구에 직접 기입한 답안은 조사원들이 데이터로 변환하였다.

조사원들은 조사 결과를 TNS가 보유하고 있는 데이터 입력 시스템인 surveycraft 프로그램을 활용하여 통계 산출이 용이한 데이터로 변환하였다. 이러한 데이터를 바탕으로, 통계 패키지 프로그램을 이용하여 객관식 문항에 대한 채점을 진행하였다.

2) 주관식 문항

주관식 문항은 말하기 영역과 쓰기 영역에서 구성되었다. 올해 본 연구진이 자체 실시한 2차 예비 조사에서는 조사 대상이 각각 200명씩으로,

말하기 답안 600개(200명×3문항), 쓰기 답안 600개(200명×3문항)가 채점 대상이 되었다. 이에 비해 본 조사에서는 조사 대상의 수가 1,500명으로, 말하기 답안 4,500개(1,500명×3문항), 쓰기 답안 4,500개(1,500명×3문항)가 채점 대상이 되었다. 이러한 방대한 양의 주관식 답안을 채점하기 위해 본 연구에서는 전문성을 갖춘 대규모 채점단을 구성하였고, 지속적인 채점단 협의를 통해 채점자 간 신뢰도를 확보하며 채점을 진행하였다.[14]

5. 수준 등급화

1) 등급화의 목적

국민의 국어능력 조사 연구는 국민의 국어능력 수준을 수치화하는 데 그치는 것이 아니라 국어 정책을 수립·시행하기 위한 기초 자료로 그 결과를 활용하는 데에 목적이 있다. 이를 위해서는 국민의 국어능력 수치가 무엇을 의미하는지 해석할 수 있는 절대적 판단 수치가 필요하다.

절대적인 능력을 판별하기 위해서는 그 정도를 구분할 수 있는 분할 점수(cut-score)를 산출하여야 한다. 예를 들어, 읽기 영역의 평균 점수가 80점으로 도출되었다고 할 때, 이 수치가 의미하는 바가 무엇인지를 해석할 수 있어야 하는데, 분할 점수를 기준으로 80점이라는 수치가 우수 등급에 해당하는 것인지, 그렇지 않은지를 판단할 수 있어야 한다.

또한 본 연구를 통해 설정된 각 영역별 분할 점수들은 향후 다음 주기 조사 결과와의 비교를 위한 기준점으로 작용하게 된다. 이러한 기준점들은 측정학적으로 조사가 시행될 때마다 계속 유지되어야 하기 때문에 본 연구에서 분할 점수를 설정하는 것은 그 중요도가 매우 크다 할 수 있다.

---

14) 예비 조사와 본 조사의 채점단을 동일한 인원으로 편성하여 채점에 대한 연속성을 확보하고자 하였다.

따라서 각 영역의 분할 점수들은 타당하고 신뢰할 수 있는 방식으로 도출되어야 한다.

2) 등급화 방식

특정 능력의 수준을 등급화하기 위해서는 우선 몇 개의 등급으로 성취수준을 분할할 것인가를 결정해야 한다. 본 연구에서는 국가 수준 학업성취도 평가 등의 대규모 국어능력 평가에서 일반적으로 활용하는 '우수, 보통, 기초, 기초 미달'의 네 단계로 등급을 설정하였다. 이렇게 볼 때, 듣기, 말하기, 읽기, 쓰기, 문법의 각 영역별 분할 점수는 '우수와 보통 사이', '보통과 기초 사이', '기초와 기초 미달 사이'의 세 수준에서 결정된다.

'우수, 보통, 기초, 기초 미달'의 4개 등급의 분할 점수를 도출하기 위한 방식으로 본 연구에서 활용한 것은 '앵고프(Angoff) 방식'이다. 이 방법은 전문가 패널 집단이 각 문항에 대한 기대 정답률을 판정하여 분할 점수를 산출하는 과정으로 진행된다. 이때 기대 정답률의 판정은 최소 능력자(Minimum Competency Person, MCP)가 각 점수를 획득할 확률이나 그가 획득할 것으로 예상되는 점수를 개별 문항별로 기록하는 방식으로 이루어진다.

앵고프 방식을 통한 분할 점수 산출 과정은 다수의 전문가들로 구성된 전문가 집단에 의해 진행되었다. 전문가 집단은 말하기-듣기, 읽기, 쓰기, 문법으로 영역을 나누어 각 영역의 내용 전문가들을 균등한 비율로 선발하였다. 각 영역별로 6명씩 총 24명의 전문가를 분할 점수 도출을 위한 패널로 선정하였다.

이러한 방식으로 분할 점수를 도출하는 앵고프 방식은 전문가 패널의 개성과 편향성에 대한 문제가 제기되기 쉽다. 이를 고려하여 본 연구에서는 전문가 집단의 구성을 국어교육의 실천가이자 연구자로서의 교사, 대학원생(박사 수료 이상), 전공 교수의 비율이 적절하게 분포되도록 하였다.

또한 3명씩 구성된 네 개의 소집단을 중심으로 실질적인 판정과 토의가 이루어지도록 하였으며, 이때 소집단 편성에 있어서도 전문가 구성 비율이 반영되도록 하였다. 즉, 하나의 소집단은 교사 1명, 박사과정 이상 연구자 1명, 교수 1명으로 구성되었다. 이와 더불어 전공 영역과 관련해서도 각 소집단에는 반드시 한 명 이상의 다른 영역의 전공자가 포함되도록 집단을 구성하였다. 예를 들어, 읽기 전문가가 두 명 있는 소집단에는 쓰기 영역의 전문가 한 명이 포함될 수 있도록 하였다.

본 연구에서 상정한 국어능력의 수준 분할에 대한 기본적 전제는 <표 7>과 같다.

<표 7> 등급별 국어능력의 일반적 수준

등급	국어능력의 수준
우수	국민 공통 기본 교육과정 국어 과목을 이수한 국민에게 기대되는 국어능력을 **대부분(80% 이상)** 획득한 수준
보통	국민 공통 기본 교육과정 국어 과목을 이수한 국민에게 기대되는 국어능력을 **상당 부분(50% 이상 80% 미만)** 획득한 수준
기초	국민 공통 기본 교육과정 국어 과목을 이수한 국민에게 기대되는 국어능력을 **부분적으로(20% 이상 50% 미만)** 획득한 수준

판정은 3번의 라운드를 거치면서 조정되는 과정을 거쳤다. 이때 3번의 라운드는 <기대 정답률 판정 → 정보 피드백 → 토의 → 기대 정답률 판정 → 정보 피드백 → 토의 → 기대 정답률 판정>을 반복하면서 진행되었다. 각 라운드의 절차 및 제공된 정보들은 <표 8>과 같다.

<표 8> 라운드별 제공 정보

객관식 문항 (듣기, 읽기, 문법)	1라운드 → 2라운드	▪ 전체 전문가 집단의 1차 판정 결과의 중앙값과 분포표 (규준 정보)
	2라운드 → 3라운드	▪ 전체 전문가 집단의 2차 판정 결과의 중앙값과 분포표 (규준 정보) ▪ 실제 정답률(실질 정보)

		▪ 2차 판정 결과를 적용했을 때 예상되는 등급별 피조사자 비율과 분포도(영향 정보)
주관식 문항 (말하기, 쓰기)	1라운드 → 2라운드	▪ 전체 전문가 집단의 1차 판정 결과의 중앙값과 분포표(규준 정보)
	2라운드 → 3라운드	▪ 전체 전문가 집단의 2차 판정 결과의 중앙값과 분포표(규준 정보) ▪ 문항의 채점 요소별 평균 점수와 채점에 적용한 오점 척도별 누적 빈도 자료(실질 정보) ▪ 상, 중, 하 수준에 속하는 피조사자의 실제 답안(녹음 음성 파일 + 답안지)(실질정보) ▪ 2차 판정 결과를 적용했을 때 예상되는 등급별 피조사자 비율과 분포도(영향 정보)

III. 연구 결과

1. 영역별 분석 결과

본 연구는 조사 도구 설계상 다섯 영역을 모두 응답한 피조사자가 존재하지 않는다. 따라서 원 점수에 대한 총점을 직접적으로 계산하는 것은 불가능하다. 이를 대체하기 위해 각 영역별 척도 변환 점수의 평균에 대한 합산으로 전체 국어능력 점수의 평균을 산출하였다. 또한, 표준 편차를 산출하기 위해서는 '가' 형과 '나' 형의 공통 문항으로 설정된 16개 문항(읽기 영역 6개, 문법 영역 10개)의 원 점수 총점의 표준 편차에서 전체 영역 총점에 대한 공통 문항 총점의 비를 곱한 값을 산출하였다. 즉, 공통 문항 16개 문항에 대한 표준 편차를 전체 문항에 대한 표준 편차로 확장하여 산출한 것이다[15]. 이러한 과정을 거쳐 산출된 평균과 표준 편차에

15) 더 엄밀한 방식은 총점의 표준 편차를 산출하거나 각 영역별 표준 편차들을 합산하여 전체 표준 편차를 산출하는 방식이지만 전술하였듯이 본 연구에서는 2,916명의

대한 분포를 따르는 전체 국어능력에 대한 척도 변환 점수를 계산한 후, 절삭(truncation) 과정을 거쳐 최종적으로 전체 국어능력에 대한 척도 변환 점수를 산출하였다.

이러한 방식을 통해 산출한 척도 변환 점수를 기준으로, 전체 국어능력(총점) 및 각 영역별 평균과 표준 편차를 제시하면 다음과 같다. 다시 말해, 본 연구의 조사 결과에서는 1,000점 만점의 국어능력에 대한 조사 도구에서 579.62점의 평균 점수(척도 변환 점수 기준)가 산출된 것이다.

<표 9> 각 영역별 및 전체 국어능력의 척도 변환 점수

	평균	표준편차
듣기 영역	108.88	29.12
읽기 영역	190.66	61.81
문법 영역	88.14	28.87
말하기 영역	69.79	19.08
쓰기 영역	122.07	37.49
총점	**579.62**	**165.38**

이러한 과정을 거쳐 산출된 전체 국어능력에 대한 척도 변환 점수와 각 영역별 척도 변환 점수 사이의 상관계수를 분석하면 다음과 같다. 전체 국어능력 점수와 각 영역별 점수 사이의 상관계수가 듣기 영역 0.659, 읽기 영역 0.843, 문법 영역 0.576, 말하기 영역 0.479, 쓰기 영역 0.582로 분석되었다. 이 결과를 통해 각 영역과 전체 국어능력 점수에 대한 상관성을 확인할 수 있었다. 이는 본 연구의 방식을 통한 전체 국어능력 점수의 산출이 각 영역별 점수와 상관성을 확보할 수 있다는 점에서 산출 방식의

피조사자가 모든 영역의 문항에 대해 응답한 결과를 분석할 수 없기 때문에 이상의 방식으로 전체 점수의 표준 편차를 대체한 것이다. 이때 본 연구의 조사 도구 설계 시 공통 문항은 읽기와 문법 영역의 대표성을 띠는 문항으로 선정하였다는 점에서 공통 문항의 표준 편차를 통한 전체 문항의 표춘 편차 산출의 객관성을 어느 정도 확보할 수 있다.

타당성을 확인할 수 있다.

<표 10> 각 영역별 및 전체 국어능력의 척도 변환 점수 상관관계

	듣기	읽기	문법	말하기	쓰기	전체
듣기	1	.540	.348	.107	-	.659
읽기	.540	1	.421	.143	.194	.843
문법	.348	.421	1	.153	.208	.576
말하기	.107	.143	.153	1	-	.479
쓰기	-	.194	.208	-	1	.582
전체	.659	.843	.576	.479	.582	1

이상의 방식으로 산출한 전체 국어능력에 대한 점수 분포를 대규모 전문가 집단에 의해 앵고프 방식으로 분할한 성취 수준 등급으로 나눈 결과[16]는 다음 <표 11>과 같다.

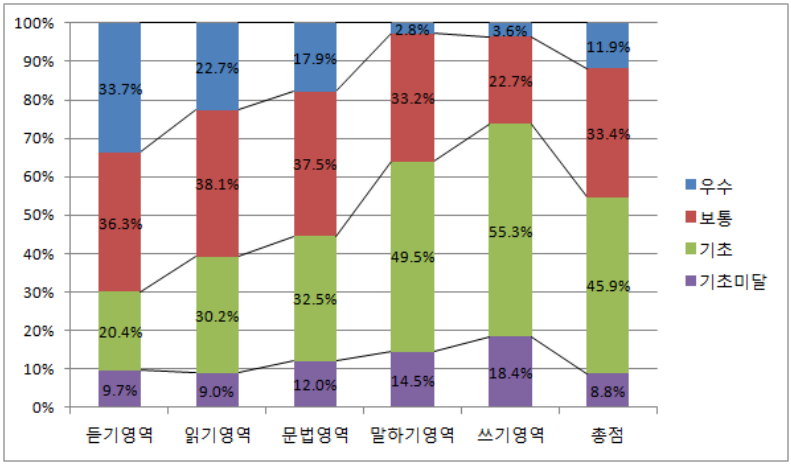
<표 11> 전체 국어능력의 수준 등급별 분포

영역	수준	분할 점수	인원(명)	비율
전체 국어능력	우수	781.2	347	11.9%
	보통	579.7	975	33.4%
	기초	372.0	1,338	45.9%
	기초 미달		256	8.8%
계			2,916	100%

이러한 전체 국어능력에 대한 수준 등급별 분포 비율을 각 영역별 수준 등급별 분포 비율과 비교하여 제시하면 <그림 2>와 같다. 다른 영역에 비해 쓰기와 말하기 영역의 기초 및 기초 미달 등급의 비율이 상대적으로 높아 국민들의 말하기, 쓰기 능력이 상대적으로 저조함을 확인할 수 있다.

16) 앵고프 방식에 의한 분할 점수의 산출은 각 영역별로 진행된바, 전체 국어능력에 대한 분할 점수는 각 영역의 분할 점수에 대한 합산으로 산출된다.

<그림 2> 각 영역별 및 총점의 수준 등급별 분포



2. 변인별 분석 결과

본 연구에서 상정한 성별, 연령, 지역 규모, 직업, 학력 다섯 가지의 변인별로 분석한 결과는 다음과 같다.

1) 성별

성별에서 평균 점수는 여성이 남성에 비해 10점 정도 높았지만 유의도[17]가 0.099로 산출되어 통계적으로 유의미한 차이는 아님을 확인할 수 있었다. 이러한 성별 차이를 수준 등급별로 비교해 보면 남성에 비해 여성이 보통 이상의 등급에 많이 분포되어 있음을 확인할 수 있다.

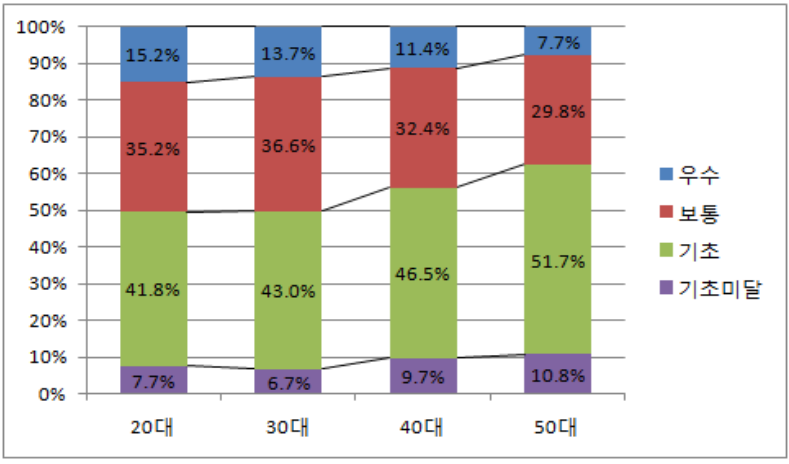
17) 일반적으로, 0.05를 기준으로 유의도가 이 수치보다 작을 경우 유의미한 차이가 있는 것으로 분석한다.

<그림 3> 국어능력 수준 등급의 성별 분포

2) 연령

연령대별로 전체 국어능력의 조사 결과의 차이를 분석해 보면, 유의도가 0.05보다 작아 통계적으로 유의한 차이가 있는 것으로 확인된다. 사후 비교를 위한 터키 검정 결과, 20대가 40대와 50대에 비해, 30대가 40대와 50대에 비해, 40대가 50대에 비해 국어능력이 우수함을 통계적으로 확인할 수 있었다. 이러한 연령대별 차이를 수준 등급으로 구획하여 비교해 보면 <그림 4>와 같다.

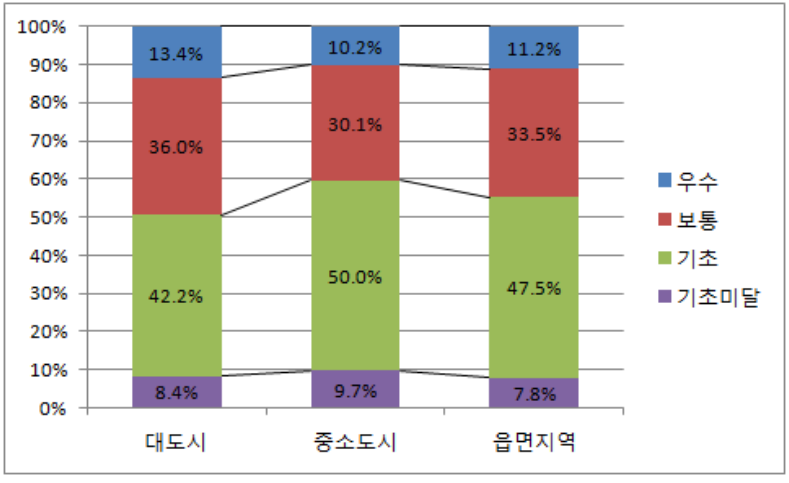
<그림 4> 국어능력 수준 등급의 연령대별 분포



3) 지역 규모

다음으로 지역 규모별 국어능력의 차이를 분석해 보면, 유의도가 0.05보다 작아 통계적으로 유의한 차이가 있는 것으로 확인되며, 사후 비교를 위한 터키 검정 결과, 대도시가 중소 도시에 비해 국어능력이 우수함을 통계적으로 확인할 수 있었다. 이러한 지역 규모별 차이를 수준 등급으로 구획하여 비교해 보면 <그림 5>와 같다.

<그림 5> 국어능력 수준 등급의 지역 규모별 분포



일반 국민을 대상으로 한 이번 조사에서 나타난 지역 규모별 국어능력 분포 차이는 고등학생을 대상으로 한 국가 수준 학업 성취도 평가 결과와 약간의 차이가 있다. 국가 수준 학업 성취도 평가 결과에서는, 전반적으로 기초와 기초 미달 등급의 비율이 읍면 지역>중소 도시>대도시의 순으로 높음에 비해, 일반 국민을 대상으로 한 본 연구의 결과에서는 중소 도시>읍면 지역>대도시의 순으로 그 비율이 높았다.[18]

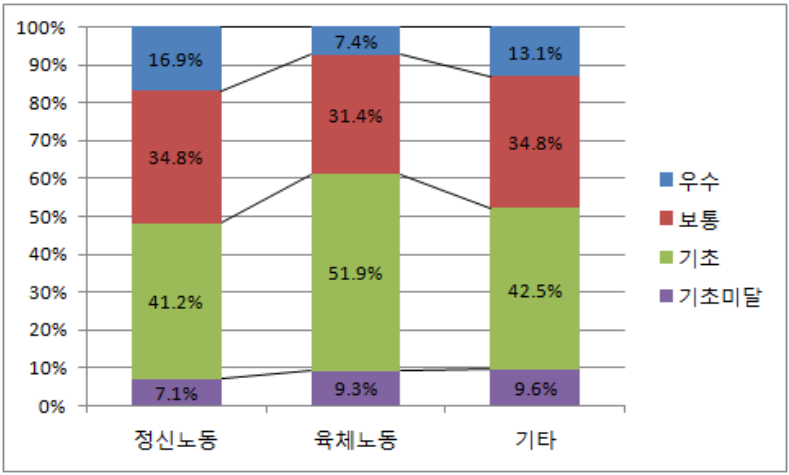
4) 직업

정신노동, 육체노동, 기타의 직업군별로 조사한 직업군별 조사 결과를 분석해 보면, 유의도가 0.05보다 작아 통계적으로 유의한 차이가 있는 것으로 확인된다. 사후 비교를 위한 터키 검정 결과, 정신노동자가 육체노

18) 기초 등급과 기초 미달 등급을 합하여 그 비율을 비교하였다.

동자에 비해, 기타가 육체노동자에 비해 국어능력이 우수함을 통계적으로 확인할 수 있었다.[19] 이러한 직업군별 차이를 수준 등급으로 구획하여 비교해 보면 <그림 6>과 같다.

<그림 6> 국어능력 수준 등급의 직업군별 분포

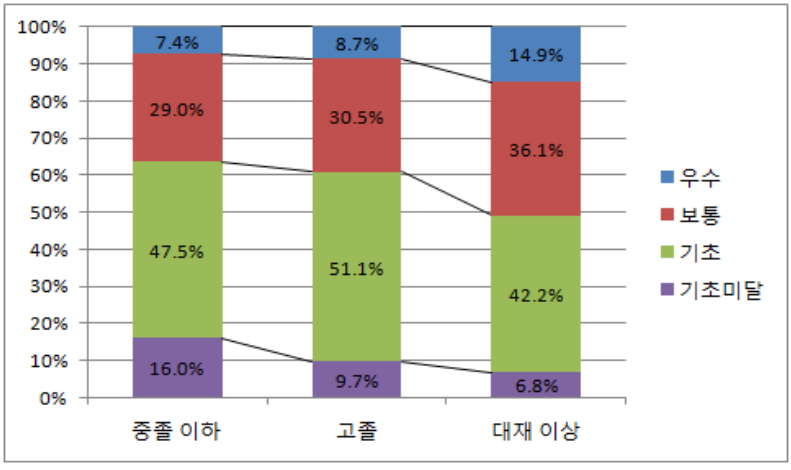


5) 학력

마지막으로 중졸 이하, 고졸, 대재 이상의 학력군별 국어능력 역시 유의도가 0.05보다 작아 통계적으로 유의한 차이가 있는 것으로 확인되었다. 또한 사후 비교를 위한 터키 검정 결과, 대재 이상이 중졸 이하보다, 대재 이상이 고졸보다 국어능력이 우수함을 통계적으로 확인할 수 있었다. 이러한 학력군별 차이를 수준 등급으로 구획하여 비교해 보면 <그림 7>과 같다.

19) 기타 직업군의 점수가 높은 것은 이 집단 안에 '학생'이 포함되어 있었기 때문으로 추정된다.

<그림 7> 국어능력 수준의 학력군별 분포



IV. 논의 및 제언

1. 본 연구의 기대 효과

본 연구는 2013년 현재 국민의 국어능력 수준을 점검하는 데에서 그치는 것이 아니다. 이러한 조사 결과를 바탕으로 하여 관련된 국어 정책을 입안하고 일반 국민 대상의 국어교육을 시행하는 데 보다 궁극적인 목적을 두고 있다. 이에 따라 본고에서는 본 연구를 통해 나타난 문제점을 보완할 수 있는 국어교육 프로그램의 개발에 대한 상(像)을 제시해 보고자 한다.

먼저, 일반 국민 대상의 국어교육 프로그램 개발을 온라인 교육과 오프라인 교육으로 나누어 생각해 볼 수 있다. 온라인 교육은 국립국어원의 홈페이지에 탑재된 '온라인 강의'의 형태로 그 내용을 구성하고, 오프라인

교육은 '국어문화 학교'를 참조하여 개발이 가능할 것으로 보인다. 다만, '쓰기' 영역과 '말하기' 영역은 온라인 교육보다는 오프라인 교육이 더 효과적일 것이며, '읽기'나 '듣기' 영역은 온라인 교육만으로도 충분히 효과적인 구성이 가능해 보인다. '문법' 영역은 온라인으로도 교육이 가능하나 충분한 설명이 요구되는 측면이 있어 온라인과 오프라인에서 서로 보완해 주는 차원으로 구성될 필요가 있다.

위와 같은 교육 프로그램들은 이후에 이루어질 후속 사업으로서 지속적으로 개발이 필요하다. 교육 프로그램을 구성할 교육 내용을 마련하는 일은 각 영역별 결과에 대한 세부적인 분석 작업이 이루어지면서 가능해질 것이다.

2. 후속 연구를 위한 제언

본 연구는 국민의 국어능력을 점검하기 위해 기초 연구에서 마련한 조사 도구를 다듬어 정비하고, 조사를 계획하여 실행한 뒤 그 결과를 분석하는 것까지를 연구의 범위로 삼고 있다. 이에 후속 연구에서는 이러한 조사 결과를 바탕으로 영역별, 변인별 심층 분석이 이루어져 국어 정책 수립을 위한 실증적 자료들을 도출해 낼 필요가 있다. 이에 따라 본고에서는 예상 가능한 후속 연구를 제안하는 것으로 결론을 갈음하고자 한다.

1) 각 영역의 하위 요소별 분석

본 연구에서 각 영역의 조사 도구는 일정한 조사 요소(평가 요소)의 설정에 기반하여 설계되었다. 말하기 영역은 '설득, 설명'으로, 듣기 영역은 '사실적 듣기, 추론적 듣기, 비판적 듣기'로, 읽기 영역은 '사실적/축자적 읽기, 추론적 읽기, 비판적 읽기'로, 쓰기 영역은 '친교적 쓰기, 설명적 쓰

기, 논증적 쓰기'로, 문법 영역은 '규범, 어휘, 문장'으로 조사 요소를 설정한 것이다. 따라서 각 조사 요소별로도 심층적으로 조사 결과를 분석할 필요가 있다.

2) 각 영역 사이의 상관관계 분석

본 연구에서는 원 점수와 척도 변환 점수를 기준으로, 각 영역별 상관관계를 일정 정도 분석하였다. 그러나 보다 심층적이고 구체적인 분석 결과 도출을 위해서는 보다 세분화된 영역별 상관관계 분석이 요구된다.

3) 쓰기, 말하기 영역에 대한 심층 분석

쓰기와 말하기 영역은 다른 영역에 비해 점수가 현저히 낮게 나왔다. 이를 통해, 우리 국민의 표현 능력이 상대적으로 저조하다는 해석이 가능한데, 이러한 조사 결과를 다층적으로 분석하여 그 원인과 대처 방안을 탐색해야 한다. 예를 들어 구체적인 국어교육 정책 내용을 마련하기 위해서는 쓰기와 말하기 영역에서 설정한 하위 요소별로 조사 결과를 분석하는 과정과 더불어 이미 확보된 대규모의 말하기와 쓰기 결과물에 대한 질적 분석을 통해 보다 심층적이고 구체적으로 조사 결과를 분석할 수 있어야 한다.

4) 변인별 격차 해소 방안 탐색

본 연구에서 상정한 변인들에 대한 차이의 원인을 심층적으로 분석할 수 있는 연구가 후속되어야 한다. 이때의 원인 분석은 단순히 추정을 통해서만이 아니라, 실태 조사 및 다른 연구 결과와의 비교 과정 등을 통해 세심하게 접근하여 보다 타당한 결과를 도출하고, 이를 토대로 이러한 격

차를 해소할 수 있는 정책적·교육적 방안을 마련해야 할 것이다.

5) 맥락 변인에 대한 심층적 조사 연구

본 연구에서는 인구학적 통계 정보의 변인만을 설정하였지만, 실상 국어능력과 관련되는 다양한 맥락 변인에 대한 연구가 후속되어야 이러한 조사 결과를 보다 심층적이고 실증적으로 분석할 수 있다. 예를 들어, 각 영역, 수준, 변인별 집단 등에서 대표성을 띠는 소수의 피조사자들을 대상으로 면담, 관찰 등의 질적 조사를 병행하는 방식도 고려할 수 있다.

6) 조사 결과에 대한 데이터베이스화

본 연구를 통해 확보한 다양한 자료들을 데이터베이스화하여 지속적으로 관리할 수 있는 방안을 마련해야 한다. 또한, 이들 자료를 학계와 공유하여 보다 다양한 결과 분석과 이에 대한 해석을 이끌어 낼 필요가 있다. 예컨대, 한국교육과정평가원에서는 대학 수학 능력 시험이나 국가 수준 학업 성취도 평가 등의 대규모 데이터들을 공모를 통해 선정된 일부 연구자들에게 제공함으로써 보다 심층적인 자료에 대한 분석을 해 오고 있다. 또한, 국제 성인역량 조사(Programme for the International Assessment of Adult competencies, PIAAC)[20]에서는 홈페이지를 통해 해당 조사를 홍보함은 물론, 데이터를 공유하고 있다.*

---

20) 24개 국가 혹은 국가의 일부 지역에서 16세에서 65세에 이르는 15만 7천 명의 성인을 대상으로 조사하는 것으로, 우리나라에서는 교육부 평생 학습 정책과와 고용부 직업 능력 개발 정책과의 협업 사업으로 진행하고 국내 조사 및 연구 총괄은 한국직업 능력 개발원이 조사 실행은 통계청에서 담당한다.

* 이 논문은 2014. 4. 30. 투고되었으며, 2014. 5. 13. 심사가 시작되어 2014. 6. 4. 심사가 완료되었고, 2014. 6. 5. 편집위원회 심의를 거쳐 게재가 확정되었음.

참고 문헌

김동환 외(2008), 『국어능력 검정 모의 시험 방법 연구』, 국립국어원.
김문오 외(2005), 『국민의 국어능력 평가를 위한 기초 연구』, 국립국어원.
김성규 외(2006), 『국어능력 조사 방법 연구 1』, 국립국어원.
김종철 외(2013), 『2013년 국민의 국어능력 평가』, 국립국어원.
김창원 외(2008), 『국민의 기초 문해력 조사』, 국립국어원.
박미영(2013), 「성인 대상 국어 재교육의 현황과 전망-국립국어원 국어문화학교 운영을 중심으로-」, 『작문연구』 19, 한국작문학회, pp.37-67.
박재현 외(2008), 『국군 장병의 국어능력 실태 조사』, 국립국어원.
송미영 · 임해미 · 최혁준 · 박혜영 · 손수경(2013), 「OECD 국제 학업 성취도 평가 연구: PISA 2012 결과 보고서」, 한국교육과정평가원.
양명희(2013), 「문해교육의 개념과 내용 분석 연구」, 『인문연구』 67, pp.319-348.
윤여탁 외(2009), 「교사의 국어능력 실태 조사」, 국립국어원.
이현희 · 박철우(2013), 「성인의 어휘 능력과 텍스트 이해의 상관관계에 대한 기초 연구」, 『사회언어학』 21-3, 한국사회언어학회, pp.249-282.
이희수 · 한유경 · 박현정 · 이세정 · 이정희 · 권재현(2001), 「한국 성인의 문해 실태 및 OECD 국제 비교 조사 연구」, 한국교육개발원.
임언 · 최동선 · 오은진(2005), 「한국의 성인 직업 기초 능력」, 한국직업능력개발원.
장경희 외(2009), 『국어능력 검정 시험에서 말하기 능력 평가 방법 개발을 위한 연구』, 국립국어원.
장소원 외(2012), 『국어능력 평가를 위한 기초 연구』, 국립국어원.

■국문초록

국민의 국어능력 조사 결과 분석

김종철 · 김봉순 · 박재현 · 박현정 · 정희창 · 주재우
박은진 · 이관희 · 이지수 · 정재미 · 민병곤

본 연구의 목적은 2013년 현재 우리 국민의 국어능력이 어느 수준에 있는지를 진단하고, 향후 5년을 주기로 이러한 조사를 정례화하기 위한 체재를 구축하는 데 있다. 연구의 조사 결과를 통해 도출된 실증적이고 구체적인 자료는 관련 국어 정책을 입안하고 일반 국민 대상 국어교육의 틀을 세우는 데 바탕이 될 것이다.

이를 위해 먼저, 장소원 외(2012)의 기초 연구에서 마련한 조사 도구를 보완하여 400명을 대상으로 예비 조사를 실시하여 문항의 난이도와 변별도를 점검하였다. 그런 다음 2단계 층화 추출(stratified sampling) 방식과 집락 추출(cluster sampling) 방식으로 조사 대상을 표집하였으며, 성, 연령대, 지역 규모, 학력군, 직업군의 5개 변인을 설정하여 각 변인별로 실제 인구 분포 비율에 비례하게 조사 대상이 표집될 수 있도록 하였다.

조사 결과는 문항 반응 이론(Item Response Theory) 모형 중 2모수 로지스틱 모형을 적용하여 척도 점수(scale score)를 산출하였다. 이에 더해, 조사 결과를 절대적인 지표로 해석하기 위해, 우수, 보통, 기초, 기초 미달로 국민의 국어능력 수준을 등급화하였다.

본 조사는 2013년 9월 1일부터 11월 15일까지 약 10주에 걸쳐 20세~59세의 성인 남녀 3,000명을 대상으로 시행되었다. 조사 대상 국민의 전체 국어능력 평균 점수는 579.62점(1000점 만점)으로 분석되어 우리 국민의 전체 국어능력은 평균적으로 '보통 등급과 기초 등급의 경계선'의 수준인 것으로 나타났다. 영역별로는 말하기와 쓰기 표현 능력이 다른 영역에 비해 낮게 나타났다.

[주제어] 국어능력, 국어능력 평가, 국민의 국어능력, 듣기 능력, 말하기 능력, 읽기 능력, 쓰기 능력, 문법 능력, 척도 점수, 앵고프

■ Abstract

Analysis of Evaluation of Korean Language Proficiency for the General Public

Kim, Jong-cheol · Kim, Bong-soon · Park, Jae-hyun
Park, Hyun-jeong · Jeong, Hui-chang · Joo, Jae-woo
Park, Eun-jin · Lee, Kwan-hee · Lee, Ji-su
Jung, Jae-mi · Min, Byeong-gon

The present study has conducted to evaluate Korean Language Proficiency for the general public on five sections, and to construct the system of regularizing this kind of quinquennial survey. On the basis of the empirical and concrete data from the result of this survey, we expect to plan the relevant Korean language policy and to set up a frame of Korean language education for the general public.

We have sampled evaluation subject in proportion to the ratio of actual population distribution by controlling 5 factors; sex, age, scale of the area, education background, occupational cluster. The main evaluation was conducted for about 10 weeks (9.1.13~11.15.13) for 3000 adult males and females. As a result of the investigation, general public have a low level in speaking and writing section.

[Key words] Korean language proficiency, evaluation of Korean language proficiency, Korean language proficiency of the general public, listening proficiency, speaking proficiency, reading proficiency, writing proficiency, grammar proficiency, scale score, Angoff