

국립국어원 2022-01-37

발간등록번호
11-1371028-000927-01

국민의 국어능력 실태 조사 개선 기초 연구

연구책임자
구 본 관

2022. 12. 16.

제 출 문

국립국어원장 귀하

국립국어원과 체결한 연구용역 계약에 따라 ‘국민의 국어능력 실태 조사 개선 기초 연구’에 관한 연구 보고서를 작성하여 제출합니다.

■ 사업 기간: 2022년 7월 ~ 2022년 12월

2022년 12월 16일

연구책임자: 구본관(서울대학교 국어교육과)

연구 기관: 서울대학교 국어교육연구소

연구책임자: 구본관

공동연구원: 민병곤, 이관희, 김승현, 박성석, 이성준, 최선희

보조연구원: 강지영, 김예원, 이수진, 나상수

조사 수행 기관: KANTAR PUBLIC

국민의 국어능력 실태 조사 개선 기초 연구

본 연구는 2023년부터 시행될 <제3차 국민의 국어능력 실태 조사>를 위해 기존 1·2차 실태 조사의 성과를 평가하여 실태 조사 고도화를 위한 계획을 수립하고, 듣기·말하기 영역의 조사 도구를 개발함과 동시에 조사 결과의 활용 방안을 마련하는 데 목적이 있다. 이를 위해 ‘① 성과 평가 및 현황 분석, ② 듣기·말하기 영역 도구 개발, ③ 조사 도구의 정교화, ④ 본 조사 시행안 제언’이라는 네 가지 세부 연구 목표를 설정하여 국민의 국어능력 실태 조사의 영역별·연차별 고도화 방안을 구체화하고, 변화된 국어 사용 환경에 따른 영역별 국어능력의 범위 및 개념을 정립하며, 조사 시행을 위한 이론적, 방법론적 기틀을 마련하고자 하였다.

제3차 국민의 국어능력 실태 조사는 1차(2013), 2차(2018) 실태 조사에서 이어지는 일련의 연속성과 단계성 속에 위치한 국가 차원의 대규모 조사로, 본 연구는 이러한 제3차 실태 조사 실시를 위한 기초 연구이자 예비 조사의 성격을 띤다. 즉 정례적 조사 체계를 마련한 1차 실태 조사와 조사 도구 개발 및 안정화를 시도한 2차 실태 조사에서 나아가 국어능력 영역별 연차 조사의 고도화 방안을 수립하고자 하는 연구로서, 분할 조사의 첫 번째 영역인 ‘듣기·말하기’ 영역에 무게를 두고 있다.

먼저 국어능력 영역별·연차별 실태 조사 고도화 방안을 수립하기 위해 국어 사용 환경 및 현황을 분석하여 국어능력의 개념을 수립하고, 자국어능력, 듣기·말하기 능력, 디지털 리터러시와 관련된 국내외 실태 조사 사례를 검토해 평가 영역 및 요소, 조사 대상 및 방법 등에 대한 시사점을 도출하였다. 특히 국제성인역량조사(PIAAC), 국제학업성취도평가(PISA) 등을 포함한 여러 조사에서 변화하고 있는 문식 환경을 반영하기 위해 디지털 리터러시를 강조하고 있는바, 본 연구에서도 이러한 지점을 적극적으로 고려하여 디지털 리터러시 관련 요소를 조사 문항의 한 내용 요소로 구성하거나 차후 별도의 조사 영역으로 설계하는 방안을 제안하고, 배경 변인 설문문을 통해 조사 결과에 대한 다각적 해석의 가능성을 확보해 두어야 함을 제시하였다.

이상의 성과 및 현황 분석 결과를 바탕으로 ‘듣기·말하기’ 조사 도구는 다음과 같이 개발되었다. 우선 듣기·말하기 능력 실태 조사는 기본적으로 2018년 실태 조사와의 일관성을 유지하기 위해 기존의 듣기와 말하기 조사 틀을 준용하되 구어 의사소통의 실재성과 협력적 소통의 필요성을 강화하는 방향으로 설계되었다. 듣기·말하기 조사 영역은 크게 ‘듣기’, ‘듣기·말하기’, ‘말하기’ 영역으로 구분되어 각각의 영역 특성에 맞게 세부 조사 영역 및 요소가 설정되었다. 구체적으로 설명하자면, 듣기 영역의 경우 기존의 사실적, 추론적, 비판적 듣기에 더하여 공감적 듣기가 중영역 평가 항목에 추가되었다. 말하기 영역의 경우 실제 언어 사용에서 차지하는 담화의 비중 및 사회적 중요성, 매체 환경의 변화 등을 두루 고려하여 기존의 ‘설득하는 말하기(주장하는 말, 요청하는 말)’에 있던 ‘요청하는 말’을 제하고, ‘설명하는 말하기(경험 또는 개념을 설명하는 말)’에 ‘프레젠테이션’을 추가 구성하는 방식으로 수정·보완되었다. 마지막으로 듣기와 말하기를 연계한 듣기·말하기 영역은 언어의 통합성과 구어 담화의 역동성을 반영하기

위해 일상 대화에 자주 나타나는 친교적 대화에 초점 맞추어 중영역 ‘친교 표현하기’를 신설하고, 그 하위 조사 영역으로 ‘배려하며 말하기’와 ‘상호교섭하기’를 설정하였다.

차년도 듣기·말하기 영역 본 조사는 기수행된 ‘2018년 국민의 국어능력 실태 조사’의 영역별 문항 수, 시간, 배점 등을 고려하여 듣기 15문항(가교 2, 신규 13), 듣기·말하기 2문항(신규 2), 말하기 3문항(가교 1, 신규 2)으로 진행될 계획이다. 이에 1차 예비 조사에는 가교 문항을 포함하여 본 조사에서 사용할 문항의 2배수를 개발하였으며, 이 중에서 1차 예비 조사 분석 결과에 근거하여 1.5배수 문항을 선별하고 이를 보정해 2차 예비 조사 도구로 활용하였다. 이후 2차 예비 조사 결과를 바탕으로 1.2배수 문항을 최종 확정하였다. 연구 진행 과정에서 개발·선별된 문항은 (가), (나)형 검사지로 구성되어 서울, 인천, 경기도에 거주하는 20~60대 남녀 총 480명(1차: 240명, 2차: 240명)을 대상으로 듣기·말하기 1·2차 예비 조사에 활용되었다. 선택형 문항으로 구성된 듣기 영역은 태블릿 PC 화면에서 특정 보기를 터치해 CAPI 시스템에 정답을 직접 입력하는 방식으로 응답 자료가 수집되었으며, 수행형 문항으로 구성된 듣기·말하기, 말하기 영역은 응답을 녹음하는 방식으로 자료가 수집되었다.

수행형 문항에 대한 채점은 1·2차 예비 조사에서 각각 3,600건(듣기·말하기: 1,440건, 말하기: 2,160건)을 대상으로 이루어졌으며, 채점자 간 신뢰도 확보를 위해 각 문항별로 국어교육 전공자 3인이 공동 채점을 할 수 있게 설계되었다. 그리고 척도 사용 수준에 대한 채점자 간 눈높이를 맞추기 위해 웹 기반 훈련(채점자 워크숍)이 1·2차 예비 조사에서 각 1회씩 실시되었다. 영역별 문항 곤란도 및 변별도는 듣기 영역(선택형)의 경우 2모수 문항반응이론 모형을 적용하여 분석하였으며, 듣기·말하기, 말하기 영역(수행형)의 경우 1차 예비 조사에서는 다국면라쉬모형 분석을 통해 문항 양호도 및 채점 특성을 파악하였으며, 2차 예비 조사에서는 2모수 문항반응이론 채점자 효과 모형을 적용한 분석을 통해 객관화된 문항양호도를 확인한 후 최종 문항을 선정하였다. 채점자 간 신뢰도 분석 결과, 1·2차 예비 조사 모두에서 급내 상관이 듣기·말하기 영역은 0.7 이상의 양호한 수준으로, 말하기 영역은 0.8 이상의 우수한 수준으로 나타났다.

각 검사 영역별로 피험자들의 t점수 분포를 보면, 1차 예비 조사(9~10월)의 경우 점수 분포는 세 영역 모두 정규분포에 가깝게 나타났다. 다만 듣기, 듣기·말하기 영역은 오른쪽으로 약간 치우친 부적 편포를, 말하기 영역은 왼쪽으로 약간 치우친 정적 편포의 모습을 보였다. 영역별 문항 곤란도 및 변별도는 듣기 영역의 문항 대부분이 양호한 경향을 보였으나 일부 문항에서 .35 미만으로 변별도가 거의 없거나 곤란도가 -2.0 미만으로 지나치게 쉬운 문항이 있는 것으로 나타나 이를 수정·삭제하는 방향으로 1.5배수 문항을 선별하였다. 듣기·말하기 영역은 전반적으로 채점 척도의 적용이 안정적으로 나타났다. 그러나 (나)형 2번 문항의 경우 채점 척도의 사용 면에서 상대적으로 일관성이 부족한 양상이 나타났고, 다른 문항이 비해 문항 곤란도와 문항 변별도가 낮게 나타났다. 게다가 채점자 훈련 과정에서도 ‘내용 선정’ 채점 요소에 대한 조정이 있었던 점을 고려하여 해당 문항을 2차 예비 조사 문항에서 배제하였다. 말하기 영역의 경우 변별도와 곤란도 사이에 역전 현상이 나타난 (가)형 3번 문항을 1.5배수 문항에서 제외하였다. 이 과정에서 해당 문항의 토론 주제(교육 목적 체벌)가 특정 연령대 혹은 직업군에 유불리 문제를 발생시킬 수 있다는 가능성도 함께 고려되었다.

다음으로 2차 예비 조사(11~12월) 결과를 살펴보면, 점수 분포는 1차 예비 조사 결과에서와

마찬가지로 세 영역 모두 정규분포에 가깝게 나타났다. 듣기 영역은 오른쪽으로 약간 치우친 부적 편포를 보였지만 정규성 검정 결과 유의확률 .064로 정규성을 만족하였다. 듣기·말하기 영역도 부적 편포가 나타나지만 정규분포에 거의 근접한 모양을 보였으며, 말하기 영역은 왼쪽으로 약간 치우친 다소 뾰족한 모양의 정적 편포를 보였다. 영역별 문항 곤란도 및 변별도 측면에서는 듣기 문항 대부분이 양호하였지만 일부 어려운 변별도와 곤란도를 보이는 문항이 나타났다. 변별도와 곤란도에 특별한 문제가 없는 경우 다양한 곤란도로 문항을 구성하는 데 초점을 두고 본 조사 문항 15개와 예비 문항 3개(총 18개)의 1.2배수 문항을 선별하였다. 듣기·말하기 영역은 전반적으로 낮은 곤란도와 양호한 변별도가 나타났다. 검사지 (나)형 준거의 곤란도 모수가 (가)형에 비해 평균에 가까운 분포를 보이고 있어 (나)형 문항을 1.2배수 최종 문항으로 선정하였다. 그리고 말하기 영역은 문항 2의 변별도와 곤란도가 (가), (나)형 모두에서 양호한 수준으로 나타나 모두 1.2배수 문항에 포함시키기로 결정하였다.

이상의 조사 결과를 토대로 본 연구에서 제안하는 몇 가지 논의점을 제시하면 다음과 같다. 첫째, 듣기·말하기 영역 본 조사 방안 수립과 관련된 부분이다. 조사 도구 구성 면에서 단일 조사 도구의 크기는 본 연구의 예비 조사와 유사하게 구성할 수 있으나 단일 조사 도구를 투입할 것인지, 이원화된 조사 도구를 투입할 것인지에 대한 추가적인 논의가 필요하다. 아울러 3, 4년 차 조사 시행 방식과 연령대를 고려한 조사 시간, 피조사자의 영역별 응시 범위, 그리고 인공지능 진단 체계 활용 여부 등에 대해 추가 논의가 필요하다. 조사 인원 및 방식 면에서는 국어능력 전체가 아닌 국어능력 각 영역별로 조사가 시행된다는 점을 감안하여 표본 수를 줄이는 방안을 제안하였다. 또한 조사 환경 및 피조사자들의 참여도 향상을 위해 기존 1:1 방문 조사 방식에서 내방형 집단 조사로 전환하는 방안을 제안하였으며, 이에 대한 적절한 유인책 마련이 필요함을 언급하였다. 그리고 성취 수준 등급화의 측면에서 2013년 1차 실태 조사에 적용했던 앵고프(Angoff) 방식을 활용해, 가교 문항을 바탕으로 한 이전 검사와의 동등화 및 분할 점수 산출을 시도할 필요가 있다.

둘째, 5년 주기 국어능력 조사의 정례화 방향 및 개선안 수립과 관련된 부분이다. 본 연구는 1) 분할 조사를 기조로 하여 국어능력 ‘영역별 연차 조사 안’을 제안하였다. 그리고 2) 국어능력 영역별, 연차별 실태 조사 고도화 방안 계획을 수립하는 과정에서 ‘인공지능 활용 국어능력 진단 체계 개발’ 과업과의 연계안을 중장기적 계획의 일환으로 고안하고자 하였다. 이에 더하여 3) 세분화된 영역별 심층 조사 및 환류 정책의 수립을 가능하게 하기 위해 후속 연구를 통한 실효성 있는 정책 수립안을 구상하였으며, 4) 인구통계학적 변인, 영역별 심층 변인과 연계하는 조사 결과 분석 및 해석을 제안하였다.

주요어: 국어능력, 국어능력 조사, 국민의 국어능력, 듣기 능력, 말하기 능력, 디지털 리터러시

차 례

제1장 서론

1.1. 연구의 목적 및 필요성	3
1.2. 연구의 범위	7
1.3. 연구의 추진 과정	8

제2장 국민의 국어능력 실태 조사 고도화를 위한 이론적 검토

2.1. 1, 2차 국민의 국어능력 실태 조사 성과 평가	13
1) 조사 도구 측면에서의 성과 분석	13
2) 조사 인원 측면에서의 성과 분석	14
3) 조사 방식 측면에서의 성과 분석	15
4) 조사 결과 해석 측면에서의 성과 분석	15
2.2. 국내외 국어능력 및 디지털 리터러시 실태 조사 사례 검토	16
1) 국내의 국어능력 실태 조사 사례	16
(1) 2018년 국민의 국어능력 실태 조사	16
(2) 2020년 국민의 언어의식 조사	17
(3) 2008년 국민의 기초 문해력 조사	17
(4) 국어능력인증시험(ToKL)	18
(5) KBS 한국어능력시험	18
(6) 한국어능력시험(TOPIK)	19
2) 국외의 자국어능력 실태 조사 사례	20
(1) 미국	20
(2) 영국	22

차 례

(3) 중국	24
(4) 일본	26
(5) 유럽	28
(6) 기타 언어능력 측정 사례	30
3) 국내외 듣기·말하기 능력 실태 조사 사례	33
(1) 듣기·말하기 능력을 포함하여 국어능력 제반을 조사한 경우	33
(2) 듣기·말하기 능력에 초점화하여 조사한 경우	40
4) 국내외 디지털 리터러시 실태 조사 사례	45
(1) PISA	45
(2) PIAAC	48
2.3. 디지털 리터러시 역량 실태 조사 방안 수립	51
1) 국어 사용 환경 및 현황 분석	51
2) 디지털 리터러시의 개념 및 조사 방향	54
(1) 디지털 리터러시의 개념	54
(2) 디지털 리터러시의 기존 조사 문항 분석	58
3) 디지털 리터러시 영역 조사를 위한 시사점	69
2.4. 국어능력 영역별·연차별 실태 조사 고도화 방안 계획	70
1) 5년 주기 실태 조사 사업의 연차별 정례화 방향	70
2) 연차별 조사 및 연구 계획	72
(1) 1안	72
(2) 2안	78
(3) 인공지능 활용 연계 방향	80

차 례

제3장 조사 도구의 개발

3.1. 조사 도구 개발의 절차	83
3.2. 영역별 조사 도구의 구성	85
1) 듣기 영역	85
(1) 세부 조사 영역 및 요소 설정	85
(2) 가교 문항 검토	88
(3) 문항 형식	92
2) 듣기·말하기 영역	93
(1) 세부 조사 영역 및 요소 설정	93
(2) 문항 형식	97
3) 말하기 영역	98
(1) 세부 조사 영역 및 요소 설정	98
(2) 가교 문항 검토	101
(3) 문항 형식	103

제4장 1차 예비 조사 시행 및 결과 분석

4.1. 1차 예비 조사 설계	107
1) 1차 예비 조사를 위한 검사지 구성	107
(1) 영역별 검사지 구성	107
(2) 영역별 검사 문항의 실제	107
2) 1차 예비 조사의 설계	108
(1) 모집단 및 표본 수	108
(2) 표본 수 배분	109
(3) 모집단 총화와 표본 수 배분	109

차 례

(4) 집락(읍·면·동) 추출과 응답자 추출	109
(5) 조사 대상자(개인) 추출 방법	110
(6) 표본 구성	112
4.2. 1차 예비 조사 진행	116
1) 조사 방법	116
2) 조사원 선발 및 교육	116
3) 1차 예비 조사 진행 일정	117
4) 조사 검증	118
5) 결과물 제출	118
4.3. 1차 예비 조사 관련 특이 사항	119
1) TAPI 활용 관련	119
2) 실사 과정상의 문제점	119
4.4. 1차 예비 조사 결과	120
1) 1차 예비 조사 결과의 채점	120
(1) 선택형 문항(듣기)의 채점	120
(2) 수행형 문항(듣기·말하기, 말하기)의 채점	120
(3) 수행형 문항(듣기·말하기, 말하기) 채점의 채점자 신뢰도 및 영향 분석	127
2) 1차 예비 조사 결과 분석	130
(1) 검사 설계상의 특징	130
(2) 영역별 점수 분포 분석	131
(3) 영역별 문항 곤란도 및 변별도 분석	133
(4) 배경 변인별 분석	136

차 례

제5장 2차 예비 조사 시행 및 결과 분석

5.1. 2차 예비 조사 설계	143
1) 2차 예비 조사를 위한 검사지 구성	143
(1) 영역별 검사지 구성	143
(2) 영역별 검사 문항의 실제	143
2) 2차 예비 조사의 설계	143
(1) 모집단 및 표본 수	143
(2) 표본 수 배분	144
(3) 모집단 층화와 표본 수 배분	144
(4) 집락(읍·면·동) 추출과 응답자 추출	144
(5) 조사 대상자(개인) 추출 방법	145
(6) 표본 구성	146
5.2. 2차 예비 조사 진행	151
1) 조사 방법	151
2) 조사원 선발 및 교육	151
3) 2차 예비 조사 진행 일정	152
4) 조사 검증	153
5) 결과물 제출	153
5.3. 2차 예비 조사 관련 특이 사항	154
1) TAPI 활용 관련	154
2) 실사 과정상의 문제점	154

차 례

5.4. 2차 예비 조사 결과	154
1) 2차 예비 조사 결과의 채점	154
(1) 선택형 문항(듣기)의 채점	154
(2) 수행형 문항(듣기·말하기, 말하기)의 채점	154
(3) 수행형 문항(듣기·말하기, 말하기) 채점의 채점자 신뢰도 및 영향 분석	159
2) 2차 예비 조사 결과 분석	163
(1) 검사 설계상의 특징	163
(2) 영역별 점수 분포 분석	163
(3) 영역별 문항 곤란도 및 변별도 분석	166
(4) 배경 변인별 분석	169
3) 최종 도구의 확정	173

제6장 제언

6.1. 듣기·말하기 영역 본 조사 방안 수립	177
1) 조사 도구 구성	177
2) 조사 인원 및 방식	178
(1) 모집단 및 표본 수	178
(2) 층화 후 집락추출	178
(3) 조사 대상자 추출	178
(4) 조사 방법	179
3) 성취 수준의 등급화	180

차 례

6.2. 5년 주기 정례화 방안 수립	183
1) 5년 주기 국어능력 조사의 정례화 방향 및 개선안	183
2) 인공지능 활용 국민의 국어능력 진단체계 개발 사업과의 연계안	185
3) 후속 연구를 통한 실효성 있는 정책 수립	185
4) 국어능력의 전체 등급화 및 특화된 배경 설문과 연계한 분석	186
(1) 인구통계학적 변인에 따른 결과 분석	187
(2) 특화된 배경 설문과 연계한 결과 분석	188
 참고 문헌	 189
 Abstract	 193
 부록	 197

표 차례

<표 I-1> 연구의 좌표	3
<표 I-2> 연구 내용 범위	7
<표 I-3> 단계별 추진 일정 및 세부 사업 내용	8
<표 II-1> 1차 실태 조사(2013년)의 조사 도구 설계	14
<표 II-2> 2차 실태 조사(2018년)의 조사 도구 설계	14
<표 II-3> 1, 2차 실태 조사의 조사 인원	15
<표 II-4> 2018년 국민의 국어능력 실태 조사 영역, 방식, 문항 수, 시간, 배점	16
<표 II-5> 2013년 조사 도구와 2018년 조사 도구의 설계 측면에서의 비교	17
<표 II-6> 2008년 국민의 기초 문해력 조사의 내용 요소	18
<표 II-7> KBS 한국어능력시험 영역별 주요 내용	19
<표 II-8> 한국어능력시험(TOPIK)의 영역별 문항 구성 및 시험 시간 ..	19
<표 II-9> NAEP의 읽기 영역과 쓰기 영역의 문항 구성	21
<표 II-10> SAT의 영역별 문항 수 및 소요 시간	22
<표 II-11> 보통화능력시험의 문항 유형, 배점, 제한 시간	24

표 차례

<표 II-12> ‘한자응용능력시험’의 문항 유형, 평가 내용, 문항 수	26
<표 II-13> ‘일본어 문장 이해·작성 능력 검정 시험’의 등급 및 난이도	26
<표 II-14> ‘일본 한자 능력 검정 시험’의 등급 및 난이도, 대상 한자 수	27
<표 II-15> 일본어 검정 시험의 등급 구분	28
<표 II-16> 독일어 능력 시험의 평가 영역 및 소요 시간	28
<표 II-17> 독일어 능력 시험의 유형에 따른 평가 영역	29
<표 II-18> 독일어 능력 시험의 등급 구분	29
<표 II-19> 대입 독일어 능력 시험의 평가 영역 및 소요 시간	30
<표 II-20> OPIc의 평가 영역	32
<표 II-21> 국어능력인증시험의 영역별 발문 예시	34
<표 II-22> KBS 한국어능력시험의 준거 집단 대상 수행 과제	36
<표 II-23> KBS 한국어능력시험의 언어 수행 능력 데이터베이스(사례)	36

표 차례

<표 II-24> 케임브리지 대학의 영어 인증 시험의 영역별 문항 수 및 소요 시간	37
<표 II-25> 케임브리지 대학의 영어 인증 시험의 듣기 영역의 내용, 문항 형식, 문항 개수	38
<표 II-26> 케임브리지 대학의 영어 인증 시험의 말하기 영역의 내용 및 과제 형식	39
<표 II-27> 중국어능력시험의 말하기 영역의 제시문(사례)	40
<표 II-28> 보통화능력시험의 문항 유형 및 평가 요소	40
<표 II-29> 보통화능력시험의 채점 방식	41
<표 II-30> TOPIK 말하기의 문항 유형	43
<표 II-31> TOPIK 말하기의 예시 문항 및 모범 답안	44
<표 II-32> TOPIK 말하기의 평가 요소 및 내용	44
<표 II-33> PISA의 DRA 평가 틀(평가 대상 인지 요소)	46
<표 II-34> PISA의 주기별 평가 틀 비교	47
<표 II-35> PISA에서 진행하는 ICT 설문의 내용	48

표 차례

<표 II-36> PIAAC의 문식력, 적응형 문제 해결력의 평가 틀(OECD, 2021: 19)	49
<표 II-37> 적응형 문제 해결력의 평가 내용(예시)	50
<표 II-38> 전 국민(10세 이상)의 하루 시간 사용(통계청, 2020: 8)	51
<표 II-39> 성인 및 초·중·고교 학생의 연간 독서율 추이(백원근 외, 2020: 10-11)	53
<표 II-40> 디지털 리터러시의 하위 구성 개념과 그 정의(정현선 외, 2016: 217-218)	55
<표 II-41> 디지털 리터러시의 하위 구성 개념과 관련된 성취기준 (정현선 외, 2016: 221-222)	55
<표 II-42> 미디어 교육의 내용 영역 비교 및 종합(장은주, 2020: 47)	56
<표 II-43> 디지털 리터러시 관련 조사	58
<표 II-44> 디지털 리터러시 평가의 평가 틀	60
<표 II-45> PIAAC의 독서 리터러시 평가 틀(박주현, 2018: 66)	61
<표 II-46> PIAAC의 디지털 독서 리터러시 평가 문항	62

표 차례

<표 II -47> PIAAC의 디지털 문제 해결 능력 평가 틀(박주현, 2018: 67)	62
<표 II -48> PIAAC의 디지털 문제 해결 능력 평가 문항	63
<표 II -49> 디지털 문제 해결 능력 관련 PIAAC 문항 사례(2)	65
<표 II -50> PISA 2018의 독서 리터러시 평가 틀(박주현, 2021: 142 참고)	66
<표 II -51> PISA의 디지털 독서 리터러시 평가 문항	67
<표 II -52> 국어능력 실태 조사 고도화 방안 수립 과정	70
<표 II -53> 1, 2차 실태 조사 성과 분석을 바탕으로 한 3차 실태 조사 시사점	71
<표 II -54> 국어능력 실태 조사 고도화 방안(1안)	72
<표 II -55> 2018년 조사의 심층 변인 및 배경 설문지 내용	73
<표 II -56> 국어능력 실태 조사 고도화 방안(2안)	79
<표 II -57> 인공지능 사업과 국어능력 실태 조사의 연계안	80
<표 III -1> 본 연구의 조사 도구 개발 절차	83
<표 III -2> 2018, 2022 세부 조사 영역 및 요소	86

표 차례

<표 III-3> 2018년 본 조사 듣기 영역 담화 유형	87
<표 III-4> 2022년 예비 조사 듣기 영역 담화 유형	87
<표 III-5> 2023년도 듣기·말하기(신설) 조사 영역	94
<표 III-6> 2022년 제1차 예비 조사 듣기·말하기 영역의 문항 정보 ..	95
<표 III-7> 친교 표현하기(배려하며 말하기) 조사 요소 ①	96
<표 III-8> 친교 표현하기(상호교섭하기) 조사 요소 ②	97
<표 III-9> 듣기·말하기 문항의 형식	97
<표 III-10> 2018년도 조사 영역과 2022년도 조사 영역의 비교	98
<표 III-11> 2022년도 예비 조사 말하기 조사 영역	99
<표 III-12> 2022년 제1차 예비 조사 말하기 영역의 문항 정보	99
<표 III-13> 설명적 말하기(프레젠테이션) 조사 요소 ①	101
<표 III-14> 설득적 말하기(주장하는 말) 조사 요소 ②	101
<표 III-15> 말하기 문항의 형식	103
<표 IV-1> 영역별 검사지 구성	107
<표 IV-2> 1차 예비 조사 집락(읍·면·동) 추출 결과	110

표 차례

<표 IV-3> 직업 구분	111
<표 IV-4> 1차 예비 조사 성×연령대×지역별 목표 표본	112
<표 IV-5> 1차 예비 조사 성×연령대×지역별 조사 완료 표본	112
<표 IV-6> 1차 예비 조사 지역×직업별 목표 표본	113
<표 IV-7> 1차 예비 조사 지역×직업별 조사 완료 표본	113
<표 IV-8> 1차 예비 조사 지역×학력별 목표 표본	113
<표 IV-9> 1차 예비 조사 지역×학력별 조사 완료 표본	113
<표 IV-10> 1차 예비 조사 검사지 유형별 목표 표본	114
<표 IV-11> 1차 예비 조사 검사지 유형별 조사 완료 표본	115
<표 IV-12> 영역별 조사 진행 방법 및 지침	117
<표 IV-13> 1차 예비 조사 진행 일정	117
<표 IV-14> 1차 예비 조사 검증 결과	118
<표 IV-15> 수행형 문항의 영역 및 문항, 채점자 구성	121
<표 IV-16> 사례: 말하기 3번 문항(S3) 채점 기준 및 세부 채점 기준 ..	122
<표 IV-17> 듣기·말하기 1번 (가)형 문항 채점 척도	124

표 차례

<표 IV-18> 듣기·말하기 1번 (나)형 문항 채점 척도	124
<표 IV-19> 듣기·말하기 2번 (가)형 문항 채점 척도	125
<표 IV-20> 듣기·말하기 2번 (나)형 문항 채점 척도	125
<표 IV-21> 말하기 1번 가교 문항 채점 척도	126
<표 IV-22> 말하기 2번 문항 채점 척도	126
<표 IV-23> 말하기 3번 문항 채점 척도	127
<표 IV-24> 1차 예비 조사 듣기·말하기 영역의 채점자 간 신뢰도 ..	127
<표 IV-25> 1차 예비 조사 듣기·말하기 영역의 채점자 영향	128
<표 IV-26> 1차 예비 조사 말하기 영역의 채점자 간 신뢰도	129
<표 IV-27> 1차 예비 조사 말하기 영역의 채점자 영향	129
<표 IV-28> 듣기 영역 검사지별 문항 구성	130
<표 IV-29> 듣기·말하기, 말하기 영역 검사지별 문항 구성	130
<표 IV-30> 영역별 점수 분포의 기술통계치	131
<표 IV-31> 1차 예비 조사 듣기 영역 문항 곤란도 및 변별도 분석	133

표 차례

<표 IV-32> 1차 예비 조사 듣기·말하기 문항 양호도	134
<표 IV-33> 1차 예비 조사 말하기 문항 양호도	135
<표 IV-34> 1차 예비 조사 듣기 영역 배경 변인별 분석	136
<표 IV-35> 1차 예비 조사 듣기·말하기 영역 배경 변인별 분석	137
<표 IV-36> 1차 예비 조사 말하기 영역 배경 변인별 분석	139
<표 V-1> 2차 예비 조사 집락(읍·면·동) 추출 결과	145
<표 V-2> 직업 구분	146
<표 V-3> 2차 예비 조사 성×연령대×지역별 목표 표본	147
<표 V-4> 2차 예비 조사 성×연령대×지역별 조사 완료 표본	147
<표 V-5> 2차 예비 조사 지역×직업별 목표 표본	148
<표 V-6> 2차 예비 조사 지역×직업별 조사 완료 표본	148
<표 V-7> 2차 예비 조사 지역×학력별 목표 표본	148
<표 V-8> 2차 예비 조사 지역×학력별 조사 완료 표본	148
<표 V-9> 2차 예비 조사 검사지 유형별 목표 표본	149
<표 V-10> 2차 예비 조사 검사지 유형별 조사 완료 표본	150

표 차례

<표 V-11> 영역별 조사 진행 방법 및 지침	152
<표 V-12> 2차 예비 조사 진행 일정	152
<표 V-13> 2차 예비 조사 검증 결과	153
<표 V-14> 수행형 문항의 영역 및 문항, 채점자 구성	155
<표 V-15> 듣기·말하기 1번 (가)형 문항 채점 척도	156
<표 V-16> 듣기·말하기 1번 (나)형 문항 채점 척도	157
<표 V-17> 듣기·말하기 2번 (가)형 문항 채점 척도	157
<표 V-18> 말하기 1번 가교 문항 채점 척도	158
<표 V-19> 말하기 2번 문항 채점 척도	158
<표 V-20> 말하기 3번 문항 채점 척도	159
<표 V-21> 2차 예비 조사 듣기·말하기 영역의 채점자 간 신뢰도 ..	159
<표 V-22> 2차 예비 조사 듣기·말하기 영역의 채점자 영향	160
<표 V-23> 2차 예비 조사 말하기 영역의 채점자 간 신뢰도	161
<표 V-24> 2차 예비 조사 말하기 영역의 채점자 영향	162
<표 V-25> 듣기 영역 검사지별 문항 구성	163
<표 V-26> 듣기·말하기, 말하기 영역 검사지별 문항 구성	163

표 차례

<표 V-27> 영역별 점수 분포의 기술통계치	163
<표 V-28> 2차 예비 조사 듣기 영역 문항 곤란도 및 변별도 분석	166
<표 V-29> 2차 예비 조사 듣기·말하기 문항 양호도	167
<표 V-30> 2차 예비 조사 말하기 문항 양호도	168
<표 V-31> 2차 예비 조사 듣기 영역 배경 변인별 분석	169
<표 V-32> 2차 예비 조사 듣기·말하기 영역 배경 변인별 분석	170
<표 V-33> 2차 예비 조사 말하기 영역 배경 변인별 분석	172
<표 V-34> 최종 조사 도구 자문 및 검토 전문가	173
<표 VI-1> 2년 차(2023년) 듣기·말하기 본 조사의 조사지 구성 방식 ·	177
<표 VI-2> 2년 차(2023년) 듣기·말하기 본 조사 방식에 따른 예산안 ·	179
<표 VI-3> 2018년 조사에서의 전체 국어능력의 성취 수준 등급화 ·	181
<표 VI-4> 국민의 국어능력 실태 조사 고도화	183
<표 VI-5> 인공지능 사업과 국어능력 실태 조사의 연계안	185
<표 VI-6> 국내외 평가 사례의 설문 배경 변인	188

그림 차례

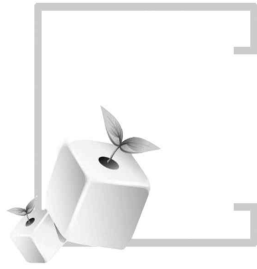
[그림 I -1] 세부 연구 목표	4
[그림 II -1] 케임브리지 대학의 영어 인증 시험(CEQ)의 영어 능력 단계	24
[그림 II -2] OPIc의 평가 영역	32
[그림 II -3] 국어능력인증시험의 문항 예시	35
[그림 II -4] 케임브리지 대학의 영어 인증 시험의 듣기 문항(예시)	38
[그림 II -5] OPIc의 문항 응답 화면(예시)	42
[그림 II -6] 1993~2020년 4대 매체 및 인터넷 이용률 추이 (한국언론진흥재단, 2020: 57)	52
[그림 II -7] 2016~2020년 인터넷 기반 미디어 이용률 추이 (한국언론진흥재단, 2020: 58)	53
[그림 II -8] 초·중등학교 국어과 교육과정 내 디지털 리터러시 관련 내용(노은희 외, 2018: 54)	57
[그림 II -9] 디지털 문제 해결 능력 관련 PIAAC 문항 사례(1)	64
[그림 II -10] 디지털 독서 리터러시 관련 PISA 문항 사례	67
[그림 II -11] 디지털 독서 리터러시 관련 PISA 문항 사례	68

그림 차례

[그림 II-12] 국어능력 실태 조사의 연차별 정례화	71
[그림 II-13] 한국교육과정평가원의 학업성취도 후속 연구 방식 결과 ..	75
[그림 II-14] 학업성취도 및 대학수학능력시험 자료 활용 분석 심포지엄	76
[그림 II-15] 한국교육과정평가원의 학업성취도 후속 연구 결과 목차	77
[그림 II-16] 국립국어원 국어 정책 자료 홈페이지	78
[그림 IV-1] TAPI 화면 예시 - 듣기 영역	108
[그림 IV-2] TAPI 화면 예시 - 듣기·말하기, 말하기 영역 공통	108
[그림 IV-3] 1차 예비 조사 조사원 지침서	116
[그림 IV-4] 1차 예비 조사의 듣기 영역 점수 분포	131
[그림 IV-5] 1차 예비 조사의 듣기·말하기 영역 점수 분포	132
[그림 IV-6] 1차 예비 조사의 말하기 영역 점수 분포	132
[그림 IV-7] 1차 예비 조사 듣기 영역 배경 변인별 분석	136
[그림 IV-8] 1차 예비 조사 듣기·말하기 영역 배경 변인별 분석	138
[그림 IV-9] 1차 예비 조사 말하기 영역 배경 변인별 분석	139
[그림 V-1] 2차 예비 조사 조사원 지침서	151

그림 차례

[그림 V-2] 2차 예비 조사의 듣기 영역 점수 분포	164
[그림 V-3] 2차 예비 조사의 듣기·말하기 영역 점수 분포	165
[그림 V-4] 2차 예비 조사의 말하기 영역 점수 분포	165
[그림 V-5] 2차 예비 조사 듣기 영역 배경 변인별 분석	169
[그림 V-6] 2차 예비 조사 듣기·말하기 영역 배경 변인별 분석	171
[그림 V-7] 2차 예비 조사 말하기 영역 배경 변인별 분석	172
[그림 VI-1] 2013년 조사에서의 영역별 및 전체 국어능력의 등급별 분포	181
[그림 VI-2] 2018년 조사에서의 영역별 및 전체 국어능력의 등급별 분포	182



제 1 장

서 론



1. 연구의 목적 및 필요성

국어기본법에서는 국민의 국어능력 향상과 국어의 발전 및 보전, 국어 환경의 개선을 위한 국어능력 조사의 필요성을 명시하고 있다. 제9조 실태 조사에 관한 법령에서는 “문화체육관광부장관은 국어 정책의 수립에 필요한 국민의 국어능력, 국어의식, 국어사용 환경 등에 관한 자료를 수집하거나 실태를 조사할 수 있다.”라고 규정하고 있으며, 제23조 국어능력의 검정에 관한 법령에서는 “문화체육관광부장관은 국민의 국어능력 향상과 창조적인 언어생활의 정착을 위하여 국어능력을 검정할 수 있다.”라고 언급하고 있다. 특히 국어 기본법 제6조 국어 발전 기본계획의 수립에 관한 법에서는 “문화체육관광부장관은 국어의 발전과 보전을 위하여 5년마다 국어 발전 기본계획(이하 "기본계획"이라 한다)을 수립·시행하여야 한다.”라고 규정하며 국어의 발전을 위한 국어능력 조사가 5년을 기점으로 주기적으로 이루어질 필요가 있음을 밝히고 있다.

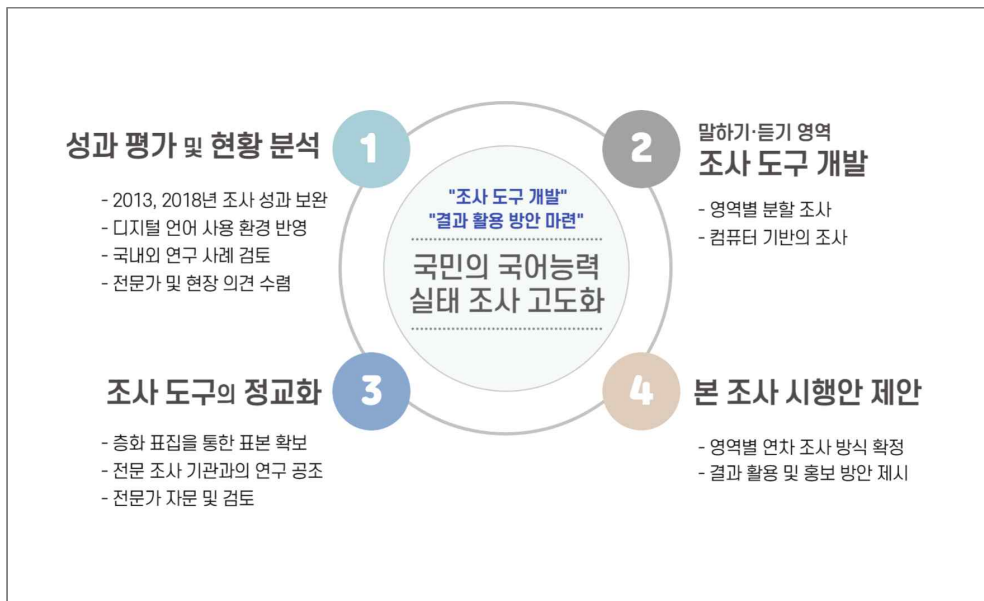
위 내용을 근거로 하여, ‘국민의 국어능력 실태 조사 개선 기초 연구’ 영역은 2023년부터 시행될 제3차 국민의 국어능력 실태 조사를 위해 기존의 1차 및 2차 실태 조사의 성과를 평가하여 실태 조사 고도화를 위한 계획을 수립하고, 말하기, 듣기 영역의 조사 도구 개발 및 조사 결과의 활용 방안을 마련하는데 연구의 목적이 있다. 즉, 본 연구는 <표 I -1>과 같이 ‘국어능력 평가를 위한 기초 연구(2012)’, ‘국민의 국어능력 평가(2013)’, ‘국민의 국어능력 실태 조사 도구 개발(2017)’, ‘국민의 국어능력 실태 조사(2018)’를 바탕으로 ‘2023년 국민의 국어능력 평가(듣기·말하기 영역)’로 이어지는 일련의 흐름 속에 놓여 있다.

<표 I -1> 연구의 좌표

1차 조사 정례적 조사 체계 마련	2013년 국민의 국어능력 평가(2013) ■ 평가 영역(5영역) 설정: 듣기, 말하기, 읽기, 쓰기, 문법 ■ 영역별 가교 문항의 설정 ■ 영역별 국어능력의 절대적 정도 측정, 국어능력 수준의 등급화	
↓		
2차 조사 조사 도구 개발 및 안정화	국민의 국어능력 실태 조사 도구 개발(2017) 평가 도구의 내용 문제 개선	2018 국민의 국어능력 실태 조사(2018) ■ 타당도, 변별도, 신뢰도 낮은 문항 재설계 ■ (문법/읽기) 디지털 문식력 측정 항목 보완 ■ 전국 단위의 표본 선정, 연령대 확대 ■ 국어능력 측정을 위한 변인 다양화 ■ 채점자 간 신뢰도 강화, 주관식 채점 정밀화 ■ 컴퓨터 기반의 평가 도구 구현 ■ 충분한 조사 시간 확보 및 사례 상향
	평가 도구 개발상의 문제 개선	
	평가 도구 구현 및 실행상의 문제 개선	
↓		
3차 조사 실태 조사 고도화	국민의 국어능력 실태 조사 개선 기초 연구(2022) ■ 영역별 ‘분할 조사’ 계획 수립: 듣기·말하기 영역 초점화 ■ 구어 의사소통의 실제성 강화 - 다양한 담화 유형 고려 - 듣기, 말하기 영역의 연계 고려	

- 동영상 활용 문항 구성 - 서답형 문항 개발
- 변화된 국어 사용 환경 반영 및 디지털 리터러시 강화

본 연구를 바탕으로 추후 2023년도에 시행될 듣기·말하기 영역의 국어능력 조사는 전 국민을 대상으로 실시되는 대규모 조사로, 그간의 국어능력 조사와는 달리 영역별 분할 조사의 첫걸음이라는 특징을 가진다. 따라서 본 연구는 국어능력에 대한 기존 연구(2005, 2006, 2008, 2009, 2012, 2013, 2017, 2018)들과의¹⁾ 연속성과 단계성을 중시하는 동시에 시대 변화를 반영한 국어능력의 변화를 측정할 수 있는 조사 도구를 개발하고, 2023년부터 시행될 국어능력의 영역별, 연차별 실태 조사 고도화 방안을 수립하고자 한다. 이때, 인공지능 활용 국민의 국어능력 진단 체계 개발 사업과의 연계 방안 역시 함께 마련한다. 이러한 연구 목적을 실현하기 위해서는 다음과 같은 세부 연구 목표가 단계적으로 달성되어야 한다.



[그림 I -1] 세부 연구 목표

[그림 I -1]에서 상정한 세부 연구 목표에 도달하기 위한 단계별 연구 추진 방향은 다음과 같다.

1) 세부 목표 ①: 성과 평가 및 현황 분석

본 연구는 5년 주기로 실시되는 국어능력 조사가 안정적이고 효율적으로 시행될 수 있는 기반을 구축하기 위하여 2013년, 2018년에 실시된 1, 2차 국민의 국어능력 실태 조사 성과를 평가하여 보완점을 마련하고, 2023년 3차 국민의 국어능력 실태 조사 계획 수립 시 주요 반영 사항을 도출한다. 또한 디지털 언어 사용 환경과 같이 변화된 국어 사용 환경에 따른 영역별 국어능력의 범위와 개념을 규정한다. 더하여, 국내외 국어능력 실태 연구 및 평가 사례, 국립국어원의 국어능력 실태 조사와 유사한 국어(언어) 능력 관련 연구 등에 대한 분석을 통해 국민의 국어능력의 영역별, 연차별 실태 조사 고도화 방안

1) 국립국어원(2005), 국민의 국어능력 평가를 위한 기초 연구, 국립국어원(2006), 국어능력 조사 방법 연구, 국립국어원(2008), 국어능력검정 모의시험 방법 연구, 국립국어원(2009), 국어능력검정시험에서 말하기 능력 평가 방법 개발을 위한 연구, 국립국어원(2012), 국어능력 평가를 위한 기초 연구, 국립국어원(2013), 국민의 국어능력 평가 연구.

계획을 수립한다.

2) 세부 목표 ②: 듣기·말하기 영역 도구 개발

영역별 분할 조사 계획에 따라 2023년에는 듣기·말하기 영역의 본 조사가 시행될 것인데, 이를 위해 본 연구에서는 안정적이고 효율적인 듣기·말하기 영역 조사 도구를 구안한다. 이때 2023년부터는 전 영역을 일시에 조사하는 방식이 아닌, 영역별로 분할하여 조사하는 방식을 택한다는 점에서 기존보다 조사 요소별 문항 수를 늘리고 검사지 종류를 세분화하여 영역에 따른 심층적인 분석을 수행할 필요가 있다. 특히 대화 상황에서 듣기와 말하기의 상호 교섭성과 역동성을 반영하고자 듣기·말하기 영역을 추가로 설정하여 실제성을 제고하고 실제의 의사소통 상황과 관련한 소통 능력을 측정할 필요가 있다.

3) 세부 목표 ③: 조사 도구의 정교화

예비 조사 및 전문가 자문을 실시하여 조사 도구를 정교화하고, 객관성 및 신뢰도를 제고한다. 또한 조사 도구의 정교화를 위해 2013, 2017, 2018의 연구 결과 및 시사점을 적극적으로 수용한다. 특히 말하기 영역에서는 실제 담화 상황 속에서의 말하기 능력을 평가할 수 있고, 디지털 리터러시를 측정할 수 있는 도구가 되도록 조사 도구를 정교화한다. 듣기 영역에서는 일상의 언어 소통 환경을 고려하여 매체를 다양하게 활용하여 담화 맥락의 실제성을 꾀할 수 있는 도구가 개발될 수 있도록 세부 보완 작업을 수행한다. 또한 언어의 통합성 및 구어 담화의 역동성을 고려하여 일상적인 대화 상황에서 청자와 화자의 역할 수행을 통해 듣기·말하기 능력을 평가할 수 있는 듣기·말하기 연계 문항이 본 연구를 통해 자리 잡을 수 있도록 조사 도구를 설계할 것이다.

4) 세부 목표 ④: 본 조사 시행안 제안

국내외 연구 및 국어 사용 현황 분석을 통해 수립한 국어능력의 영역별, 연차별 실태 조사 고도화 방안 계획을 기반으로 본 연구에서 실시한 예비 조사 결과를 반영하여 영역별 연차 조사 평가 방식을 구체적으로 확정한다. 또한 본 조사 결과를 다각도로 활용하여 유관 정책 수립에 반영하고, 이를 홍보할 수 있는 방안을 함께 제안한다.

이러한 세부 목표의 달성에 따라 본 연구가 가지는 의의는 다음과 같은 세 가지 방향에서 찾을 수 있다.

첫째, 국민의 국어능력 실태 조사의 영역별·연차별 고도화 방안을 구체화한다. 국민의 국어능력 실태 조사는 국어기본법 제23조에 따라 5년 주기로 시행되는 국가 차원의 국어능력 측정 조사이다. 본 연구는 1차(2013), 2차(2018) 국민의 국어능력 실태 조사에 대한 검토뿐만 아니라 국내외 국어능력 실태 조사 사례를 검토하여 앞으로 영역별·연차별로 이루어질 실태 조사의 시행 방안을 일정, 시기, 조사 방법 등으로 구체화하여 제안하고, 조사 결과를 다각도로 활용하고 홍보할 수 있는 방안을 상세화한다는 의의를 지닌다.

둘째, 본 연구는 국민의 국어능력과 관련된 국어 사용 환경 및 현황을 분석하여, 변화된 국어 사용 환경에 따른 영역별 국어능력의 범위 및 개념을 정립한다는 의의를 가진다. 또한 2023년에 본 조사를 실시할 2개 영역(듣기, 말하기)에 초점을 맞추어 조사 도구의 내용적인 측면뿐만 아니라 시행적인 측면을 모두 고려하여 국어 사용 환경의 변화를 반영한 문항을 개발하되 충분한 검증 절차를 통해 조사 도구의 안

정화 및 정교화를 도모하여 추후 이루어질 영역별 연차 조사에 참조 지점을 마련할 수 있다는 의의를 지닌다.

셋째, 본 연구는 그간 두 차례 이루어진 국가 차원의 대규모 조사가 지속적이고 안정적으로 시행되는데 기여하되 영역별, 연차별 실태 조사를 시행하기 위한 조사 고도화의 이론적 기반과 방법론을 제시한다는 의의를 가진다. 이를 통해 영역별 조사 도구 개발과 대규모 실태 조사의 시행이 체계적이고 연계적으로 작동하여 국어능력 지표에 대한 연속적이며 초점화된 논의가 가능해질 것이다. 이는 국내의 다양한 국어능력 평가에 영향을 미치는 큰 틀로서 작용하게 될 것이며, 국제적인 대규모 조사 사업과 위상을 견주며 국어능력에 대한 내실 있는 논의를 촉발하는 데에도 기여할 것이다.

2. 연구의 범위

본 연구에서 수행하는 연구의 범위는 다음의 <표 I -2>와 같이 범주화될 수 있다.

<표 I -2> 연구 내용 범위

성과 평가	■ ‘국민의 국어능력 평가(2013)’, ‘국민의 국어능력 실태 조사(2018)’의 성과 및 한계 분석 ■ 2023 실태 조사 계획 수립을 위한 보완점 및 반영 사항 도출
국어 사용 환경 및 현황 분석	■ ‘디지털 리터러시’ 관련 선행 연구 및 통계 지표 검토 ■ 변화된 국어 사용 환경에 따른 영역별 국어능력의 범위 및 개념 확립
실태 조사 사례 검토	■ 국내외 국어능력 실태 조사 사례 검토 ■ 유사 국어(언어)능력 조사 연구 검토 ■ 조사 영역 및 요소, 조사 대상 및 방법 관련 시사점 도출
실태 조사 고도화 계획 수립	■ 연차별, 영역별 조사 시기, 내용, 방법 등에 대한 구체적 방안 마련 ■ 인공지능 활용 국민의 국어능력 진단체계 개발 사업과 연계 방안 마련
조사 도구 평가 및 설계 방안 마련	■ ‘국민의 국어능력 평가(2013)’, ‘국민의 국어능력 실태 조사(2018)’의 듣기·말하기 영역 조사 도구의 성과 및 한계 분석 ■ 3차 국어능력 실태 조사 듣기·말하기 영역의 조사 도구 설계 및 시행 방안 수립
듣기·말하기 영역의 조사 도구 개발	■ 듣기, 말하기, 듣기·말하기(연계) 영역의 문항 및 검사 설계 : 듣기, 말하기 영역의 핵심 역량 분석, 세부 조사 항목 및 조사 요소 설정 ■ 1차 예비 조사를 위한 조사 도구 개발
1차 예비 조사 시행 및 결과 분석	■ 1차 예비 조사 표본 집단(240명) 추출 ■ 1차 예비 조사 시행 ■ 문항 양호도 분석: 정답률, 변별도, 신뢰도, 타당도 분석 등 ■ 1차 예비 조사 응답 결과에 대한 채점 진행
조사 도구 보정	■ 조사 도구 수정·보완(1.5배수 문항)
2차 예비 조사 시행 및 결과 분석	■ 2차 예비 조사 표본 집단(240명) 추출 ■ 2차 예비 조사 시행 ■ 문항 양호도 분석: 정답률, 변별도, 신뢰도, 타당도 분석 등 ■ 2차 예비 조사 응답 결과에 대한 채점 진행 ■ 조사 도구 정교화 및 확정(1.2배수 문항)
본 조사 시행 및 결과 활용 방안 제안	■ 국민의 국어능력 실태 조사 시행 및 결과 활용 방안 마련 : 2023년 국민의 국어능력 조사 시행을 위한 제언 : 조사 결과 활용 및 홍보 방안 제안

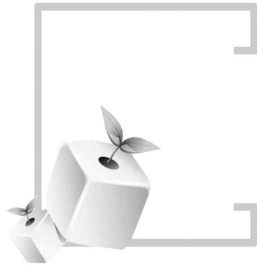
3. 연구의 추진 과정

본 연구의 추진 과정은 다음 <표 I-3>과 같다.

<표 I-3> 단계별 추진 일정 및 세부 사업 내용

단계	추진 일정	세부 사업 내용
1단계 1, 2차 국민의 국어능력 실태 조사 성과 평가	착수 ~ 6월 30일	■ 계약 및 연구 착수 ■ ‘국민의 국어능력 평가(2013)’, ‘국민의 국어능력 실태 조사(2018)’ 성과 및 한계 분석 ■ 조사 도구 및 실행 관련하여 보완점 마련 ■ 2023 실태 조사 계획 수립을 위한 반영 사항 도출 [국립국어원 보고] ■ 착수 보고
2단계 국어 사용 환경 및 현황 분석에 따른 국어능력의 개념 수립	7월 1일 ~ 7월 8일	■ 국민의 국어능력 관련 국어 사용 환경 및 현황 분석 ■ 디지털 리터러시 관련 선행 연구 및 각종 통계 지표 분석 ■ 디지털 환경에서 국민의 듣기, 말하기, 읽기, 쓰기 활동, 국어능력의 수요와 활용 분석 ■ 변화된 국어 사용 환경에 따른 영역별 국어능력의 범위 및 개념 수립
3단계 국내외 국어능력 실태 조사 사례 검토	7월 9일 ~ 7월 20일	■ 미국, 영국, 중국, 일본, 유럽 연합 등 국내외 국어능력 실태 조사 검토 ■ 국립국어원 국어능력 실태 조사 외 유사 국어(언어)능력 조사 ■ 평가 영역 및 요소, 조사 대상, 조사 방법 관련 시사점 도출
4단계 국어능력 영역별·연차별 실태 조사 고도화 방안 계획 수립	7월 21일 ~ 8월 10일	■ 국민의 국어능력 실태 조사 고도화 방안 수립 ■ 영역별 조사 시기, 조사 내용, 조사 방법, 조사 결과 활용 및 정책 환류에 관한 구체적 방안 마련 ■ 인공지능 활용 국민의 국어능력 진단체계 개발 사업과 연계 방향 마련
5단계 실태 조사 도구 분석 및 조사 설계 방안 마련 (듣기, 말하기)	8월 11일 ~ 8월 20일	■ 2013, 2018 실태 조사의 듣기, 말하기 영역 조사 도구의 성과 및 한계 분석 ■ 제3차 국어능력 실태 조사의 듣기, 말하기 영역 조사 도구 개발 및 방향 수립

6단계 국민의 국어능력 조사 도구 개발 (듣기, 말하기)	8월 21일 ~ 9월 15일	■ 국어능력 조사를 위한 듣기, 말하기 영역 문항 설계 및 검사 설계 ■ 듣기, 말하기 영역의 핵심 역량 분석, 세부 평가 영역 및 요소 설정 ■ 1차 예비 조사를 위한 조사 도구 개발 ■ 듣기, 말하기, 듣기·말하기(연계) 영역 문항 개발 (가교 문항 포함, 최종 문항의 2배수 개발) ■ 영역별(듣기, 말하기) 논의 및 전체 회의 실시
7단계 1차 예비 조사 실시	9월 16일 ~ 11월 1일	■ 예비 조사 표본 집단 추출: 240명 ■ 예비 조사 요원 훈련 ■ 1차 예비 조사 시행 ■ KANTAR PUBLIC과의 공조를 통한 조사 실시
8단계 1차 예비 조사 결과 분석 및 조사 도구 보정	11월 2일 ~ 11월 21일	■ 문항 양호도(정답률, 변별도, 신뢰도, 타당도 등) 분석 ■ 말하기 영역 주관식 채점 진행 ■ 피험자의 거부감, 피로도 분석 ■ 조사 시행의 난점, 적합성 정도 분석 ■ 조사 도구 수정/보완 ■ 영역별(듣기, 말하기) 논의 및 전체 회의 [국립국어원 중간 보고(11월 중순)] ■ 1차 예비 조사 개요 및 결과 보고 (인원, 소요 예산, 평가지 등 포함)
9단계 2차 예비 조사 실시	11월 22일 ~ 12월 9일	■ 2차 예비 조사 표본 집단 추출: 240명 ■ 예비 조사 요원 훈련 ■ 2차 예비 조사 시행
10단계 2차 예비 조사 결과 분석 및 조사 도구 정교화·확정	12월 9일 ~ 12월 12일	■ 문항 양호도(정답률, 변별도, 신뢰도, 타당도 등) 분석 ■ 말하기 영역 주관식 채점 진행 ■ 조사 도구 정교화 및 확정(1.2배수 문항)
11단계 본 조사 시행 및 결과 활용 방안 제안	12월 13일 ~ 12월 16일	■ 국민의 국어능력 실태 조사 시행 및 결과 활용 방안 마련 ■ 2023년 국민의 국어능력 조사 시행을 위한 제언 ■ 조사 결과 활용 및 홍보 방안 제안 [국립국어원 최종 보고] ■ 용역 사업 최종 보고



제 2 장

국민의 국어능력
실태 조사 고도화를
위한 이론적 검토



1. 1, 2차 국민의 국어능력 실태 조사 성과 평가

1) 조사 도구 측면에서의 성과 분석

본 연구는 5년 주기로 시행되는 국민의 국어능력 실태 조사의 제3차 실행(2023년)을 위한 기초 연구에 해당되며, 이를 위해 우선 1차 실태 조사(2013년)와 2차 실태 조사(2018년)의 성과를 평가하고 부분적 한계에 대한 보완을 통해 3차 실태 조사(2023년)의 시행 계획을 고도화할 필요가 있다. 먼저, 국민의 국어능력에 대한 제1차 실태 조사(2013년)와 제2차 실태 조사(2018년)는 다음과 같은 방향성을 유지하는 가운데 시행되었다.

제1차 실태 조사(2013년)	제2차 실태 조사(2018년)
① 2012년에 개발된 조사 도구를 전면적으로 재설계 ② 종이 검사지를 통해 구현 ③ 400명 규모의 예비 조사 시행 ④ 3,000명 규모의 본 조사 시행 ⑤ 본 조사 결과의 정밀한 분석 ⑥ 5년 주기 국어능력 조사의 추이 분석을 위한 기반 마련 ⑦ 국어능력 수준의 등급화를 위한 앵고프 연구 시행 ⑧ 결과 분석을 통한 국어정책 방향 수립 입론	① 2017년에 개발된 조사 도구를 ‘문법’, ‘읽기’ 중심 (디지털 영역 추가)으로 보완적으로 재설계 ② 태블릿 피시를 통해 구현 ③ 조사 대상의 연령 확대를 통한 노인 문식력의 측정 ④ 300명 규모의 예비 조사 시행 ⑤ 조사 도구의 삼원적 설계 및 조사비의 상향 ⑥ 국어능력 측정을 위한 변인의 다양화 ⑦ 본 조사 시행 규모 확대 ⑧ 워크숍을 통한 주관식 채점의 정밀화 ⑨ 변화된 국어능력의 추이 분석 ⑩ 학계와의 연계를 통한 국어 정책 방향 제언

아울러 제1차 실태 조사와 제2차 실태 조사에서는 다음과 같은 영역별 통합 조사 도구를 활용함으로써 조사에 참여하는 국민들의 피로도가 높아지고 집중도가 떨어짐에 따라 조사 결과의 신뢰성을 낮추는 문제가 제기되었다. 다양한 변인을 갖춘 일반 국민들을 대상으로 하여 조사가 진행됨에 따라 1시간 이상 여러 영역의 다양한 형태의 문항에 답을 하는 과정의 한계를 노정하는 것이다. 물론, 이원화(1차 조사) 및 삼원화된(2차 조사) 조사 도구를 마련한 후 공통 문항 비동등 집단 설계(common item non-equivalent design)를 통해 이러한 문제를 최소화하고자 하였으나 완벽하게 이 문제를 해결하기는 어려웠던 것으로 보인다. 이에 따라, 제3차 실태 조사(2023년 이후)에서는 연차별로 상이한 영역의 조사를 초점화하여 시행함으로써 조사 도구의 신뢰도를 높일 필요가 있다.

<표 II-1> 1차 실태 조사(2013년)의 조사 도구 설계

〔‘가’ 형〕				〔‘나’ 형〕			
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 II-2> 2차 실태 조사(2018년)의 조사 도구 설계

〔‘가’ 형〕				〔‘나’ 형〕				〔통합형 ‘가’+‘나’ 형〕			
조사 대상자 수: 2,500명				조사 대상자 수: 2,500명				조사 대상자 수: 500명			
영역	문항 수	시간 (분)	배점	영역	문항 수	시간 (분)	배점	영역	문항 수	시간 (분)	배점
듣기	10	15	150	쓰기	3	30	250	듣기	10	15	150
말하기	3	15	150	읽기	12	24	180	말하기	3	15	150
읽기	12	24	180	문법	10	10	150	읽기	20	40	300
문법	10	10	150	디지털	2	6	30	문법	10	10	150
디지털	2	6	30	누계	27 문항	70분	610	쓰기	3	30	250
누계	37 문항	70분	660					디지털	2	6	30
								누계	48 문항	116분	1030

2) 조사 인원 측면에서의 성과 분석

제한된 예산 및 상기와 같은 통합형 조사 도구의 문제로 인해 1차 실태 조사와 2차 실태 조사에서는 다음과 같이 조사에 참여하는 표본의 수가 크지 않았다. 물론, 특정 변인에 따른 층화 표본 추출(stratified sampling) 방식에 따라 선정된 표본 집단이라는 점에서는 통계적 대표성을 충분히 확보할 수 있지만, 보다 정교한 조사 결과를 도출하고 국민의 국어능력 조사에 대한 관심을 제고하기 위해서는 축소된 조사 도구를 통해 보다 다양한 국민들이 조사에 참여하게 할 필요가 있다.

<표 II -3> 1, 2차 실태 조사의 조사 인원

	예비 조사	본 조사
1차 실태 조사	수도권에 거주하는 20세~59세의 성인 남녀 400명	20세~59세의 성인 남녀 ‘가’ 형: 1,416명 ‘나’ 형: 1,500명
2차 실태 조사	수도권에 거주하는 20세~69세의 성인 남녀 300명	20세~69세의 성인 남녀 통합형: 500명 ‘가’ 형: 2,470명 ‘나’ 형: 2,500명

3) 조사 방식 측면에서의 성과 분석

1차 실태 조사(2013년)에서는 숙달된 조사원들이 표본 집단에 대한 1:1 방문 면접을 통해 지면 형태의 조사지로 조사를 수행하였다. 2차 실태 조사(2018년)에서는 기본적으로 태블릿 PC를 활용한 디지털 검사지를 통해 조사를 수행하고 태블릿 PC에 어려움을 겪는 표본 집단을 대상으로는 지면 검사지로 보완하는 병행 방식으로 조사를 수행하였다. 그런데 이러한 조사 방식은 기본적으로 가구 방문을 통한 1:1 조사라는 점에서 조사 환경 및 피조사자들의 적극적 참여도 측면에서 한계가 있을 것으로 분석된다.

따라서 표본 집단을 특정한 장소로 방문하게 하여 동일한 형식으로 CBT로 실태 조사를 수행하는 방법의 병행을 적극적으로 고려할 필요가 있다. 물론, 이에 대한 적절한 유인책을 마련하기 위해서는 관련 예산의 증대 편성이 무엇보다 요구된다.

4) 조사 결과 해석 측면에서의 성과 분석

영역 통합형 조사 도구의 문제로 인해 조사 도구를 최소화해야 한다는 점에서 1차 및 2차 실태 조사에서는 각 영역의 하위 요소별 분석을 구체적이고 심층적으로 시행하기 어려웠다. 듣기 영역과 말하기 영역을 예로 들면, 1차 및 2차 실태 조사에서는 문항 수의 제한으로 인해 말하기 영역은 ‘설득하는 말하기, 설명하는 말하기’로, 듣기 영역은 ‘사실적 듣기, 추론적 듣기, 비판적 듣기’로 하위 요소를 상정하였으나 이것만으로는 실태 조사 결과에 대한 심층적 분석은 한계가 있다. 따라서 연차별로 영역을 달리 하여 초점화된 실태 조사를 시행한다면 각 영역의 세분화된 하위 요소에 대한 문항별 분석을 시행함으로써 국민들이 각 영역에서 겪는 언어적 어려움의 원인을 진단하고 이를 타개하기 위한 정책적 방향성을 입론할 수 있을 것이다.

2. 국내외 국어능력 및 디지털 리터러시 실태 조사 사례 검토

1) 국내의 국어능력 실태 조사 사례

(1) 2018년 국민의 국어능력 실태 조사

‘2018 국민의 국어능력 실태 조사’는 ‘국어기본법’에 의거하여 국립국어원에서 주관하고 KANTAR Korea가 조사를 수행하였다. 국민의 국어능력과 관련된 전반적인 실태를 조사하여, 국민의 국어능력 증진을 위한 정책 주안점을 찾는 데 목적이 있다. 전국 17개 시도에 거주하는 만 20세~69세의 남녀 5,500명을 대상으로 층화 집락 추출과 응답자 할당 추출의 2가지 단계로 표본을 추출한 후, 1:1 가구 방문 면접 조사 및 CAPI(Computer-Aided Personal Interview)를 통해 자료를 수집하였다. 민병곤 외(2017)의 기초 연구에서 구안한 내용을 토대로 하여, 48개 문항, 116분, 1,030점 만점의 조사 도구를 재설계하였고, 조사 실행의 편의를 위해 해당 조사에서는 조사 도구를 ‘가’형, ‘나’형, ‘통합형’의 세 가지로 나누어 설계하되, ‘가’형과 ‘나’형에서 읽기 4문항, 문법 10문항, 디지털 2문항의 16문항을 공통 문항으로 설정하여 조사 도구를 동등화하였다. 구체적인 조사 영역 및 방식, 문항 수와 시간, 배점은 <표 II-4>와 같다.

<표 II-4> 2018년 국민의 국어능력 실태 조사 영역, 방식, 문항 수, 시간, 배점

조사 영역	조사 방식	문항 수	시간	배점
읽기	선택형(객관식)	20문항	40분	300점
문법	선택형(객관식)	10문항	10분	150점
쓰기	서답형(논술, 주관식)	3문항	30분	250점
듣기	선택형(객관식)	10문항	15분	150점
말하기	수행형(주관식)	3문항	15분	150점
디지털	선택형(객관식)	2문항	6분	30점
누계		48문항	116분	1,030점

이후 조사 결과는 분할 점수에 따라 4수준, 3수준, 2수준, 1수준으로 성취 수준이 등급화되었다. 이때, 2013년 조사 결과와 2018년 조사 결과 사이의 종단적 비교가 가능하도록 2013년의 분할 점수를 준용하여 조사 도구 간 동등화를 거쳐 새로이 도출한 분할 점수를 적용하였다.

이 조사는 ‘2013 국민의 국어능력 조사’에 이어 실시되었지만, 일부 차이가 존재한다. 상술한 CAPI(Computer-Aided Personal Interview) 방식, 조사 도구의 삼원적 설계 외에 조사 도구의 설계 측면에서 일부 차이가 있었다. 또한, ‘2013 국민의 국어능력 조사’의 말하기, 쓰기, 듣기, 읽기, 문법의 다섯 영역에 ‘디지털 영역’을 추가하였다는 점도 상이하다.

<표 II -5> 2013년 조사 도구와 2018년 조사 도구의 설계 측면에서의 비교

연도 설계	2013	2018
조사 인원	3,000명 대상	6,000명 대상
조사 인원의 연령대	20세~59세의 성인 남녀	20세~69세의 성인 남녀
조사 도구의 설계	이원적 설계	삼원적 설계
문항 분석 틀	고전검사이론	고전검사이론 및 문항반응이론
조사 결과 등급화	앵고프(Angoff) 방식을 통한 분할 점수 산출	이전 주기 분할 점수 준용(2013년의 산출된 분할 점수를 2018년 원 점수 척도로 변환하여 사용)

(2) 2020년 국민의 언어의식 조사

‘2020년 국민의 언어의식 조사’(김창영 외, 2020)는 2005년에 조사를 처음 실시한 후, 네 번째로 실시된 조사로서 전국 17개 시군 만 20세 이상 70세 미만의 성인 남녀 5,000명을 층화 2단 추출법을 통해 표집한 후, 가구 방문에 의한 1:1 개별 면접 조사하는 방식으로 시행되었다. 여기서는 국민들의 국어에 대한 관심과 문제의식, 언어 사용 의식, 언어 교육과 언어 정책에 관한 의식을 조사하였다.

이 조사는 국민의 국어능력 실태를 조사한 연구는 아니지만, 5년 주기로 조사를 시행함으로써 비교 가능한 자료를 축적해 왔다는 점, 다양한 설문 배경 변인을 고려해 결과를 해석하고 있다는 점이 참고할 만하다.

(3) 2008년 국민의 기초 문해력 조사

‘2008년 국민의 기초 문해력 조사’(김창원 외, 2008)는 국민의 기초 문식력을 측정하기 위한 조사이다. 국립국어원에서 주관하고 한국 갤럽 조사 연구소가 조사를 수행하였다. 현실에 기초한 국민의 문식력에 대한 정확한 통계 자료를 확보하여 이를 문해 교육 정책 개발의 기반 자료로서 삼는 것을 목적으로 한다. 조사는 만 19세 이상 80세 미만의 대한민국 국적을 가지고 있는 자를 대상으로 하여 가구 방문 면접 조사로 진행되었다. 조사 대상이 되는 표본은 4단 층화 확률 비례 추출법으로 추출하였으며, 최종적으로 조사에 응한 표본의 크기는 7,033명이었다.

조사 도구로 사용된 질문지는 총 25문항으로, 현대 사회에서 한국 국민으로서 일상생활을 해 나가는 데 필요하다고 판단되는 실용적이고 교양적인 글, 서식 등의 다양한 언어 자료를 바탕으로 한 최소한의 읽고 쓰는 능력(기초 문식력)을 측정할 수 있도록 작성되었다. 질문지 문항 설계의 기본적인 틀은 <표 II -6>과 같다.

<표 II -6> 2008년 국민의 기초 문해력 조사의 내용 요소

	내용	과정(인지)	상황
하위 요소	· 산문(문학/비문학) · 문서(도표, 그림)	· 사실적 문해 · 추론적 문해 · 비판적 문해	· 실용(일상생활, 공적 생활, 미디어) · 교양(인문/예술, 사회/문화, 과학/기술)

이후, 문해력 점수를 크게 ‘문해력 부진’과 ‘기초 문해력 보유’로 나누고 전자를 다시 ‘0수준’, ‘1수준’, 후자를 ‘2, 3, 4수준’으로 나누어 모두 5단계로 구분하여 조사 결과를 처리하였다. 이 조사는 문식력에 국한된 조사로서 국민의 전체적인 국어능력을 조사하는 것과는 다소 거리가 있다. 그러나 다수의 국민들을 대상으로 한 조사라는 점에서 표본 추출이나 통계 처리와 같은 조사 방법론의 측면에서 참고할 수 있다.

(4) 국어능력인증시험(ToKL)

‘국어능력인증시험(ToKL)’은 재단법인 한국언어문화연구원이 주관하며, 한국어를 모국어로 하는 학생 및 일반인을 대상으로 말하고, 듣고, 읽고, 쓸 줄 아는 총체적인 국어능력을 평가하는 것을 목적으로 한다.

평가 영역은 크게 ‘언어 기초 영역’, ‘언어 기능 영역’, ‘사고력 영역’으로 구분되며, 각각의 세부 평가 내용은 ‘수행기반능력(어휘), 언어규범능력(어법, 어문 규정), ‘청해능력(듣기), 독해능력(읽기), 작문능력(쓰기)’, ‘사실적 이해, 추론, 비판, 창의’로 구분된다. 문항 수는 객관식 80문항과 주관식 10문항이다. 답안지 채점의 경우, 객관식은 OMR 리더기로 채점 후 150점 이상의 고득점자와 100점 이하의 저득점자에 한해서는 역채점 방식을 취한다. 주관식은 동일 답안에 대해 3차에 걸친 채점 실시 후 오차율 30% 이상의 결과에 대해서는 재채점을 실시한다.

이 시험은 듣기와 읽기와 같은 이해 영역에서 사고력의 범주를 구분하여 ‘이해, 추론, 비판, 창의’에 해당하는 범주를 묻고 있는 것이 특징적이다. 주관식 채점의 타당도를 높이기 위한 방법적 측면 또한 참고할 만하다.

(5) KBS 한국어능력시험

‘KBS 한국어능력시험’은 KBS 한국어연구회와 자문위원들이 출제, 감수, 관리를 맡고 KBS 한국어진흥원이 주관하는 시험이다. 원만하고 능률적인 언어생활에 요구되는 국어 사용 능력과 국어 관련 교양을 측정하는 것을 목표로 한다. 지원자를 대상으로 한 지필 평가로서 문항의 형식은 오지선다 객관식이다. 각 답안지의 채점은 객관식은 OMR 리더기로 채점하여 처리한다. 평가 영역은 문법, 이해, 표현, 창안, 국어문화 영역으로 나뉘며, 구체적인 영역별 주요 내용은 <표 II -7>과 같다.

<표 II -7> KBS 한국어능력시험 영역별 주요 내용

영역 구분	주요 내용
문법 영역 (어휘·어법)	한글 맞춤법, 표준어 규정, 외래어 표기법, 로마자 표기법 등
이해 영역 (듣기·읽기)	어휘력, 사실적 이해 능력, 추리·상상적 이해 능력, 비판적 이해 능력, 논리적 이해 능력 등
표현 영역 (쓰기·말하기)	보고서 작성 및 발표·토론 능력, 협상 및 대인 설득 능력, 논증 능력, 표준 화법 등
창안 영역 (창의적 언어능력)	창의적 표어 제작, 제목 추출, 아이디어 창안, 수사법, 고사성어, 속담 활용 능력 등
국어문화 영역 (국어교과의 교양 지식)	국어 상식 및 국어 문화에 대한 교양적 이해 능력

이 시험의 경우 기초적인 문식 능력을 요구하는 이해와 표현 영역 이외에도 창안 영역이나 국어문화 영역 등, 창의적 사고나 언어 문화와 관련된 영역을 묻고 있다는 점이 특징적이다.

(6) 한국어능력시험(TOPIK)

국립국제교육원(National Institute for International Education)에서 주관하는 평가이다. 한국어를 모국어로 하지 않는 재외동포와 외국인의 한국어 학습 방향을 제시하고 한국어 보급을 확대하는 데 목적이 있다. 또한, 한국어 사용 능력을 측정·평가하여 그 결과를 국내 대학 유학 및 취업 등에 활용하고 있다. 한국어를 모국어로 하지 않는 재외동포 및 외국인으로서, 한국어 학습자 및 국내 대학 유학 희망자, 국내외 한국 기업체 및 공공기관 취업 희망자, 외국 학교에 재학 중이거나 졸업한 재외국민이 주로 응시한다. 지원자를 대상으로 한 지필 평가 방식으로 진행되며, TOPIK I의 경우에는 듣기, 읽기 영역을, TOPIK II의 경우에는 듣기, 읽기, 쓰기 영역을 평가한다. 문항은 선택형과 서답형으로 구성되어 있고, 서답형 문항은 TOPIK II에만 포함된다. TOPIK II의 서답형 문항은 문장 완성형(단답형) 2문항, 작문형 2문항(200~300자, 600~700자 각 1문항씩)으로 이루어져 있다.

TOPIK I은 총 100분, TOPIK II는 총 180분간 실시되며, 배점은 각 영역별 100점씩으로 TOPIK I은 200점 만점, TOPIK II는 300점 만점이다. 영역별 문항 구성 및 시험 시간은 다음과 같다.

<표 II -8> 한국어능력시험(TOPIK)의 영역별 문항 구성 및 시험 시간

시험 수준	영역	문항 수	평가 시간
TOPIK I	듣기	30	40분
	읽기	40	60분
TOPIK II	듣기	50	60분
	쓰기	4	50분
	읽기	50	70분

2) 국외의 자국어능력 실태 조사 사례

(1) 미국

가. 성인 문식력 조사(National Assessment of Adult Literacy, NAAL)²⁾

미국 국립 교육 통계 센터(National Center of Educational Statistics)가 주관한 조사로, 전국적 차원에서 16세 이상의 미국 성인을 대상으로 영어 문식력의 수준을 조사하였다. 조사 대상은 약 19,000명으로, 미국 거주 인구 중 3만 5천 개 이상 가구를 선택해 이 가구에 거주하는 16세 이상 성인 1만 8천 명 이상에 대해 직접 인터뷰를 시행하고 전국 107개 교도소에 수감 중인 수감자 중 1,200명을 추출해 조사를 시행하였다. 표본 추출은 4단 층화 확률 비례 추출법을 사용하였으며, 조사는 구조화된 질문지를 활용한 방문 면접 조사 방식으로 진행되었다.

이 조사에서 문식력은 산문 문식력(prose literacy), 문서 문식력(document literacy), 수리 문식력(quantitative literacy)의 세 유형으로 구분된다. 이에 더해, 단어 해독 및 인지 능력(the Fluency Addition to NAAL: FAN), 글자와 숫자 이해 조사(the Adult Literacy Supplemental Assessment: ALSA)도 포함하여 조사를 실시하였다. 조사 결과는 우수(Proficient), 보통(Intermediate), 기초(Basic), 기초 미달(Below Basic)의 네 개 등급으로 나뉜다. 또한 1992년과 2003년의 결과 비교를 통해 국가 수준의 문식력의 변화를 비교하여 조사 대상의 문식력과 인구학적 특성 및 교육 수준의 관계, 교도소에 수감 중인 수감자들의 문식력 정도 평가, 주별 성인 문식력 평가, 성인이 건강 관련 정보나 서류에 대해 이해할 수 있는 문식력 정도, 성인의 기초 독해력 평가, 문자와 숫자를 구별하고 단순한 산문체나 서류를 이해할 수 있는 성인의 최소 문식력을 분석하였다.

나. 국가 수준 교육 성취도 평가(National Assessment of Educational Progress, NAEP)

미국 국립 교육 통계센터(National Center of Educational Statistics)에서 주관하며, 미국 내 학생들의 학업 성취도를 비교하기 위한 시험이다. 교육 정책 마련을 위해 실시되므로 개인이나 학교 간 성적을 비교하기 위한 시험은 아니다. 지역, 성(性), 계층, 인종에 따른 비교를 통해 교육 격차를 파악한다. 수학과 읽기는 자주 시험을 치르고, 쓰기, 과학, 미국사, 사회, 지리, 예술 등은 정기적으로 시험을 치르며 그 시기는 2~3년으로 균일하지 않다. 행렬 표집(matrix sampling) 방식으로 조사 도구를 구성하는 것도 특징이다. 국어능력 평가와 직접적으로 관련 있는 영역은 읽기와 쓰기 영역이다.

평가 틀은 교과별, 학년별로 구분되고, 각 학년에서 학생들이 알아야 할 것, 할 수 있어야 할 것으로 구성되어 있다. 정부위원회의 감독하에 개발되는 평가 틀은 내용 전문가, 학교 행정가들, 정책 결정자들, 교사들, 학부모 등을 포함하는 교육 관련 인사들의 지식과 경험을 반영한다. 국가학업성취도운영위

2) 'literacy'의 번역어 혹은 이를 일컫는 단어로 '문해', '문해력', '문식성', '문식력', '리터러시' 등 다양한 단어가 사용되어 있어 언급이 필요하다. '문해'는 문자를 읽고 쓰는 기초적인 문식성의 의미로 사용되는 경우가 대부분이며(길호현, 이지선, 2022: 191), '문해력'의 경우 '문해'에 비해서는 기능이나 역량과 관련되기는 하지만 여전히 기초 문식성의 의미를 내포하고 있다. 하지만 'literacy'에 대한 연구가 활발해지면서 이 개념은 문자 해독 능력이라는 좁은 의미에서 사회문화적인 측면까지 포괄하는 넓은 의미로 확장되어 왔는데(김윤주, 2021: 254), '문식', '문식성', '문식력'이라는 용어는 이처럼 기초 문식 활동에서 더 나아가, 다양한 문식 활동을 포괄할 수 있다는 장점이 있다. 이러한 이유로 학문 분야에 따라 기초 문식성을 넘어서는 'literacy'를 일컬을 경우 '문해', '문해력'보다는 '문식', '문식성', '문식력'을 주로 채택하고 있어, 이러한 관행을 따라 번역어를 선택하였다.

위원회(National Assessment Governing Board, NAGB)에 의해 평가 틀이 완성되면 NCES가 문항 개발을 한다.

이후 표본이 되는 평가 대상을 선정하여 이들을 대상으로 평가를 실시한다. 평가는 지필 평가와 수행 평가로 나누어진다. 지필 평가 문항은 선택형과 서답형으로 나누어지는데 서답형 문항은 응답 길이가 짧은 단답형과 응답 길이가 비교적 긴 서술형 문항이 포함된다. 평가 대상은 4, 8, 12학년 학생들이고, 대상자들은 표집을 통해 선발된다. 표집 방법은 학생 수, 성비, 인종적 비율 등을 고려해 무선 추출법을 따른다. 조사는 하루 안에 끝내는 것을 원칙으로 한다. 읽기 영역과 쓰기 영역의 문항 구성은 다음 <표 II-9>와 같다.

<표 II-9> NAEP의 읽기 영역과 쓰기 영역의 문항 구성

문항 구성	
읽기 영역	■ 문항의 성격: 내용 이해, 통합과 해석, 비판과 평가 ■ 질문의 유형: 선다형, 단답형, 서술형
쓰기 영역	■ 문항의 성격: 설득/설명/문학적인 글쓰기 ■ 질문의 구성: 서술형(각 쓰기 목적별 3문항 출제, 각각 1문항씩 선택) ■ 평가 기준: 내용의 깊이와 다양성, 내용 조직의 체계성, 표현의 정확성과 다양성 등

선택형의 경우 답안지를 스캔하여 컴퓨터로 채점하고, 정답과 오답 구분만 하여 부분 점수를 허용하지 않는다. 서술형 문항의 경우는 답안의 질과 내용에 따라 차등 점수가 배정된다. 이때, 답안지를 이미지화하여 채점자에게 직접 전달하는 온라인 채점 시스템(image-based scoring system)을 사용한다. 이는 채점자의 부담을 줄이고, 장소의 제약을 탈피하기 위해서 도입한 것으로, 전통적인 채점 훈련과 채점 과정을 모두 온라인으로 실시하는 방식이다.

검사 점수는 문항반응이론 모형을 활용하여 척도화된다. NAEP는 평균이 250이고, 표준편차가 50이며 0점~500점까지의 범위가 가능한 척도 점수 체계를 가지고 있는데, 척도 점수는 교과별로 산출되어 교과목 간 점수 비교는 불가능하다.

교과별 성취도는 0점~300점 또는 0점~500점의 척도 점수와 보통(Basic), 우수(Proficient), 최우수(Advanced) 등의 성취 수준으로 판정한다. 또한 결과는 저, 중, 고로 학생들의 성취를 나누고, 변화 추이는 5개의 백분위 점수(10, 25, 50, 75, 90)별로 나누어 판정한다.

다. 미국 대학 입학 능력 자격 시험(Scholastic Aptitude Test, SAT)

미국의 대학 입학 능력 시험으로, 미국 대학 위원회(College Board)가 감독 및 실시하며, ETS(Educational Testing Service)에서 개발·편찬·채점한다. 기존 SAT는 근거 기반 읽기/쓰기(reading & writing), 수학(math), 에세이(Essay)의 세 영역으로 구성된 SAT와, 교과목을 선택하여 응시하는 시험인 SAT Subject test 두 축으로 이루어졌다. 하지만 2016년 개편 이후 SAT의 영역은 읽기(Reading), 쓰기 및 언어(Writing and Language), 수학(Math) 세 가지로 변경되었으며, 에세이는 선택 응시로 바뀌었다. 또한, 2021년부터는 SAT Subject test와 에세이가 폐지되었다. 이에 대부분의 문항이 선택형이며, 수학 영역에서 일부 서답형 문항이 존재한다.

<표 II-10> SAT의 영역별 문항 수 및 소요 시간

시험	문항 수	소요 시간
읽기 (Reading)	52개	65분
쓰기 및 언어 (Writing and Language)	44개	35분
수학 (Math)	58개	80분
계	154개	180분

읽기 영역의 경우 장문의 지문 5개(근현대 문학 작품 1개, 사회과학 지문 1개, 역사 지문 1개, 과학 지문 2개)를 제시하여 증거에 대한 이해, 문맥 내 단어에 대한 이해, 역사/사회 및 과학 교과 기반 이해를 평가한다. 쓰기 및 언어 영역의 경우 약 300자에서 400자 정도의 지문을 4개 제시하며, 각 지문마다 11개의 객관식 문항으로 구성된다. 문항은 크게 내용의 표현 측면에서의 질문과, 문장 구조, 문법, 구두점의 오류를 인식하고 수정하는 질문으로 이루어진다.

SAT의 총 점수는 400점에서 1600점 사이이다. 총 점수는 읽기 및 쓰기, 수학 두 섹션 점수의 합이며, 이 두 섹션 점수의 가능한 범위는 각각 200점~800점이다. 섹션 점수는 각 섹션의 원 점수를 기반으로 하며, 이는 다시 200~800점 범위의 척도 점수로 변환된다. 척도 점수를 통한 점수 동등화는 매회 시험의 점수가 일관된 의미를 갖도록 조정한다.

읽기 및 쓰기 섹션의 경우 읽기 시험이 10~40점, 쓰기 및 언어 시험이 10~40점으로, 둘의 총합에 10을 곱해 800점 만점의 섹션 점수를 부여한다. 수학 섹션의 경우에는 10~40점인 수학 시험 점수에 20을 곱하여 섹션 점수를 계산한다. 또한 총점수에 더해 교차 시험 점수(Cross Test Score)와 하위 점수(Subscores)라는 보조 점수가 부과되어 학습자의 강점을 보다 세밀하게 평가하는 역할을 한다.

SAT는 등급을 따로 설정하지 않고 척도 점수와 더불어 점수 범위(score range), 백분위(percentiles)가 보고된다. 점수 범위는 수험생이 취득한 점수의 상하 30~40점으로 표현된 점수 구간으로, 동일 응시자가 다시 시험을 치를 때 예상되는 점수 구간을 의미하며, 이 범위의 점수는 동일한 능력을 나타내는 것으로 해석한다.

(2) 영국

가. A-level(General Certificate of Education Advanced Level) 언어 시험

A-level(Advanced-level) 언어 시험은 AEB(Associated Examining Board), 케임브리지 대학, 런던 대학, 옥스퍼드 대학, NEAB(Northern Examinations and Assessment Board) 등 대학과 지역 기관 등 다양한 기관에서 주관하여 치러지는 평가이다. 1951년 처음으로 소개되어 지금까지 영국을 비롯하여 미국, 유럽, 아시아, 아프리카 등 전 세계에서 대학 지원 시 학생들의 학습 능력 평가를 위한 기준으로 널리 인정받아 오고 있다.

이 시험은 평가에 지원한 학생들을 대상으로 진행된다. 시험은 모두 주관식으로서 보통 10문제 중 3문제를 선택하여 서술한다. 한국 기준으로 대학 1학년 정도 수준의 논술형이며, 평가 문항은 ‘평가하라, 논하라, 비교·대조하라, 비평적으로 고찰하라, 어느 범위까지 동의하겠는가’ 등으로 제시된다. 시험 소요 시간은 과목당 3시간이며, 보통 3~5개의 과목을 선택한다. 하지만, 시험 일정이 과목별로 상이하므로 총 한 달여의 시간이 소요된다.

언어적 능력으로는 주로 어학적 능력과 문학적 능력을 측정한다. 이 중 어학적 능력에서는 주어진 텍스트에 대한 반응, 기술, 설명, 논평 능력, 언어 연구에 필요한 체계적인 틀을 알고 사용하는 능력, 맥락에 따라 언어 형식과 의미가 편차를 보이는 것을 이해하는 능력, 언어 사용과 관련된 개념을 이해하는 능력, 목적과 청자에 따르는 다양성에 대해 기술하는 능력, 언어 연구에 적합한 지식, 이해, 통찰을 명료하게 전달하는 능력 등을 평가한다.

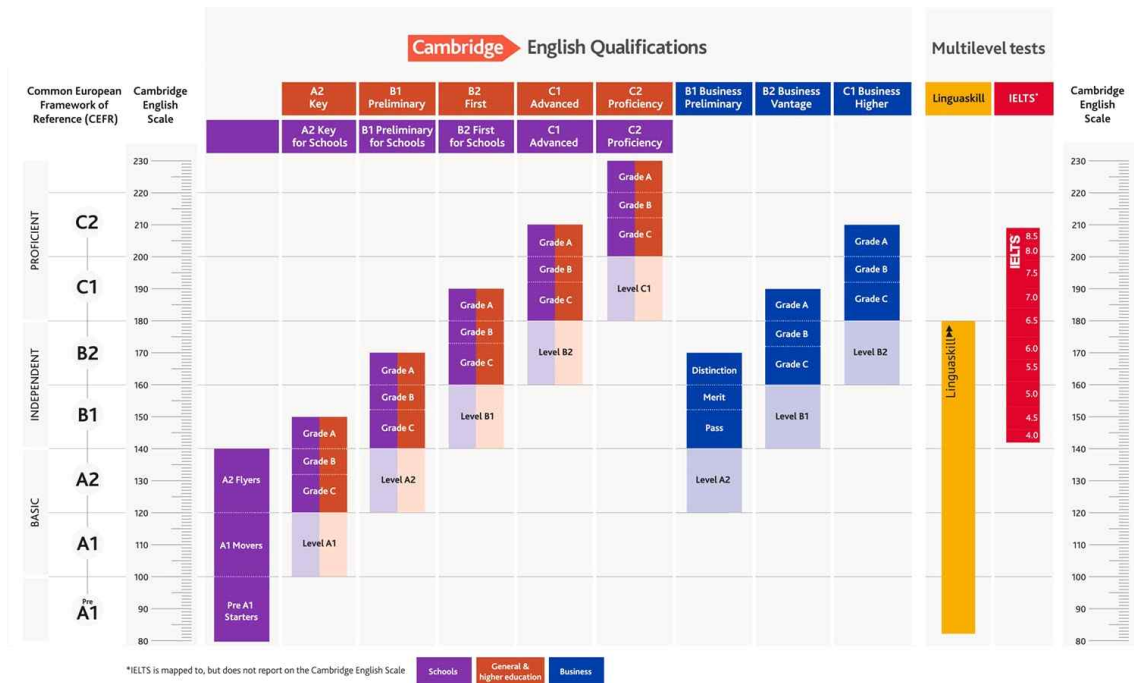
성적 평가는 국가 시험 기관에 의해 치러지는 시험과 코스 과정 내의 과제나 프로젝트에 대한 내부 평가로 이루어진다. 채점에 있어서는 ‘지식’, ‘비판력’, ‘구성력’, ‘자료의 연관성’ 등에 대한 비중치가 명시되어 있고, 외부 평가가 함께 이루어지기도 한다. 채점의 공정성이 문제가 되면서 내부 평가와 외부 평가 비율을 명시한다. 또한, 서로 다른 평가 위원회의 성적을 비교하기 위해 UMS(Uniform Mark Scale)을 사용한다. 각 과목의 성적은 A*, A, B, C, D, E로 평가되며 점수로 환산하기도 한다.

나. 케임브리지 대학의 영어 인증 시험(Cambridge English Qualifications, CEQ)

케임브리지 대학 출판·평가원(Cambridge University Press & Assessment)에서 주관하는 시험으로, 모국어 화자와 유학생 모두를 대상으로 일상생활에서 필요한 가장 기초적인 의사소통의 능력을 평가한다. 수준에 따라 ‘YLE Starters(Pre A1 Starters) - YLE Movers(A1 Movers) - YLE Flyers(A2 Flyers) - KET(Key English Test, A2) - PET(Preliminary English Test, B1) - FCE(First Certificate, B2) - CAE(Certificate in Advance English, C1) - CPE(Certificate of Proficiency, C2)’의 8단계로 이루어져 있으며, 뒤로 갈수록 난도가 높은 것에 속한다. 또한, 응시자가 학생인지 직장인인지에 따라서도 시험이 구분된다. KET, PET, FCE의 경우 일반 응시자가 아닌 학생을 대상으로 하는 학교용(for schools) 시험이 따로 존재한다. 그리고 직장인을 대상으로 하는 시험의 경우, ‘BEC Preliminary(B1 Business Preliminary) - BEC Vantage(B2 Business Vantage) - BEC Higher(C1 Business Higher)’의 순서로 단계를 구분한다.

영역은 말하기, 어휘 및 읽기, 듣기, 쓰기로 나뉜다. 결과는 응시자의 점수에 따라 A, B, C 등급으로 세분화하여 제공된다. 응시자의 점수가 응시한 단계의 한 단계 위의 수준에 준하면 A등급을, 한 단계 밑의 수준에 속하면 C등급을 받게 된다. 한 단계 밑의 수준에도 미치지 못하면 불합격으로 처리된다.

각 시험별로 국제적인 언어능력 검증의 기준으로 사용되는 유럽연합 공동 외국어 등급 기준(Common European Framework of Reference for Language, CEFR)에 따라 영어 능력 단계가 부여되는데, 구체적인 내용은 다음과 같다.



[그림 II -1] 케임브리지 대학의 영어 인증 시험(CEQ)의 영어 능력 단계

(3) 중국

가. 보통화능력시험(普通话水平测试)

교육부의 ‘보통화능력시험에 관한 시행 규칙’에 의거하여 국가시험기관(국가어문위원회, 국가교육위원회, 방송영상부)에서 주관하는 시험으로, 표준 중국어의 사용 및 숙련도를 평가한다. 이 시험은 말하기 과정에서 응시자의 표준 발음, 표준 어휘 및 문법의 사용을 평가하며, 웅변에는 초점을 두지 않는다. 시험은 총 100점 만점이며, 4개 문항으로 구성된다.

<표 II -11> 보통화능력시험의 문항 유형, 배점, 제한 시간

문항 유형	배점	제한 시간
단음절 단어 100개 읽기	10점	3분 30초
2음절 단어 50개 읽기	20점	2분 30초
400단어의 짧은 글을 소리 내어 읽기	30점	4분
말하기	40점	3분
계	100점	13분

‘단음절 단어 100개 읽기’는 응시자가 단어의 성모, 운모, 성조를 표준 중국어에 맞게 읽는지를 평가하며, ‘2음절 단어 50개 읽기’는 성모(聲母), 운모(韻母), 성조의 발음을 평가할 뿐만 아니라 상성(성조의 제3성)의 변동, 일화음, 경성(중국어에서 일정한 소리의 높낮이가 없고 짧고 약하게 내는 성조)의 발

음을 평가한다. ‘400단어의 짧은 글을 소리 내어 읽기’는 응시자의 발음, 음운의 변동, 억양 등의 평가에 중점을 두고 중국어로 작성된 자료를 읽는 능력을 평가한다. ‘말하기’는 주어진 자료 없이 응시자의 표준 중국어 사용을 평가한다. 주로 단방향 말하기를 수행하며 필요에 따라 검사자와 검사자 간의 양방향 대화로 보완한다.

이 평가에서 측정하는 숙련도는 3등급(1등급, 2등급, 3등급)으로 나뉘며, 각각이 다시 A와 B 두 등급으로 나뉘어 총 6등급으로 나뉜다.

나. 중국어능력시험(汉语综合应用能力水平测试, HANYU NENGLI CESHI, HNC)

이 시험은 교육부와 국가 언어 및 쓰기 위원회를 필두로 교육부 시험 센터에서 조직하고 시행하는 국가 종합 중국어 응용 능력 시험이다. 국어능력 시험으로서 중국어를 생활, 학업, 업무의 기본 언어로 사용하는 사람들의 중국어 응용 능력을 평가하는 것을 목적으로 한다.

응시자가 자신의 중국어 응용 능력을 이해하고 향상시키는 데 참고 자료가 되며, 직원의 중국어 능력이 요구되는 다양한 교육 기관에서 중국어 교육 및 훈련에 대한 참고 자료를 제공한다. 이 시험은 주로 내국인을 대상으로 하지만 중국에 있는 외국인도 응시할 수 있다. 하지만 외국인을 주 대상으로 설계되지는 않았다.

이 시험은 종합적인 언어능력을 평가하는 시험으로, 중국어 구어와 문어의 이해 능력과 표현 능력의 두 측면에서의 평가를 모두 포함한다. 따라서 구어 이해, 구어 표현, 문어 이해, 문어 표현의 4개 항목으로 구분하여 듣기, 말하기, 읽기, 쓰기의 4가지 측면에서 응시자의 중국어 응용 능력을 검사한다.

시험은 등급에 따라 달리 제시되는데, 등급은 총 6개로 나뉜다. 1~3급은 구두 능력 요구 사항에 중점을 둔 낮은 등급이고, 4~6급은 문어의 비중이 높은 등급이다. 시험은 오디오 및 비디오 시나리오의 형태로 컴퓨터에 제시되는 컴퓨터 기반 평가 방식으로 실시된다. 뉴스 방송의 클립이나 영화 및 TV 드라마의 대화가 제시되는 등 실제적인 문항으로 구성된다.

평가 결과는 중국어 구어 이해 능력(듣기), 중국어 구어 표현 능력(말하기), 중국어 문어 이해 능력(읽기), 중국어 문어 표현 능력(쓰기)의 4가지 응용 능력의 총점으로 산출될 뿐 아니라, 각 영역에 대한 상세한 평가도 제시된다. 또한, 응시자의 듣기, 말하기, 읽기, 쓰기에서 강점과 약점을 지적하고 방향 및 제안을 제시하는 진단 평가 보고서를 제공한다.

다. 한자응용능력시험(汉字应用水平测试)

한자응용능력시험은 교육부와 국가언어위원회가 주관하고 시행하는 표준화된 언어능력 시험이다. 이 시험은 중급 이상의 학력을 가진 사람들이 표준화된 한자를 매개로 읽기, 쓰기 표현 등의 활동에서 한자 숙달 및 사용 수준을 평가한다. 구체적인 기준은 국가에서 고시한 한자의 형태와 발음 규범에 따라 표준화된 한자를 정확하게 식별하여 사용할 수 있는지, 표준 한자를 중국어 발음과 현대 중국어에 따라 정확하게 식별하여 사용할 수 있는지 여부이다. 시험 범위는 ‘한자응용능력시험 문자표’에 있으며, 한자 상용도와 해당 집단의 인식률에 따라 A, B, C 3개의 하위표로 구분된다.

시험은 빈칸 채우기와 객관식 문항을 포함하는 필답 시험으로 제공된다. 문제지와 답안지가 분리되어 있으며, 문제지에는 답안의 조건, 예제 문제와 함께 시험 문제가 있고, 응시자는 답안지에 응답하는 방식이다. 총 130개 문항으로, 전체 응답 시간은 80분이다.

평가 영역은 ‘한자응용등급시험의 문자표’에 있는 모든 한자의 형태, 발음, 의미, 용법이다. 파트별 문

항의 유형 및 내용은 다음과 같다.

<표 II-12> ‘한자응용능력시험’의 문항 유형, 평가 내용, 문항 수

문항 유형	평가 내용	문항 수
파트 1: 빈칸 채우기 문항(쓰기)	문맥과 발음 프롬프트에 따라 한자 채우기	30개
파트 2: 객관식 문항	문맥에서 글꼴의 옳고 그름을 판단	40개
파트 3: 빈칸 채우기 문항(선택)	문장을 채울 올바른 한자 선택	10개
파트 4: 객관식 문항	발음의 정확성 또는 유사점과 차이점을 판단	30개
파트 5: 객관식 질문	단어의 의미 또는 유사점과 차이점을 판단	20개
계		130개

(4) 일본

가. 일본어 문장 이해·작성 능력 검정 시험(文章読解・作成能力検定)

일본 한자 능력 검정 협회(日本漢字能力検定協会)에서 주관하며, 자신의 생각을 문장으로 만드는 능력을 측정하는 기능 검정 시험이다. 2013년 하반기에 처음으로 실시되었다.

지원자를 대상으로 한 지필 평가 방식으로, 기초 능력, 이해 능력, 작성 능력의 세 분야로 나누어져 있다. 기초 능력은 어휘와 문법, 이해 능력(독해력)은 의미 내용, 자료 분석, 문장 구성을, 작성 능력은 구성, 표현, 종합 능력을 다루고 있다.

이 시험은 응시자가 작성하는 의견문, 논설문을 객관적으로 평가하며, 반드시 2인 이상의 채점자가 별도로 채점한다. 채점 결과가 불일치할 경우에는 다른 채점자가 판단하여 공정성을 유지한다. 평가 등급은 2~4급으로 나뉘며, 2급은 2급과 준2급으로 나뉘어 총 4개 등급으로 구성된다.

<표 II-13> ‘일본어 문장 이해·작성 능력 검정 시험’의 등급 및 난이도

등급	난이도
4급	읽기, 쓰기를 원활히 할 수 있으며, 기초적인 지적 언어활동을 위해 필요한 문장 독해력과 문장 작성력을 가진 정도
3급	고등학교에서의 적절한 이해, 표현 활동, 지적 언어활동을 위해, 사회에서의 커뮤니케이션 활동을 위해 필요한 문장 독해력과 문장 작성력을 가진 정도
준2급	더 높은 학습을 목표로, 사회에서의 유효한 커뮤니케이션을 위해 필요한 문장 독해력과 문장 작성력을 가진 정도
2급	고등교육에서 높은 교양을 주체적으로 체화하기 위해, 사회인으로 요구되는 문장 작성이 가능하도록 필요한 종합적 문장 독해력과 문장 작성력을 가진 정도

나. 일본 한자 능력 검정 시험(日本漢字能力検定)

일본한자능력검정협회(日本漢字能力検定協会)가 시행하는 한자 능력에 관한 검정 시험이다. 단순히

한자를 읽고 쓰는 지식만을 측정하는 것이 아니라, 한자의 의미를 이해하고 문장 속에서 적절히 사용하는 능력도 측정하고 있다. 1975년부터 시작되었으며, 1992년부터 문부 과학성이 인정하는 자격 제도가 되어 학점 인정이나 입학 우대에 사용하는 고등학교, 전문대학, 대학이 늘어나 널리 활용되고 있다.

이 시험은 일상생활에서도 통용되는 종합적인 한자 능력을 측정하고 있으므로, 기술식 중심으로 되어 있다. 한자 읽기, 쓰기 문제와 관용구의 구성, 의미, 오자 정정, 한자의 부수와 필순 등 넓은 분야에서 출제되고 있다. 시험에서의 문항 수는 80~120개 정도로 다양하며, 출제 문항 수의 제한이 없다.

필획을 바르고 정확하게 쓴 글자만을 채점 대상으로 하며, 흘려 쓰거나 난잡하게 쓴 글자는 채점 대상에서 제외된다. 글자의 종류와 글자의 체는 2~10급의 경우 상용한자표(2010년)를 기준으로 하며, 1급과 준1급은 한자 능력 검정 요람1/준1급대상(일본한자능력검정협회 발행)이 가리키는 표준자체, 허용자체, 구자체 일람표에 의해 채점된다.

읽기의 경우 2~10급은 상용한자표(2010년)에 의해 채점된다. 가나(仮名, 가타가나와 히라가나)의 경우 현대판 가나 사용법에 따라 채점되며, 오쿠리가나(送り仮名, 한문을 혼독하기 위해 한자의 오른쪽 아래에 다는 가나)의 경우 국가에서 공시한 오쿠리가나 사용법에 따라 채점한다. 부수는 한자능력검정요람 2~10급 대응(일본한자능력검정협회 발행)에 수록된 부수 일람표와 부수별 사용 한자를 기준으로 채점되며, 필순은 문부성이 만든 필순지도안내(1958년)에 의해, 상용한자의 필순은 한자능력검정요람 2~10급 대응에 수록된 상용한자필순일람을 따른다.

1급에서부터 10급까지로 되어 있고, 1급은 다시 1급과 준1급, 2급은 2급과 준2급으로 나누어져 모두 12단계로 되어 있다.

<표 II-14> ‘일본 한자 능력 검정 시험’의 등급 및 난이도, 대상 한자 수

등급	난이도	대상 한자 수
10급	초등학교 1학년 과정 수료자 정도	80자
9급	초등학교 2학년 과정 수료자 정도	240자
8급	초등학교 3학년 과정 수료자 정도	440자
7급	초등학교 4학년 과정 수료자 정도	640자
6급	초등학교 5학년 과정 수료자 정도	825자
5급	초등학교 6학년 과정 수료자 정도	1,006자
4급	중학교 재학자 정도	1,322자
3급	중학교 졸업자 정도	1,607자
준2급	고등학교 재학자 정도	1,940자
2급	고등학교 졸업자, 대학생, 일반인 정도 (상용한자를 모두 읽고 쓰고 활용하는 수준)	2,136자
준1급	대학생, 일반인 정도	약 3,000자
1급	대학생, 일반인 정도	약 6,000자

다. 일본어 검정 시험(日本語検定)

일본어 검정 위원회(日本語検定委員会)가 주관하는 시험으로, 일본어 지식과 운용 능력을 측정하는 검정 시험이다. 2007년부터 시행되었으며, 피험자 수는 연간 10만 명에 이른다. 2011년 7월에 문부과학성의 후원을 받았다. 이 자격 제도를 입시에서 우대하는 학교나 각종 연구의 일환으로 실시하는 기업들이 매년 늘어나고 있다.

이 시험은 지원자를 대상으로 한 지필 평가 방식으로 진행되며, 평가 영역은 일본어 운용에 필요한 6

개 영역인 어휘, 표기, 높임말, 한자, 단어의 의미, 문법으로 구성되어 있다. 총 문항은 90~120개이다.

평가 결과는 객관식 문항에 대한 정답표를 OMR 리더기로 채점하며, '정답인 점수의 합계/총 점수'로 산출되는 총합 득점과, '정답인 영역별 점수의 합계/영역별 총 점수'로 산출되는 영역별 득점으로 제시한다. 또한, 영역별 점수와 문제별 점수로 구분된다. 시험은 합격과 불합격으로 나뉘며, 인정, 준인정으로 표시되면 합격이다. 다음과 같은 총 7개의 등급으로 나뉜다.

<표 II-15> 일본어 검정 시험의 등급 구분

등급	수준
7급	초등학교 저~중학년 수준
6급	초등학교 중~고학년 수준
5급	중학생, 초등학교 고학년 수준
4급	고등학생, 중학생 수준
3급	사회인, 대학생, 고등학생 수준
2급	사회인, 대학생 수준
1급	사회인 수준

이때, 3~7등급의 경우는 총합득점률이 70% 이상이면 인정, 60~70%면 준인정이 되며, 1~2등급의 경우 총합 득점률이 80% 이상이면 인정, 70~80%이면 준인정이 된다. 그러나 6개 영역에서 균형 있게 점수를 받는 것이 필요하며, 7급을 제외하고 득점률이 일정 기준을 충족하지 못하는 영역이 있을 경우 불합격으로 판정된다.

(5) 유럽

가. 독일어 능력 시험(Zertifikat Deutsch: ZD)

세계 어느 나라에서든 인정되는 자격시험이다. 대표적 독일어권 국가인 독일, 오스트리아, 스위스의 세 기관이 협력하여 만든 시험으로, 독일어에 대한 기본 지식을 갖추고 실제 생활에 있어서 기본적인 대화를 할 수 있도록 검증하는 데 목적이 있다.

지원자를 대상으로 하여 필기시험과 구술시험으로 나누어 진행된다. 필기시험에 응시한 사람만 구술 시험을 치를 수 있다. 필기시험 및 구술시험의 평가 영역 및 소요 시간은 다음과 같다.

<표 II-16> 독일어 능력 시험의 평가 영역 및 소요 시간

시험 유형	평가 영역	소요 시간
필기시험 (Schriftliche Prüfung) : 총 150분	읽기(Leseverstehen) - 75점	70분
	문법(Sprachbausteine) - 30점	20분
	듣기(Höverstehen) - 75점	30분
	쓰기(Schriftlicher Ausdruck) - 45점	30분
구술시험 (Mündliche Prüfung) : 10~20분	처음 만났을 때의 대화(Kontaktaufnahme)	3분
	주제에 대한 대화(Gespräch über ein Thema)	6분
	과제 해결을 위한 대화(Lösen einer Aufgabe)	6분

채점은 모두 독일 현지에서 이루어진다. 총 300점 만점으로, 객관식 문항은 문항에 따라 5점, 2.5점,

1.5점의 배점이며, 쓰기 시험의 경우에는 채점 항목에 따라 부분 채점이 이루어진다. 말하기 평가에서는 시험 직후 시험관이 구두로 평가 결과를 전달해 준다. 쓰기 및 말하기의 채점 기준은 아래와 같다.

<표 II-17> 독일어 능력 시험의 유형에 따른 평가 영역

시험 유형	평가 영역
필기시험 (Schriftliche Prüfung)	지시 사항에 얼마나 충실한가에 따라 최고 5점이 주어진다. 특히 우수한 문장이 있을 경우 1~2점의 가산점이 주어진다. 내용 전개 및 문장의 자연스러움과 논리적인 순서, 그리고 적절한 호칭, 끝맺음, 날짜 등이 평가의 대상이다. 최고 5점이 주어진다. 문법, 맞춤법 등이 평가 대상이며 최고 5점이 주어진다.
구술시험 (Mündliche Prüfung)	어휘력, 주제에 맞는 표현능력 등에 따라 최고 4점이 주어진다. 대화에 얼마나 적절하게 참여하는가, 자기 의견을 얼마나 논리적으로 표명하는가 등에 따라 최고 4점이 주어진다. 문법에 얼마나 맞게 말하는가에 따라 최고 4점이 주어진다. 발음과 억양에 따라 최고 3점이 주어진다.

등급은 아래와 같이 네 등급으로 나뉘며, 179.5점부터는 불합격이다. 필기시험과 구술시험에서 각각 60% 이상의 점수를 받아야 합격할 수 있다. 즉, 필기시험에서 135점 이상, 구술시험에서 45점 이상을 받아야 한다.

<표 II-18> 독일어 능력 시험의 등급 구분

점수	등급
300~270	매우 우수(sehr gut)
269.5~240	우수(gut)
239.5~210	만족할만한(befriedigend)
209.5~180	충족한(ausreichend)
179.5~0	불합격(nicht bestanden)

나. 대입 독일어 능력 시험(DSH, Deutsche Sprachprüfung für den Hochschulzugang)

독일의 대학 입학시험에 해당하며, 독일의 각 대학에서만 칠 수 있다. 중급의 독일어 수업을 이수한 수료증을 제시해야 시험 자격이 주어진다.

일정한 자격을 갖춘 응시자의 신청에 의해 시험이 이루어진다. 필기시험과 말하기 시험으로 나누어 진행되며, 필기시험에서 57점 이상의 점수를 받았을 경우에만 말하기 시험을 칠 수 있다. 말하기 시험은 주어진 3~4개의 지문 중에서 본인이 원하는 지문을 선택하여 15분 동안 읽고 메모한 후, 2명의 시험관에게 자기 소개와 함께 문제에 대해 설명하고 질문과 대답을 하는 방식으로 진행된다.

필기시험은 읽고 이해하기와 문법(Leseverstehen und Wissenschaftssprachliche Strukturen), 듣고 이해하기(Höerverstehen), 작문(Textproduktion)의 세 영역으로 나누어진다. 따라서 실질적인 평가 영역은 읽고 이해하기와 문법, 듣고 이해하기와 작문, 말하기의 다섯 영역으로 볼 수 있다.

문항 수는 주관하는 대학에 따라 차이가 있다. 지필 평가의 총점은 100점이고, 총점과 영역별 점수를 함께 제시한다. 영역별 점수 배점은 아래와 같고, 문법 영역의 점수 비중이 다른 영역의 절반에 해당하는 점이 특징적이다. 말하기 평가의 총점도 100점이다.

<표 II-19> 대입 독일어 능력 시험의 평가 영역 및 소요 시간

시험 유형	등급(점수 비중)	소요 시간
지필 평가 (100점)	읽고 이해하기(28.57%)와 문법(14.28%)	90분
	듣고 이해하기(28.57%) (두 번의 듣기 대본 낭독 후 40분의 시간이 주어짐)	40분
	작문(28.57%)	70분
말하기 평가 (100점)	하나의 텍스트에 대한 말하기	20분

이 시험은 다수의 객관식 문항이 아니라 철저하게 과제 수행형 주관식으로 구성되어 있다. 따라서 컴퓨터 전산 처리 방식이 아닌 채점자의 직접 채점에 의해 채점이 이루어진다. 이로 인한 가장 큰 문제는 채점의 객관성 확보가 어렵다는 점이다. 각 대학에서는 문제 출제와 동시에 과제별로, 그리고 과제를 해결해 나가는 과정별로 상세한 채점 지침을 마련함으로써 이러한 문제에 대응하고 있다. 등급은 DSH 1, DSH 2, DSH 3으로 나뉘는데, 각각 필기시험과 말하기, 쓰기 시험에서 57% 이상, 67% 이상, 82% 이상 획득하였을 경우 얻는 등급이며, 67% 이상, 즉 67점 이상을 획득해야 대학 입학 자격이 주어진다.

(6) 기타 언어능력 측정 사례

가. OECD 국제 성인 문식력 조사(IALS, International Adult Literacy Survey)

국제 성인 문식력 조사(IALS)는 OECD에서 주관한 국제 규모의 성인 문식력 조사 프로그램이다. 1994년, 1996년, 1998년 3단계에 걸쳐 캐나다, 미국, 네덜란드, 폴란드, 스위스 등을 포함하여 20여 개 국가들의 16~65세 성인 68,755명을 대상으로 문식력을 조사하여, 국제 비교하였다.

각 가정에서 16세~65세의 연령대에 있는 성인 1명을 추출하고 다음으로 연령대, 교육 정도 등에 따라 분류하고 이러한 범위 안에서 표본이 지역을 대표할 수 있도록 구조화하였다. 전체적으로 볼 때 각 영역 안에서 대표성을 확보하면서 조직적으로 선정되도록 하였다.

조사 방식은 방문을 통한 면접 조사(배경 변인에 대한 설문과 문식력에 대한 면접 조사)를 사용하였다. 응답자들은 조사에 앞서 약 20분간 배경 설문지를 작성하게 되는데, 이 설문지는 참가자에 대한 일반적인 인구통계학적 정보(예: 성별, 연령대, 인종/민족)와 교육 경험, 노동 경험 및 문식력 관련 활동 등과 관련된 구체적인 정보를 수집하도록 고안되었다.

조사 문항은 예비 조사 6개 문항, 본 조사 3개 영역 108개 문항으로 이루어져 있다. 예비 조사는 본 조사에 들어가기 전에 실시하는 기초 단계 수준의 조사로, 조사에 응하는 모든 응답자가 공통적으로 작성한다. 예비 조사 2문항 이상을 수월하게 응답한 응답자는 이어서 본 조사를 작성한다. 만약 예비 조사 문항을 제대로 수행하지 못하면 본 조사에 들어가지 않고 문식력 조사는 종료된다. 본 조사는 7권으로 구성되어 있으며, 응답자는 7권 중 한 권을 작성하게 되어 있다. 각각의 조사지는 3개의 영역(산문 문식력, 문서 문식력, 수리 문식력)을 모두 포함하며, 대략 45분이 소요되도록 약 45개의 문항으로 구성

되었다. 조사 문항 설계 및 조사지 구성 시 불완비 블록 설계(balanced incomplete block)를 사용하여, 본 조사 문항 108개를 3개의 문식력 영역을 모두 포함하고 거의 동일한 응답 시간이 소요되는 7개의 블록으로 구획하여, 3개의 블록으로 각 조사지를 구성하였다.

문식력은 문항 반응 이론을 활용하여 점수화되어 처리되었다. 응답 결과를 점수화하여 단계별(level 1-level 5단계), 영역별(산문 문식력, 문서 문식력, 수리 문식력)로 비교 분석하였다.

나. OECD 국제 성인 문식력 및 생활 능력 조사(ALL, Adult Literacy and Life skills Survey)

국제 성인 문식력 및 생활 능력 조사(ALL)는 국제 성인 문식력 조사(IALS)의 후속 프로그램이다. 참가 국가의 16~65세 성인들을 대상으로 문식력 및 생활 기술 문식력이 조사되었으며, 이를 토대로 국제 비교가 이루어졌다. 2003년, 2006~2008년 두 단계에 걸쳐 이루어졌는데, 2003년에는 버뮤다, 캐나다, 이탈리아, 노르웨이, 누에보 레온(멕시코 북부), 스위스, 미국의 7개 국가가 참여하였고, 2006~2008년에는 호주, 헝가리, 네덜란드, 뉴질랜드 4개 국가가 참여하였다.

이 조사는 층화 다단계 확률 표집을 적용하여 표본을 추출하였는데, 해당 조사를 수행한 국가당 총 5,400개의 표본 수가 권장되었으며, 문식력의 영역 중 문제 해결 능력을 제외한 경우에는 3,420개의 표본 수가 권장되었다. 조사를 수행한 전체 국가의 응답자를 누적하면 총 74,650명이다. IALS와 마찬가지로 가정 방문을 통한 면접 조사로 진행되었으며, 6개의 예비 조사 문항을 수월하게 응답한 응답자만 이어서 본 조사를 작성하였다. 하지만, 본 조사의 구체적인 방식이나 조사지의 구성에서 다소 차이를 보인다. 배경 설문 또한 기본적인 인구통계학적 정보뿐 아니라 교육 성취도, 가정과 직장에서의 문식력 관련 활동, 노동 경험 정보, 정보 통신 기술 사용, 성인 교육 참여 및 문식력 자가 평가 등 기존 설문보다 항목이 조금 더 추가된 모습을 보인다.

본 조사는 IALS의 3개 영역(산문 문식력, 문서 문식력, 수리 문식력)에 ‘문제 해결력’을 더하여 총 4개의 영역으로 이루어졌으며, 산문 문식력 문항은 52개, 문서 문식력 문항은 48개, 수리 문식력 문항은 41개, 문제 해결력 문항은 19개로 총 160개의 본 조사 문항이 설계되었다. 이 문항들은 다시 8개의 블록으로 나누어졌는데, 100개의 문식력 문항(산문 문해와 문서 문해)은 4개의 블록으로, 41개의 수리 문식력 문항 및 19개의 문제 해결력 문항은 각각 2개의 블록으로 분할되었다. 이렇게 나누어진 블록을 2개씩 결합하여, 8개의 블록으로 총 28개의 소책자를 구성하였다. 블록 쌍의 조합은 1) 문식력 블록 2개, 2) 문식력과 수리 문식력 블록 각 1개, 3) 문식력과 문제 해결력 블록 각 1개로 구성되었으며, 문제 해결력과 수리 문식력을 함께 포함하는 소책자는 없었다. 그 대신, 동일한 블록 조합으로 구성된 소책자를 4쌍(8권) 설정하여, 응답자가 8개 중 2개 블록에 답하더라도 문식력의 4개 영역 모두에서 점수를 산출할 수 있도록 하였다. 각 응답자에게는 한 권의 소책자를 제공하였고, 각 소책자는 동일한 수의 응답자에게 제공하였으며, 소책자는 무작위로 배정되었다.

조사 결과는 문항 반응 이론을 활용하여 처리되었으며, 점수화하여 단계별(level 1-level 5단계), 영역별(산문 문식력, 문서 문식력, 수리 문식력, 문제 해결력)로 비교 분석하였다.

다. OPIC(Oral Proficiency Interview-Computer)

ACTFL(American Council on the Teaching of Foreign Languages)에서 주관하는 시험으로, 총 7개 어종 외국어(영어, 중국어, 스페인어, 러시아어, 한국어, 일본어, 베트남어)의 말하기 능숙도 평가를 목적으로 한다. 특정 분야에 치우치지 않는 언어 수행 능력을 중심으로 한 총체적인 언어 수행 능력 평

가를 지향한다.

이 시험은 지원자의 신청을 통해 평가가 이루어진다. 컴퓨터를 통해 진행되는 반(半) 직접 평가 방식을 활용하고 있어서, 응시자 개개인이 질문에 대한 답을 녹음하고 그 자료를 미국의 평가 서버에 전송하면 채점이 이루어져 결과가 통보된다.



[그림 II -2] OPIc의 평가 영역

OPIc은 4개 영역(Global Tasks/Functions, Context/Content, Accuracy/Comprehensibility, Text Type)에 걸쳐 언어능력을 총체적으로 측정하며, 문법(Grammar), 어휘(Vocabulary), 발음(Pronunciation) 등의 요소는 위 평가 영역 중 한 영역에 포함된다.

시험은 총 40분이 소요되며, 이에 앞서 20분 정도의 시간이 소요되는 오리엔테이션을 실시한다. 오리엔테이션은 배경 설문, 자기 평가, 사전 점검, 답변 연습의 네 가지로 이루어진다. 배경 설문은 직업, 경험, 관심 분야, 선호도 조사 등을 조사하기 위한 것이며, 이를 바탕으로 개인 맞춤형 문항을 출제한다. 자기 평가는 시험의 난이도 결정을 위한 것이며, 사전 점검은 질문의 청취 및 답변 녹음 기능을 점검하기 위한 것이다. 이어서 청취 및 답변 방법을 안내받은 후 답변 연습을 수행한다. 문항은 직업, 여가생활, 취미, 관심사, 스포츠, 여행 등에 대한 주제로 이루어지며, 개인별 맞춤 차등 문항으로 총 12~15개의 문항이 출제된다.

평가 방식은 절대 평가 체제이며, ACTFL 공인 평가자가 'ACTFL Proficiency Guidelines Speaking (Revised 2012)' 말하기 기준에 따라 평가한다. 'ACTFL Proficiency Guidelines Speaking'은 말하기 능숙도(Oral Proficiency)에 대한 ACTFL의 공식 언어능력 기준으로, 이에 따라 다음과 같은 9개의 등급을 구분한다.

<표 II -20> OPIc의 평가 영역

Level		등급별 특성
Advanced	Advanced Low	사건을 서술할 때 일관적으로 동사 시제를 관리하고, 사람과 사물을 묘사할 때 다양한 형용사를 사용한다. 적절한 위치에서 접속사를 사용하기 때문에 문장 간의 결속력도 높고 문단의 구조를 능숙하게 구성할 수 있다. 익숙하지 않은 복잡한 상황에서도 문제를 설명하고 해결할 수 있는 수준의 능숙도이다.
Intermed	Intermedi	개인에게 익숙하지 않거나 예측하지 못한 복잡한 상황을 만날 때,

iate	ate HIGH	대부분의 상황에서 사건을 설명하고 문제를 효과적으로 해결하곤 한다. 발화량이 많고, 다양한 어휘를 사용한다.
	Intermedi ate MID	일상적인 소재뿐 아니라 개인적으로 익숙한 상황에서는 문장을 나열하며 자연스럽게 말할 수 있다. 다양한 문장 형식이나 어휘를 실험적으로 사용하려고 하며, 상대방이 조금만 배려해 주면 오랜 시간 대화가 가능하다.
	Intermedi ate LOW	일상적인 소재에서는 문장으로 말할 수 있다. 대화에 참여하고 선호하는 소재에서는 자신감을 가지고 말할 수 있다.
Novice	Novice HIGH	일상적인 대부분의 소재에 대해서 문장으로 말할 수 있다. 개인정보에 대해 질문을 하고 응답을 할 수 있다.
	Novice MID	이미 암기한 단어나 문장으로 말하기를 할 수 있다.
	Novice LOW	제한적인 수준이지만 영어 단어를 나열하며 말할 수 있다.

* Intermediate Mid의 경우 Mid < Mid2 < Mid3으로 세분화하여 제공.

3) 국내외 듣기·말하기 능력 실태 조사 사례

이상의 국내외 국어능력 제반을 조사하거나 평가한 사례들을 검토한 결과, 주로 읽기 능력을 중심으로 한 서답형 평가가 많았음을 확인할 수 있다. 듣기나 말하기 능력은 국어능력을 구성하는 다른 능력들과 함께 다루어진 경우가 대다수였으며, 듣기나 말하기 능력만을 대상으로 한 조사는 극히 일부에 불과하였다. 비록 그 수가 많지는 않지만, 듣기나 말하기 능력을 조사한 사례들은 본 연구와 깊은 관련을 맺기에 해당 사례들만을 이하 구체적으로 살펴보고자 한다.

(1) 듣기·말하기 능력을 포함하여 국어능력 제반을 조사한 경우

가. 2018년 국민의 국어능력 실태 조사

‘2018 국민의 국어능력 실태 조사’에서 듣기 능력과 말하기 능력은 전체 국어능력의 일부로서 다루어졌다. 총 116분이 소요되는 48개 문항 중 듣기 문항은 10개, 말하기 문항은 3개였으며, 각각 15분간 능력을 평가하였다. 또한, 듣기와 말하기 영역의 배점은 각각 150점으로, 총점이 1,030점임을 고려한다면 큰 비중을 차지하지는 않는 것으로 볼 수 있다.

듣기 영역에서는 듣기 능력을 크게 ‘사실적 듣기’, ‘추론적 듣기’, ‘비판적 듣기’로 나누고, 이를 다시 ‘지시의 이해 능력, 세부 정보의 파악, 핵심 정보의 파악’, ‘내용의 추리, 심리 도는 태도의 추리, 적용’으로 나누어 이들 능력이 고루 평가하고자 하였다. 이때, ‘지시, 뉴스, 면담, 대화, 강연, 뉴스, 토론’의 담화 유형 또한 골고루 반영되도록 하였다. 말하기 영역에서는 말하기 능력을 ‘설득하는 말하기’, ‘설명하는 말하기’로 나눈 뒤, 설득하는 말하기는 ‘주장’과 ‘요청’으로 세분화하였다. 말하기 영역 또한 담화 유형(일대일 대화, 토론), 말하기 상황(공적, 개인적)을 비교적 고르게 반영하였다.

듣기는 선택형(객관식) 문항으로 구성되었으며, 말하기는 수행형(주관식)으로, 조사 대상자가 직접 말한 것을 가구 방문 조사원이 녹음하여, 이를 전문가가 채점하는 방식으로 이루어졌다. 채점자 간 신뢰도

제고를 위해 10인으로 구성된 채점단을 대상으로 워크숍을 실시하였으며, 문항당 2인의 교차 채점 방식의 타당성과 적합성을 검증하기 위해 교차 채점 및 공통 채점 방식으로 진행하였다. 채점 중 중간 점검을 통해 채점자 간 신뢰도를 확인하고, 부족한 지점을 보완하는 방식으로 채점 방식의 타당성을 보완하였다.

이 조사에서는 듣기와 말하기 능력을 세부적으로 나누고, 이를 여러 가지 담화 유형과 함께 골고루 측정하고자 하였다는 점, 말하기 채점에 있어서 신뢰도를 높이기 위하여 혼합된 채점 방식을 사용하고 있다는 점에서 참조가 된다.

나. 국어능력인증시험(ToKL)

‘국어능력인증시험(ToKL)’은 총체적인 국어능력을 평가하는 것을 목적으로 하며, 평가 영역을 크게 ‘언어 기초 영역’, ‘언어 기능 영역’, ‘사고력 영역’으로 구분하고 있다. 이 중에서 듣기 능력은 읽기 능력, 쓰기 능력과 함께 ‘언어 기능 영역’에 포함된다.

전체 문항은 객관식 80문항과 주관식 10문항으로 이루어져 있으며, 듣기 영역의 문항 수는 총 15개이다. 듣기 영역은 모두 객관식으로 구성되어 있으며, 배점은 모두 2점씩이다. 듣기 능력의 세부 평가 영역은 크게 1) 사실적 이해, 2) 추론, 3) 비판, 4) 창의로 나뉘며, 다양한 상황 설정을 통한 듣기 능력을 종합적으로 평가하는 것을 목표로 한다. ‘사실적 이해’는 다시 ‘내용의 파악, 발화 상황의 이해’, ‘추론’은 다시 ‘의도 및 상황의 파악’, ‘생략된 정보의 추론’, ‘구체적 상황에 적용하기’, ‘비판’은 ‘내용의 적절성 평가’, ‘근거의 적절성 평가’, ‘창의’는 ‘적용 및 대안 탐색’으로 이루어진다. 각 세부 영역별 발문의 예시는 <표 II-21>과 같으며, 세부 영역 중 ‘(1) 사실적 이해’의 ‘① 내용의 파악’ 영역의 예시 문항을 보이면 다음과 같다.

<표 II-21> 국어능력인증시험의 영역별 발문 예시

영역		발문 예시
(1) 사실적 이해	① 내용의 파악	- 강연의 핵심적인 주장으로 가장 적절한 것은? - 남자의 주장으로 가장 핵심적인 것은?
	② 발화 상황의 이해	- 이 대화의 전개 양상으로 설명으로 가장 적절한 것은? - 이 강연자의 주장이 아닌 것은?
(2) 추론	① 의도 및 상황의 파악	- 여자의 마지막 말이 의미하는 것으로 적절한 것은? - 토론자들이 공통적으로 전제하고 있는 것은? - 남자의 주장의 근거로 가장 적절한 것은?
	② 생략된 정보의 추론	- 이 토론에 이어질 여자의 말로 가장 적절한 것은? - 이 강연의 뒤에 이어질 내용으로 적절한 것은? - 강연에서 생략된 말로 가장 적절한 것은? - 이 방송에서 삽입될 인터뷰 내용으로 가장 적절한 것은?
	③ 구체적 상황에 적용하기	- 이 강연의 주제와 관련이 깊은 실례로 가장 적절한 것은? - 남자의 질문에 답하기 위해 제시해야 할 자료로 적절하지 않은 것은? - 담화 내용을 듣고 보인 반응으로 가장 적절한 것은?
(3) 비판	① 내용의 적절성 평가	- 남자의 발언에서 모순된 내용으로 적절한 것은? - 이 토론의 문제점으로 가장 적절한 것은?

	② 근거의 적절성 평가	- 연사의 태도의 문제점으로 가장 적절한 것은? - 남자의 주장에 대한 비판으로 가장 적절한 것은?
(4) 창의	① 적용 및 대안 탐색	- 강연자의 주장에 반박하는 내용을 하나의 주제문으로 쓰시오. - 제시된 이야기에 대해서 주인공의 입장을 옹호하거나 비판하는 근거를 쓰시오.

예제
※ 다음은 대화의 일부입니다. 잘 듣고 물음에 답하세요. 남> 이번 영화, 최희진이 나온다지? 보러 가고 싶은데, 영화가 어떤지 모르겠어? 여> 최희진이라면 뭐 비슷하겠지, 분위기나 스타일이 늘 비슷하잖아. 남> 그것 때문에 보는 거야. 뭘 해도 예쁘게 나오니까... 여> 그런 게 불만이지, 난. 영화 보는 게 아니라 티비 토크쇼 보는 것 같아. 배우의 외모는 튀어나오지만, 그 덕분에 극 속의 주인공은 묻히잖아. 캐릭터가 없어서 싫어. 남> 배우 그 자체로 캐릭터인 거지.. 여> 외모로 밀고 나가는 것도 한 두 번이지, 역할은 다른데, 항상 같은 모습으로만 나오니까, 저걸 연기라고 해야 하나 싶은 생각도 들어. 영화도 별로 성공할 것 같지도 않고. 남> 극 중 캐릭터는 없어도, 배우 최희진의 캐릭터만 있으면 됐지..(웃음) 여> 너 같은 관객들이 보는 영화일 수도 있겠네. 하지만, 그 배우가 늙으면 그걸로 연기 인생은 끝이겠구나, 휴.. 잘 들으셨지요? 이 대화를 통해서 알 수 있는 것은 무엇입니까? ① 여자는 배우의 외모에 불만을 가지고 있다. ② 여자는 배역에 몰입된 배우의 연기를 보고 싶어 한다. ③ 남자는 티비 토크쇼에 여배우가 출연하는 것을 싫어한다. ④ 남자는 여배우의 캐릭터가 영화의 흥행요소라고 생각한다. ⑤ 남자는 배우의 연기보다는 영화의 내용에 관심을 가지고 있다. 「정답 및 해설」 여배우의 외모를 중심으로 한, 내용 없는 영화에 대해 남자는 긍정적으로 평가한 반면, 여자는 부정적으로 보고 있다. 여자의 경우 외모만 내세운 영화에는 진지한 연기가 없다고 판단해서이다.

[그림 II -3] 국어능력인증시험의 문항 예시

다. KBS 한국어능력시험

이 시험은 원만하고 능률적인 언어생활에 요구되는 국어 사용 능력과 국어 관련 교양을 측정하는 것을 목표로 한다. 80~100개의 문항이 출제되며, 모두 객관식 5지선다형이다. 문항의 수준에 대해 주최 측은 한국의 고교 수준의 국어교육을 정상적으로 받은 사람이 풀 수 있는 정도의 수준이라고 밝히고 있다. 평가 영역은 문법, 이해, 표현, 창안, 국어문화 영역으로 나뉘는데, 듣기는 ‘이해’ 영역에, 말하기는 ‘표현’ 영역에 속한다. 듣기 능력의 문항들은 강의, 강연, 뉴스, 토론, 대화, 인터뷰 자료 등 다양한 구어 담화를 듣고 문제를 해결하는 방식으로 구성된다. 말하기 능력의 문항들은 발표, 토론, 협상, 설득, 논증, 표준화법(언어예절, 호칭어와 지칭어 사용 등) 등의 다양한 말하기 상황과 관련된 능력을 평가한다. 정확한 발음과 관련하여 표준발음법 관련 문항도 포함되어 있다.

한편, 이 시험은 등급의 부여 과정에서 말하기 수행 과제를 참고하고 있다는 점이 주목된다. 모든 국어능력을 객관식으로 평가하고 있으므로, 준거 집단을 추출하여 이들의 언어 수행 능력을 평가함으로써 객관식 시험의 등급화에 참조하는 방식이다. 이 과정은 1) 준거 집단 추출, 2) 언어 수행 평가 실시, 3) 정성 지표(등급표) 구안, 4) 언어 수행 능력 데이터베이스 구축, 5) 등급 처리 부여의 5단계 절차를 거친다.

우선, ‘준거 집단(reference group) 추출’ 단계에서는 KBS 한국어능력시험에 3회 이상 응시한 대단

위 응시자 중에서 언어 수행 능력 평가 데이터베이스를 구축하기 위하여 주기적으로 준거 집단(reference group)을 추출한다. 점수 변동이 안정된 응시자를 대상으로 각 점수대별로 고르게 분포하도록 추출하며, 단순 무작위추출을 원칙으로 한다. 이후, 이들을 대상으로 실제 직무와 관련된 말하기 및 쓰기 과제를 부여하여 실제 직무를 수행할 수 있는 언어 수행 능력을 평가하는데, 이것이 ‘언어 수행 평가 실시’ 단계이다. 과제는 다음과 같이 네 가지 영역으로 구분하며 특정 영역에 종사하는 응시자가 유리하지 않도록 수행 과제를 선정한다.

<표 II -22> KBS 한국어능력시험의 준거 집단 대상 수행 과제

구분	과제 영역	과제 사례
쓰기	조건 제시형 과제	기획서 쓰기(쓰기①)
	자유 과제	자유 작문(쓰기②)
말하기	조건 제시형 과제	설명(프레젠테이션)(말하기①)
	자유 과제	설득(대담)(말하기②)

다음으로, 수행 평가 심사관으로 참여한 전문가 집단 협의를 통해 언어 수행 능력의 성취 수준을 기술하는 정성 지표(등급표)를 구안하는 ‘정성 지표(등급표) 구안’ 단계를 거친다. 언어 수행 능력 등급표는 하위 단계에 대한 기능의 습득이 상위 단계의 기능 습득으로 발전하는 구조를 취하되 등급별로 성취해야 할 지식 수준과 기능 수준을 구별하여 기술하며, 이러한 정성 지표는 사회적 요구를 반영하여 주기적으로 갱신할 수 있다. 이후, ‘언어 수행 능력 데이터베이스 구축’ 단계에서는, 준거 집단의 언어 수행 능력을 체크리스트로 계량화하여 다수 평가관의 평균 점수를 부여한다. 평가자 간 편차가 큰 점수를 보정하는 과정을 거쳐 다음과 같이 언어수행능력 데이터베이스(DB)를 구축한다.

<표 II -23> KBS 한국어능력시험의 언어 수행 능력 데이터베이스(사례)

관리번호	국어능력시험 점수 (990점 만점)			언어 수행 능력 점수 (100점 만점으로 환산)			
	3회	4회	5회	쓰기①	쓰기②	말하기①	말하기②
A-001	780	810	805	85.33	90.67	78.00	82.33

마지막으로, ‘등급 처리 부여’ 단계에서는 이전 단계에서 구축한 정성 지표 및 준거 집단의 국어능력 시험 점수와 언어 수행 능력 시험 점수를 분석하여 선형 방정식을 도출한다. 결국 객관식 문항을 통해 평가한 듣기·읽기(Listening and Reading) 능력과 수행 평가를 통해 평가한 말하기·쓰기(Speaking and Writing) 사이의 관계를 하나의 식으로 도출하는 것이 핵심이다. 이를 바탕으로 전체 응시자의 국어능력시험 점수(990점 만점)에 해당하는 추정 성취도를 도출한 다음, 정성 지표에 맞게 국어능력 등급을 부여한다.

이 시험은 비록 듣기, 말하기 능력을 객관식 문항만으로 평가하였다는 점에서 일정한 한계가 존재한다. 하지만 이러한 한계를 보완하고자, 국어 사용 능력 평가 점수와 등급이 실제 언어 수행 능력을 얼마나 예측할 수 있느냐를 토대로 등급화를 실시한다는 것이 특징이다. 그리고 이를 위하여, 준거 집단을 대상으로 수행형 평가를 실시하고, 누적된 응시자 점수와 실제 언어 수행 능력 사이의 상관관계를 DB로 구축한다는 점이 주목된다. 또한, 매회 시험마다 문항의 특성이 상이함을 반영할 수 있도록 등급을 나누

는 경계 점수를 이와 연동시키는 등 쉽게 출제된 시험에 응시한 수험생이 자신의 능력보다 높은 등급을 받는 것, 어렵게 출제된 시험에 응시한 수험생이 자신의 능력보다 낮은 등급을 받는 것 등의 모순을 해결하고, 동점자라도 변별력이 높은 문항의 정답률이 높은 수험생이 더 높은 등급을 받도록 주의를 기울였다는 점을 참조할 수 있다.

라. 한국어능력시험(TOPIK)

이 시험은 한국어를 모국어로 하지 않는 재외동포와 외국인의 한국어 사용 능력을 측정한다. 지필 평가 방식으로 진행되며, TOPIK I의 경우에는 듣기, 읽기 영역을, TOPIK II의 경우에는 듣기, 읽기, 쓰기 영역을 평가한다. 문항은 선택형과 서답형으로 구성되어 있고, 서답형 문항은 TOPIK II의 쓰기 영역에만 포함된다. 즉, 듣기 능력은 모두 서답형 문항을 통해 측정되는데, TOPIK I의 경우에는 40분이 소요되는 30개 문항, TOPIK II의 경우에는 60분이 소요되는 50개 문항으로 평가가 진행된다.

마. 케임브리지 대학의 영어 인증 시험(Cambridge English Qualifications, CEQ)

모국어 화자와 유학생 모두를 대상으로 일상생활에서 필요한 가장 기초적인 의사소통의 능력을 평가하는 시험이다. 수준에 따라 ‘YLE Starters(Pre A1 Starters) - YLE Movers(A1 Movers) - YLE Flyers(A2 Flyers) - KET(Key English Test, A2) - PET(Preliminary English Test, B1) - FCE(First Certificate, B2) - CAE(Certificate in Advance English, C1) - CPE(Certificate of Proficiency, C2)’의 8단계로 이루어져 있으며, 뒤로 갈수록 난도가 높은 것에 속한다. 여기서는 가장 높은 단계라고 할 수 있는 ‘CPE’ 시험의 듣기, 말하기 영역을 자세히 살펴본다.

‘CPE’ 시험의 영역은 읽기 및 영어 사용, 쓰기, 듣기, 말하기로 나뉜다. 각 영역의 문항 수 및 소요 시간은 다음과 같다.

<표 II -24> 케임브리지 대학의 영어 인증 시험의 영역별 문항 수 및 소요 시간

영역	문항 수	소요 시간
읽기 및 영어 사용 (Reading and Use of English)	7개 파트, 53개 문항	90분
쓰기 (Writing)	2개 파트	90분
듣기 (Listening)	4개 파트, 30개 문항	40분
말하기 (Speaking)	3개 파트	응시자 한 쌍당 16분

듣기 영역은 다음과 같이 다양한 녹음 자료와 과제가 제시된다. 파트 1, 파트 3, 파트 4의 문장은 선택형, 파트 3의 문항은 서답형이다. 파트에 상관없이 녹음은 모두 2번씩 들려 준다.

<표 II -25> 케임브리지 대학의 영어 인증 시험의 듣기 영역의 내용, 문항 형식, 문항 개수

파트	내용	문항 형식	문항 개수 (배점)
파트 1	녹음 자료: 3개의 녹음(각 1분) 과제: 각 녹음마다 3지선다의 객관식 질문 2개에 응답.	선택형	6개(1점)
파트 2	녹음 자료: 독백(3~4분) 과제: 불완전한 형태의 문장이 제시되며, 녹음을 듣고 문장의 빈칸을 채움.(1개~3개 단어)	서답형 (빈칸 완성형)	9개(1점)
파트 3	녹음 자료: 상호 작용하는 화자들의 대화(3~4분) 과제: 녹음을 듣고 4지선다의 객관식 질문 4개에 응답.	선택형	5개(1점)
파트 4	녹음: 5개의 짧은 주제 독백(약 35초) 과제: 두 가지 다중 연결 문항이 제시되며, 녹음을 듣고 각 문항에서 8개 목록 중 5개의 올바른 선택지를 연결.	선택형 (다중 짝 맞추기형)	10개(1점)

이 중 파트 4의 문항 형식이 다중 짝 맞추기(Multiple matching)로 다소 독특하다. 5개의 짧은 주제 독백에 대하여 2개의 문항이 제시되는데, 각 문항에서 응답자는 총 8개의 목록 중에 올바른 5개의 선택지를 골라 연결하여야 한다.

Part 4

You will hear five short extracts in which different people are talking about taking a gap year – the time which some young people take off from their studies to gain other experience.

TASK ONE

For questions 21 – 25, choose from the list (A – H) what each speaker did during their gap year.

TASK TWO

For questions 26 – 30, choose from the list (A – H) which benefit of having a gap year each speaker mentions.

You will hear the recording twice. While you listen, you must complete both tasks.

A taught a language

B travelled widely

C worked in an office

D did voluntary work

E taught a sport

F worked in a hotel

G worked as a tour guide

H went on an organised expedition

Speaker 1 21

Speaker 2 22

Speaker 3 23

Speaker 4 24

Speaker 5 25

A making long-lasting friendships

B overcoming financial problems

C increased independence

D improved linguistic skills

E dealing with disappointment

F a renewed sense of adventure

G a more mature approach to studying

H a chance to consider future plans

Speaker 1 26

Speaker 2 27

Speaker 3 28

Speaker 4 29

Speaker 5 30

[그림 II -4] 케임브리지 대학의 영어 인증 시험의 듣기 문항(예시)

말하기 영역은 세 파트로 구성되어 있으며 다른 응시자와 함께 응시한다. 2명의 심사위원이 시험을 주재하는데, 한 명은 시험을 진행하는 역할을, 다른 한 명은 응시자의 말을 듣고 필기하는 역할을 수행한다.

<표 II -26> 케임브리지 대학의 영어 인증 시험의 말하기 영역의 내용 및 과제 형식

파트	내용	과제 형식	소요 시간
파트 1	심사위원과 각 응시자 간의 대화. 심사위원은 각 응시자에게 차례로 일련의 질문을 하여 응시자 본인에 대해 이야기할 기회를 줌.	인터뷰	2분
파트 2	심사위원이 몇 가지 말하기 지침과 하나 이상의 사진을 제시. 1) 사진에 대한 응시자의 반응에 초점을 맞춘 질문에 대답(1분). 2) 다른 응시자와의 의사 결정 진행.	의사 결정	4분
파트 3	1) 심사위원이 응시자 한 명에게 질문과 몇 가지 아이디어가 적힌 카드를 제공하고, 응시자는 말하기를 수행(약 2분). 2) 다른 응시자는 응시자의 말하기에 대해 논평. 3) 심사위원은 응시자들에게 같은 주제에 대해 질문. 4) 다른 응시자도 1)~3)의 동일한 절차를 진행. 5) 이후 심사위원은 두 사람과 함께 토론을 진행.	긴 말하기와 토론	10분(각 응시자별로 2분씩 말하기를 진행한 후 약 6분간 토론 진행).

바. 중국어능력시험(汉语综合应用能力水平测试, HANYU NENGLI CESHI, HNC)

이 시험은 주로 내국인을 대상으로 하지만 중국에 있는 외국인도 응시할 수 있다. 하지만 외국인을 주 대상으로 하여 설계된 것은 아니다. 총 듣기, 말하기, 읽기, 쓰기의 4가지 측면에서 응시자의 중국어 응용 능력을 측정한다.

시험은 등급에 따라 달리 제시되는데, 등급은 총 6개로 나뉜다. 1~3급은 구두 능력 요구 사항에 중점을 둔 낮은 등급이고, 4~6급은 문어의 비중이 높은 등급이다. 시험은 오디오 및 비디오 시나리오의 형태로 컴퓨터에 제시되는 컴퓨터 기반 평가 방식으로 실시된다. 뉴스 방송의 클립이나 영화 및 TV 드라마의 대화가 제시되는 등 실제적인 문항으로 구성된다.

듣기 영역은 들은 문단의 내용을 기반으로 하여 단항 선택이나 다항 선택 문항을 제시하며, 대화나 독백을 제시하는 일반적인 영어 듣기 평가와 유사하다. 말하기 영역에서는 특정 주제를 담은 동영상을 제시하고, 이것의 내용을 기반으로 하여 구두로 설명하는 과제가 제시된다. 제시문의 예시는 다음과 같으며, 아래에서 볼 수 있듯 제시문에서 주제를 해석하기 위해서는 어느 정도의 읽기 능력과 사고력이 요청된다.

<표 II -27> 중국어능력시험의 말하기 영역의 제시문(사례)

제시문
"원숭이 5마리를 우리에게 가두었고 실험자는 바나나를 우리에게 넣고 원숭이가 있을 때마다 모든 원숭이에게 수돗물을 뿌렸다. 천천히 원숭이들은 바나나를 가지러 가는 것을 멈췄다. 그런 다음, 새장에 있는 원숭이를 새 원숭이로 교체하였다. 새 원숭이가 바나나를 가져가면 다른 4마리 원숭이가 달려가서 바나나를 때렸다. 결과적으로 새로운 원숭이는 감히 바나나를 가져갈 수 없었다. 이렇게 해서 원래의 5마리의 원숭이는 점차 교체되었고, 새로운 5마리의 원숭이 중 누구도 감히 바나나를 가져갈 수 없었지만 이유를 알지 못했다."

응시자들은 위 제시문을 바탕으로 금기, 전통, 집단 무의식, 전체주의 등에 대해 반성하고 자신의 이해에 대해 이야기하도록 요청받는다. 이처럼 제시문 이해에 다소의 읽기 능력을 동원하도록 하고, 이에 대한 자신의 의견을 설명하도록 한다는 점이 특징적이다.

(2) 듣기·말하기 능력에 초점화하여 조사한 경우

가. 보통화능력시험(普通话水平测试)

이 시험은 말하기의 내용을 평가하기보다는 표준 중국어(보통화)의 숙달과 사용을 평가한다. 즉, 말하기 과정에서 응시자의 표준 발음, 표준 어휘 및 문법의 사용을 평가한다.

구술 평가(단어 읽기, 글 읽기, 말하기) 방식으로 진행되며, 총 100점 만점의 4개 문항으로 이루어진다. 각 문항은 '단음절 단어 100개 읽기', '2음절 단어 50개 읽기', '400 단어의 짧은 글을 소리 내어 읽기', '말하기'로, 각 문항이 평가하는 요소를 제시하면 다음과 같다.

<표 II -28> 보통화능력시험의 문항 유형 및 평가 요소

문항 유형	평가 요소
단음절 단어 100개 읽기 (10점, 제한 시간 3분 30초)	응시자가 단어의 성모, 운모, 성조를 표준 중국어에 맞게 읽는지를 평가함. (단어는 두 번 읽을 수 있음. 즉, 수험자가 첫 번째 발음에서 실수하면 다시 읽고 두 번째 발음에 따라 판단.)
2음절 단어 50개 읽기 (20점, 제한 시간 2분 30초)	성모(聲母), 운모(韻母), 성조의 발음을 평가할 뿐만 아니라 상성(성조의 제3성)의 변동, 열화음, 경성(중국어에서 일정한 소리의 높낮이가 없고 짧고 약하게 내는 성조)의 발음을 평가
400 단어의 짧은 글을 소리 내어 읽기 (30점, 제한 시간 4분)	응시자의 발음, 음운의 변동, 억양 등의 평가에 중점을 두고 중국어로 작성된 자료를 읽는 능력을 평가
말하기 (40점, 최소 3분 이상, 최대 4분)	주어진 자료 없이 응시자의 표준 중국어 사용을 평가. (단방향 말하기가 주된 방법이며 필요에 따라 검사자와 검사자 간의 양방향 대화로 보완.)

각 문항의 채점 방식을 정리하면 다음과 같다. 아래 <표 II-29>를 보면 알 수 있듯, 만점을 기준으로 오류의 횟수를 계량하여 감점하는 방식으로 채점을 진행한다.

<표 II-29> 보통화능력시험의 채점 방식

문항 유형	채점 방식
단음절 단어 100개 읽기	전체 점수의 10%, 10점 단어의 성모, 운모, 성조를 잘못 발음한 경우: 0.1점 감점 발음에 결함이 있는 경우: 단어당 0.05점 감점
2음절 단어 50개 읽기	전체 점수의 20%, 20점 음절의 성모, 운모, 성조를 잘못 발음한 경우: 0.2점 감점 발음에 명백한 결함이 있는 경우: 매회 0.1점 감점
400 단어의 짧은 글을 소리 내어 읽기	전체 점수의 30%, 30점 각 자료의 처음 400 단어(구두점 제외)를 기준으로 다음을 누적하여 채점. - 발음 오류 1개당: 0.1점 감점 - 단어 누락 시: 0.1점 감점 방언 억양에 대하여 다음과 같은 기준으로 1회 감점 - 심각한 경우: 3점 감점 - 명확한 경우: 2점 감점 - 경미한 경우: 1.5점 감점 멈춤 및 부적절한 문장 분할 시 1점 감점 너무 빠르거나 느린 말은 1회 2점 감점
말하기	전체 점수의 40%, 40점 (1) 음성의 정확성: 20%, 20점 음성 오류와 방언 사용이 없는지를 평가 음성 측면에서 2급(또는 2급 미만)으로 판정되면, 총점이 96점을 초과하더라도 1급 A에 진입할 수 없음. (2) 어휘 문법 표준화 정도: 10%, 10점 (3) 자연스럽고 유창한지의 정도: 10%, 10점

나. OPIC(Oral Proficiency Interview-Computer)

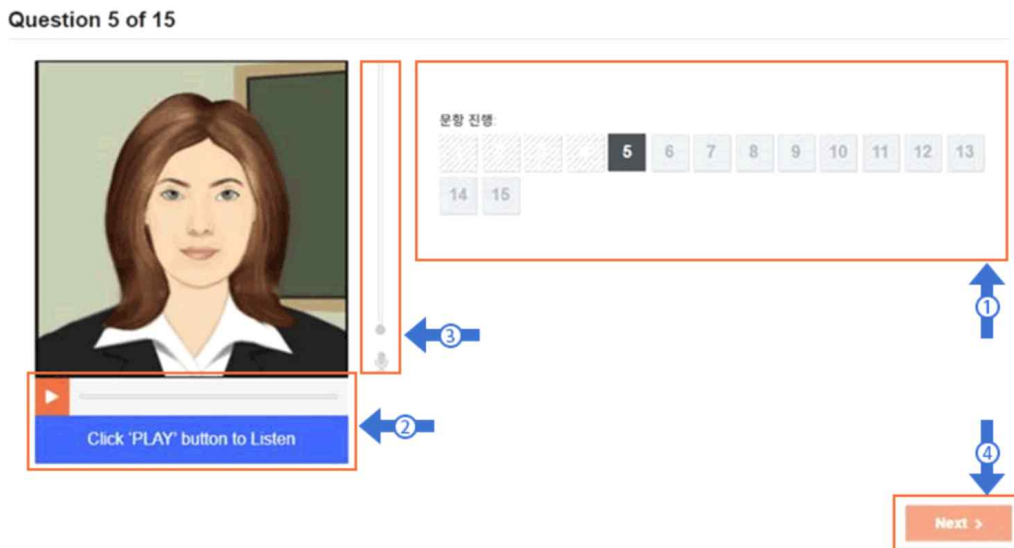
ACTFL(American Council on the Teaching of Foreign Languages)에서 주관하는 시험으로, 총 7개 어종 외국어(영어, 중국어, 스페인어, 러시아어, 한국어, 일본어, 베트남어)의 말하기 능숙도 평가를 목적으로 한다. 특정 분야에 치우치지 않는 언어 수행 능력을 중심으로 한 총체적인 언어 수행 능력 평가를 지향한다.

이 시험은 지원자의 신청을 통해 평가가 이루어진다. 컴퓨터를 통해 진행되는 반(半) 직접 평가 방식을 활용하고 있어서, 응시자 개개인이 질문에 대한 대답을 녹음하고 그 자료를 미국의 평가 서버에 전송하면 채점이 이루어져 결과가 통보된다.

OPic은 4개 영역(Global Tasks/Functions, Context/Content, Accuracy/Comprehensibility, Text Type)에 걸쳐 언어능력을 총체적으로 측정하며, 시험은 총 40분이 소요된다.

시험에 앞서 20분 정도의 오리엔테이션을 실시한다. 오리엔테이션은 배경 설문, 자기 평가, 사전 점검, 답변 연습의 네 가지로 이루어진다. 배경 설문은 직업, 경험, 관심 분야, 선호도 조사 등을 조사하기 위한 것이며, 이를 바탕으로 개인 맞춤형 문항을 출제한다. 자기 평가는 시험의 난이도 결정을 위한 것이며, 사전 점검은 질문의 청취 및 답변 녹음 기능을 점검하기 위한 것이다. 이어서 청취 및 답변 방법을 안내받은 후 답변 연습을 수행한다. 문항은 직업, 여가 생활, 취미, 관심사, 스포츠, 여행 등에 대한 주제로 이루어지며, 개인별 맞춤 차등 문항으로 총 12~15개의 문항이 출제된다.

시험 시간 40분 동안 총 두 번의 세션으로 나누어 진행된다. 첫 번째 세션에서는 개인 맞춤형 문항 약 7개가 진행되며, 질문은 총 2번 들을 수 있고, 답변 시간에는 제한이 없다. 세션이 끝난 후에는 난이도를 재조정하는 단계를 거친다. 다시 한번 자기 평가가 진행되며, 이후에 출제될 문항을 쉬운 질문, 비슷한 질문, 어려운 질문 중 선택하도록 하고 있다. 이와 같이 난이도 재조정 절차가 끝나면 첫 번째 세션과 동일한 구조의 두 번째 세션이 진행된다. 문항에 응답 시 평가자가 보게 되는 화면의 예시를 보이면 다음과 같다.



[그림 II-5] OPic의 문항 응답 화면(예시)

[그림 II-5]에서 볼 수 있듯, 질문 듣기 버튼(②), 문제 이동 버튼(④)을 통해 문항에 응답하고 다음 문항으로 넘어갈 수 있으며, 시험의 진행도(①)뿐만 아니라 답변 녹음 시 녹음이 잘 되고 있는지를 녹음 게이지(③)로 쉽게 확인할 수 있다. 그리고 질문 청취 시 그림 속 감독관의 입이 움직이며 직접 질문을 들려 주는 것과 같은 효과를 주고 있어, 전반적으로 실제적이고 직관적인 시험 환경을 구축하고자 노력한 것으로 보인다.

평가 방식은 절대 평가 체제이며, ACTFL 공인 평가자가 말하기 능숙도(Oral Proficiency)에 대한 ACTFL의 공식 언어능력 기준인 ‘ACTFL Proficiency Guidelines Speaking (Revised 2012)’ 말하기 기준에 따라 평가한다.

다. 한국어능력시험(TOPIK) 말하기 평가

한국어능력시험(TOPIK)의 말하기 평가 시험은 ‘TOPIK 말하기’로 따로 설정되어 있다. 이 시험은 한국어능력시험(TOPIK)과 별개의 시험으로, 따로 응시료를 내고 응시해야만 한다. 2022년 11월 19일에 처음으로 제1회 시험이 치러질 예정이며, 2023년 6월에 제2회, 8월에 제3회 시험이 실시될 예정이다.

<표 II -30> TOPIK 말하기의 문항 유형

문항	문항 유형	준비 시간	응답 시간
1	질문에 대답하기	20초	30초
2	그림 보고 역할 수행하기	30초	40초
3	그림 보고 이야기하기	40초	60초
4	대화 완성하기	40초	60초
5	자료 해석하기	70초	80초
6	의견 제시하기	70초	80초

각 문항 유형에 대하여 공부 방법과 예시 문항, 모범 답안을 공개하고 있는데, 이때 ‘질문에 대답하기’, ‘그림 보고 역할 수행하기’는 초급 수준의 문제, ‘그림 보고 이야기하기’, ‘대화 완성하기’는 중급 수준의 문제, ‘자료 해석하기’, ‘의견 제시하기’는 고급 수준의 문제임을 밝히고 있다. 이를 통해, 문항을 해결하는 데 요구되는 능력의 수준을 구분하고, 각 능력의 수준당 2개씩의 문항을 설정하되, 능력의 수준이 낮은 것, 즉 쉬운 문항이 앞에 오도록 문항을 배열하였음을 추측해 볼 수 있다. 이 중 ‘의견 제시하기’ 유형의 예시 문항 및 모범 답안을 보이면 다음과 같다. ‘의견 제시하기’ 문항은 ‘전문 분야나 추상적인 내용, 사회 문제 등에 대해 자신의 견해를 논리적으로 제시하거나 찬성 또는 반대 입장에서 자신의 견해를 제시하는 고급 수준의 문제’로 소개되고 있다.

<표 II -31> TOPIK 말하기의 예시 문항 및 모범 답안

예시 문항 및 모범 답안	
* 예시 문항	잘못에 대한 사과는 인간관계에서 발생한 갈등을 해결하는 중요한 실마리가 됩니다. 사과란 무엇입니까? 그리고 인간관계에서 사과는 왜 중요하고, 사과를 하는 올바른 태도는 무엇인지 자신의 생각을 말하십시오.
* 모범 답안	사과란 인간관계에서 어떠한 문제가 발생했을 때 잘못을 솔직하게 인정하고 그 잘못으로 인해 상처를 입거나 피해를 당한 사람에게 용서를 구하는 것을 의미합니다. 인간관계에서 제대로 사과를 하는 것이 중요한 이유는 잘못을 한 사람과 상처를 입은 사람 모두에게 관계를 회복할 수 있는 기회를 주기 때문입니다. 잘못을 한 사람은 자신이 한 실수를 바로잡고 상대방에게 용서를 받을 수 있는 기회를 얻을 수 있습니다. 그리고 상처를 입은 사람은 사과를 받음으로써 마음의 상처를 치유하고 위로를 받는 계기가 될 수 있습니다. 사과를 할 때 순간의 위기를 모면하기 위해 보여 주기 식의 성의 없는 사과를 하거나 자신의 잘못에 대한 변명으로 일관하는 것은 좋은 태도가 아닙니다. 자신이 무엇을 잘못했는지 명확히 밝히고 상대방이 인정할 수 있는 진실한 태도로 사과의 마음을 전달하는 것이 제대로 된 사과입니다. 인간관계의 회복은 이러한 진심 어린 사과에서 시작됩니다.

평가 요소 및 평가 내용은 다음과 같으며, 평가 결과는 문항반응이론을 적용하여 개발된 척도 점수로 산출된다. 척도 점수는 0점부터 200점까지 범위를 가진다.

<표 II -32> TOPIK 말하기의 평가 요소 및 내용

평가 요소	내용
내용 및 과제 수행	과제에 적절한 내용으로 표현하였는가? 주어진 과제를 풍부하고 충실하게 수행하였는가? 담화 구성이 조직적으로 잘 이루어졌는가?
언어 사용	담화 상황에 적합한 언어를 사용하였는가? 어휘와 표현을 다양하고 풍부하게 사용하였는가? 어휘와 표현을 정확하게 구사하였는가?
발화 전달력	발음과 억양이 어느 정도 이해 가능한가? 발화 속도가 자연스러운가?

4) 국내외 디지털 리터러시 실태 조사 사례

(1) PISA(Programme for International Student Assessment, 국제학업성취도평가)

이 조사는 경제협력개발기구(Organization for Economic Co-operation and Development, 이하 OECD)가 주관하는 교육 체제의 질을 점검하기 위한 평가이다. 만 15세 학생들의 읽기, 수학, 과학 영역에 대한 소양(literacy) 수준을 파악하고, 소양 수준에 영향을 주는 배경 변인과의 연계 분석을 통해 각국 교육 정책 수립의 기초 자료를 제공하는 것을 목적으로 한다. 이때 ‘소양(literacy)’이란 지식을 상황과 목적에 맞게 활용하는 능력을 의미한다. PISA는 2000년에 시작되었으며, 3년을 주기로 하여 실시된다. 첫 주기에는 32개국으로 시작하였으나, 주기를 거듭하며 참여하는 국가의 수가 늘어나 2018년에는 79개국이 참여하였다. 매 주기마다 핵심 영역인 읽기, 수학, 과학 중 하나를 주 영역으로 설정하여 평가하며, 이는 읽기, 수학, 과학 순으로 순환된다.

대부분의 나라에서 의무교육이 종료되는 시점인 만 15세 학생을 대상으로 평가가 이루어지며, 비례 층화 표집 방법으로 표본을 추출한다. 각 영역의 문항이 서로 다르게 배합된 검사지를 활용하여 평가하는데, 주 영역과 보조 영역의 문항 수 비중이 다르다. 시험은 선택형, 단답형, 폐쇄형 서술형, 개방형 서술형 등 다양한 유형으로 구성된다.

PISA는 2015 시행부터 컴퓨터 기반의 평가 방식으로 운용되고 있으며, 이에 따라 컴퓨터를 통해 채점이 가능한 선다형 문항과 구성형(C, Computer scored) 문항은 자동 채점 방식으로 채점을 진행한다. 나머지 구성형(H, Human coded) 문항은 국제본부에서 제시한 10개의 채점 설계 중 과학 채점자 8명, 읽기 채점자 6명, 수학 채점자 4명으로 설계된 표준 채점 방식에 따라 전문가가 단수 채점과 복수 채점(double and triple)을 병행하여 진행한다. 학업 성취도 수준을 진단하는 데에서 나아가 학업 성취도에 영향을 미치는 변인을 수집함으로써 교육맥락적 상황을 분석하고 교육 성과의 의미를 부여한다. 장기간에 걸쳐 능력 수준을 진단하고 교육 성과를 제고하기 위해 필요한 정책과 사례를 연구하는 심층 연구들이 후속적으로 이루어진다.

조사 결과 처리에 있어서는 문항반응이론을 활용한다. PISA 점수는 평균 500이고 표준편차 100인 척도 점수를 활용하며, 등급 구분은 우수, 보통, 기초, 기초 미달의 네 등급으로 구분한다. 이와 더불어 PISA 2006부터는 95% 신뢰 수준에서 각 국가의 순위에 대한 범위를 제공하고 있다. 이에 따라 PISA 2009에서는 각 국가별로 평균 점수에 해당하는 등수를 제공하는 대신, 95% 신뢰 수준에서 그 국가가 위치할 수 있는 최고 등수와 최하 등수를 추정하여 제공하고 있다.

본 연구에서 특히 PISA의 여러 영역 중 눈여겨 볼 만한 영역은 ‘읽기’ 영역이다. PISA의 ‘읽기’ 영역은 지속적으로 디지털 리터러시를 포함하는 방향으로 변화하여 왔다. PISA 2000은 읽기 소양 평가 틀 설계에서 “전자 텍스트는 추후 조사에서 사용될 수 있겠으나, 본 조사에서는 시간 및 접근상의 문제로 인해 포함하지 않는다.”라고 언급함으로써 디지털 리터러시 조사의 배경을 논의하였다.

또한, PISA 2009에서는 디지털 세계에 맞는 읽기 소양 측정 도구의 필요성을 인정하고 디지털 읽기 평가(Digital Reading Assessment, DRA)를 새롭게 도입했다. OECD는 DRA 도입의 배경에 대해 인쇄 매체와 디지털 매체 간 읽기의 심리적 과정에 큰 차이가 있고, 인쇄 매체 환경에서의 읽기 능력이 우수하다고 하여 디지털 환경에서의 읽기 능력이 우수할 것이라 단정할 수 없다고 설명하였다. PISA 2009는 평가 틀에 전자적 독해(electronic reading)를 포함했으며, 이는 독자가 텍스트에 접근하기 위해 사용하는 인지 과정 및 텍스트 그 자체에 대한 재정의의 이끌어 냈다. PISA의 DRA 평가 틀 구조는 <표 II-33>과 같다.

<표 II -33> PISA의 DRA 평가 틀(평가 대상 인지 요소)

범주	범주별 세부 항목
접근 및 확인	• 검색하기 • 가상 공간에서 위치를 파악하고 탐색하기(Navigation) • 탐색 도구와 텍스트 구조(메뉴) 활용하기 • 정보를 선택하고 배열하기
통합 및 해석	• 스크롤하며 통합하기 • 해석하기 • 폭넓게 이해하기
성찰 및 평가	• 정보에 대한 사전 평가하기(예 : 메뉴 사용, 웹페이지 훑어보기, 페이지/사이트의 신뢰성과 유용성 점검하기) • 정보의 신뢰성 평가하기 • 내용의 타당성 평가하기 • 응집성과 일관성 평가하기 • 개인적 경험에 비추어 성찰하기

한편, PISA 2015에서는 모든 영역에 컴퓨터 기반 평가(Computer Based Assessment, CBA)를 도입했으며, 시험 운용 방식이 지필 평가에서 컴퓨터 기반 평가로 바뀔에 따라 디지털 텍스트 대신 ‘역동적인 텍스트(dynamic text)’라는 개념을 도입하였다.

이에 나아가, PISA 2018에서는 읽기 소양의 정의 자체에 디지털 리터러시를 반영하고 있다. PISA 2018에서는 읽기 소양을 “자신의 목적을 성취하고 지식과 잠재적 능력을 계발하며, 사회에 참여하기 위해 텍스트를 이해·활용·평가·성찰하고, 다양한 텍스트 읽기 활동에 관여하는 능력(OECD, 2019: 28)”으로 정의하고 있는데, 이처럼 이전 주기에 비해 ‘평가’가 강조된 것은 현대 사회에서의 읽기 환경이 디지털 매체 기반 읽기 능력을 요구한다는 점과 관련이 있다(조성민 외, 2019: 41). 즉, 디지털 매체 환경에서는 늘어나는 정보의 양만큼 독자의 정보 선별 능력이 더욱 중요하다는 점을 고려하여 ‘평가’를 읽기 소양을 구성하는 하위 요소에 도입한 것이다(구자옥 외, 2016: 20). 여기에서 ‘평가’는 읽기와 과제 수행의 목적에 비추어 여러 정보 가운데 보다 필요한 정보가 무엇인지, 어떤 정보가 신뢰할 만한 것인지를 판단하는 읽기 능력(구자옥 외, 2016: 20)을 일컫는다.

또한, 디지털 리터러시와 관련된 요소들이 평가 틀에 새롭게 추가된 점도 눈에 띈다. 조성민 외(2019: 42)에서는 PISA의 주기별 읽기 평가 틀을 정리하고 있어 참고할 만하다. 여기에서는 ‘상황’ 차원은 제외하고 ‘과정’ 차원과 ‘텍스트’ 차원만을 이하 제시한다.

<표 II -34> PISA의 주기별 평가 틀 비교

PISA 2018			PISA 2009/2012/2015		
차원	범주	종류	차원	범주	종류
과정	텍스트	정보 찾기, 이해하기, 평가와 성찰하기	양상	접근과 확인	
	과정	목표 설정과 계획 수립, 점검, 조정		통합과 해석	
	과제 관리	유창하게 읽기		성찰과 평가	
텍스트	출처	단일, 다중	텍스트	체제	연속, 비연속, 혼합 다중
	조직과 탐색	정적, 동적			
	체제	연속, 비연속, 혼합		유형	기술, 서사, 설명, 논증, 지시
	유형	기술, 서사, 설명, 논증, 지시, 상호 작용			

위 <표 II -34>를 보면, ‘텍스트 범주’의 ‘정보 찾기’에서는 ‘텍스트 검색과 선택’, ‘이해하기’에서는 ‘축자적 의미 표상’과 ‘통합과 추론 생성’, ‘평가와 성찰하기’에서는 ‘정보의 질과 신뢰성 평가’ 및 ‘갈등의 발견과 조정’ 등과 같은 요소들이 새로 추가되었는데, 이들은 디지털 매체를 고려한 새로운 형태의 읽기를 반영하고 있다.

또한 ‘텍스트’ 차원에 ‘조직과 탐색’ 범주가 신설되었는데, 이는 조직된 방식에 따라 텍스트를 탐색하는 방법이 다르다는 이유에서였다. 이는 PISA의 읽기 영역에서 다루는 텍스트의 범위가 인쇄 매체뿐 아니라 디지털 매체 환경을 반영한 텍스트까지로 확장되었음을 시사한다. 이때, 인쇄 텍스트, 디지털 텍스트를 막론하고 스크롤링, 찾기 기능 등 단순 상호 작용을 위한 최소한의 도구 모음이 함께 제공되는 경우는 ‘정적’ 텍스트로 분류된다. 반면 상호 작용성이 더 높은 하이퍼링크, 핵심어 색인을 제공하거나 대화형 텍스트 기반 통신 미디어의 특성이 반영된 경우는 ‘동적’ 텍스트로 분류된다(조성민 외, 2019: 42). 즉, 해당 텍스트의 매체와 관계없이 텍스트의 상호 작용 정도를 바탕으로 정적 텍스트와 동적 텍스트를 구분하는 것이다.

한편, PISA에서는 비인지적 영역의 성과를 파악하기 위하여 학생 배경(문화적, 인종적, 사회경제적 배경), 교수 실제, 학교 정책, 학교 체제 등과 같은 제도적 맥락에 대한 정보를 설문 조사를 통해 수집한다는 점이 특징인데, 이러한 설문 문항에서도 디지털 리터러시와 관련한 고려들을 여럿 읽어 낼 수 있다. 우선, 학생들의 학교 밖 읽기 경험을 조사할 때, 읽기 자료를 얼마나 자주 읽는지 물어보는 과정에서 인쇄물과 디지털 기기를 모두 포함하여 묻기도 하고, 온라인 매체 읽기 활동의 경험을 묻기도 하는 등 온라인 읽기를 적극적으로 고려하는 모습을 볼 수 있다. 또한, ICT 친숙도에 대해 조사하는 것 역시 디지털 리터러시와 깊은 관련이 있다. ICT 설문의 내용은 다음의 <표 II -35>를 참고할 수 있다(구자옥 외, 2016: 80).

<표 II -35> PISA에서 진행하는 ICT 설문문의 내용

설문 모듈	내용
읽기 교수의 실제	학교 수업에서 디지털 기기 사용 정도, 학교 수업 이외의 학습에서 디지털 기기 사용 정도, 학교 수업에서 학생이나 교사의 디지털 기기 사용 여부
학교 밖 읽기 학습 경험	학교 밖에서의 인터넷 사용 정도(주중), 학교 밖에서의 인터넷 사용 정도(주말)
학생의 SES 등 가정 배경	집에 구비된 디지털 기기의 사용 여부
학생 성향 및 학교에 대한 태도	디지털 기기를 처음 사용한 시기(나이), 인터넷에 처음 접속한 시기(나이), 학교에서의 인터넷 사용 정도(주중), 학교 밖에서의 디지털 기기 사용 빈도, 다음 활동을 위한 학교 밖에서의 디지털 기기 사용 빈도, 다음 활동을 위한 학교 내에서의 디지털 기기 사용 빈도, 디지털 기기 경험에 대한 견해, 디지털 기기 활용 역량, 디지털 기기 활용에 대한 흥미
학교 맥락과 자원	학교에 구비된 디지털 기기의 사용 여부

(2) PIAAC(Programme for the International Assessment of Adult Competencies, 국제성인역량조사)

국제성인역량조사는 40여 개의 OECD 회원국에서 실시되는 대규모 조사로, 개인이 사회와 경제 부흥에 기여하기 위해 필요한 핵심적인 인지적 능력 및 직업 역량의 양상을 조사하는 데 주 목적을 둔다. 조사는 10년마다 실시되며, 첫 번째 주기의 조사는 2011년~2018년 동안 세 차례의 데이터 수집이 진행되었다. 2018년에는 조사의 두 번째 주기가 시작되었으며 이 주기의 결과는 2024년에 발표될 예정이다.

이 조사에서는 성인들의 문식력, 수리력, 컴퓨터 기반 환경에서의 문제 해결력을 측정함으로써, 일상 생활 및 직업 생활을 영위하기 위해 필요한 역량이 사회에 어떻게 분포하고 있고 어떻게 활용되고 있는지를 파악한다. 이 조사의 결과는 교육 및 훈련 체계가 어떻게 이러한 능력을 계발할 수 있는지에 대한 이해를 제공하고, 교육자, 정책 입안자, 노동경제학자 등의 관계자가 성인의 역량을 강화하기 위한 경제, 교육, 사회 정책을 발전시키는 데 필요한 기초 자료를 수집하는 데 목적을 두며, 각 참여 국가당 16세부터 65세의 일반인 5천 명을 대상으로 실시된다. 이 조사는 직업 역량에 영향을 미치는 다양한 배경변인을 조사함으로써 역량 수준과 해당 변인 간의 관계를 분석하여 정책적 시사점을 마련한다. 배경 설문은 컴퓨터를 이용하여 조사원이 응답자에게 문항을 읽어서 제시한 후, 응답 내용을 조사원이 입력하는 방식으로 진행되며, 배경 설문 완료 후 컴퓨터 평가 또는 서면 평가 중 택일하여 응답자가 문항에 응답하는 방식으로 조사가 수행된다.

PIAAC은 문식력(literacy), 수리력(numeracy), 컴퓨터 기반 문제 해결력(problem solving in Technology-rich Environment)을 핵심 정보 처리 역량으로 보고, 세 영역의 능력을 측정한다. 그중 컴퓨터 기반 문제 해결력은 정보를 획득·평가하고 타인과 의사소통하며 실제적인 과업을 수행하기 위해 디지털 기술, 커뮤니케이션 도구 및 네트워크를 활용하는 능력으로 정의된다. 이 능력의 측정은 컴퓨터와 컴퓨터 네트워크를 통해 적절한 목표와 계획을 수립하고, 정보에 접근 및 이를 활용함으로써 개인적인 업무 및 시민의 목적을 위해 문제를 해결할 수 있는지 여부에 초점을 둔다. 즉, 컴퓨터 기반 문제 해결력은 ‘컴퓨터 언어능력(ICT 도구 및 응용프로그램 사용 능력)’과 문제 해결에 요구되는 인지 기술이 교차되는 부분에서 나타나는 역량의 한 영역으로, ICT 도구 및 응용 프로그램을 사용하여 정보에 효과

적으로 접근하고 적절하게 처리·평가·분석하는 능력을 의미한다. 나아가, 2018년부터 시작되는 2주기 조사에서는 ‘컴퓨터 기반 문제 해결력’ 대신 ‘적응형 문제 해결력(Adaptive Problem Solving)’을 신설하여, 보다 더 심화된 컴퓨터 기반 문제 해결력을 다루고자 하였다. 즉, 컴퓨터 기반 문제 해결의 상황에서 여러 문제를 동시에 해결하는 능력을 평가할 수 있도록, 문제 목록에 접근하는 순서를 관리하고 다양한 문제 세트를 해결하기 위해 발생하는 기회를 모니터링하는 능력을 평가하는 것을 목표로 하고 있다.

또 다른 핵심 정보 처리 역량인 문식력의 경우 문서화된 글을 이해, 평가, 활용하며, 글로써 소통할 수 있는 능력을 의미한다. 문식력은 글로 표현된 단어와 문장을 이해하는 것에서부터 복잡한 문자의 이해, 해석, 평가에 이르기까지의 다양한 기술을 포함하며, PIAAC에서는 종이에 인쇄된 텍스트 외에 디지털 텍스트 독해를 평가에 포함함으로써 응답자의 능동적 작업 수행을 살펴보고자 한다.

<표 II-36>은 PIAAC의 문식력, 적응형 문제 해결력의 평가 틀로써(OECD, 2021: 19), 이 평가 틀은 PISA DRA 및 PISA의 2018의 읽기 평가 틀과 일부 유사한 면이 있다.

<표 II-36> PIAAC의 문식력, 적응형 문제 해결력의 평가 틀(OECD, 2021: 19)

	문식력 (literacy)	적응형 문제 해결력 (Adaptive Problem Solving, APS)
정의	문식력은 자신의 목표를 달성하고 지식과 잠재력을 개발하며 사회에 참여하기 위해 문어 텍스트에 접근·이해·평가·성찰하는 것이다.	적응형 문제 해결력은 해결 수단을 즉시 사용할 수 없는 역동적인 상황에서 목표를 달성하는 능력을 포함한다. 다양한 정보 환경과 맥락에서 문제를 정의하고, 정보를 검색하며, 해결책을 적용하기 위한 인지 및 메타인지 과정을 의미한다.
인지 전략	• 텍스트에 접근하기 • 이해하기 • 평가하기	• 정의 • 검색 • 적용
내용	다음과 같은 특징이 있는 텍스트: • 유형(기술, 서사, 설명, 논증, 지시, 거래) • 형식(연속, 비연속, 혼합) • 조직(정보의 양과 내용 표현 및 접근 장치의 밀도) • 출처(단일 및 다중 텍스트)	작업의 차원 • 문제의 구성 • 상황의 역동성 • 환경의 특징 • 정보의 환경
맥락	• 일과 직업 • 개인 • 사회 및 시민	• 개인 • 일 • 사회/공동체

<표 II-36>의 평가 틀과 OECD(2021: 167-169)에 따르면, 적응형 문제 해결력의 평가 내용에는 네 가지의 작업의 차원이 포함된다. ‘문제의 구성(problem configuration)’은 초기 문제 설정 및 목표 상태와 관련된다. 이들은 문제의 요소, 관계 및 자원(resource)/운영자(operator)를 포함한다. 문제에는 더 많거나 더 적은 요소가 있을 수 있으며 이러한 요소는 서로 상호 작용하거나 상대적으로 독립적일 수 있다. 또한, 문제의 요구 사항에는 단 하나 또는 여러 목표의 달성이 포함될 수 있다. 이와 같이 적응형 문제 해결력에서는 문제 구성의 이러한 모든 특성을 반영하고자 하였다. ‘상황의 역동성’은 문제

상황 내의 변경(또는 변경 없음)과 시간에 따른 문제 제약, 그리고 이것이 문제 구성에 미치는 영향을 나타낸다. 문제 상황에 변경이 발생하면, 어려움과 교착 상태를 유발할 수 있다. 따라서 ‘문제의 역동성’ 또한 적응형 문제 해결력을 평가하는 내용이 될 수 있다. ‘환경의 특징’은 환경의 다양한 특징(feature)과 환경에서 사용할 수 있는 정보 및 자원(resource)을 의미한다. 문제가 설정되고 풀리는 환경이 정보가 풍부한지, 그리고 해당 정보가 문제 해결과 다소의 관련이 있는지, 환경이 구조화되어 있는지 등이 곧 환경의 특징과 관련되며, 이들은 적응형 문제 해결력에 영향을 준다. 정보의 환경은 문제를 해결하는데 사용할 수 있는 자원(resource)의 출처와 관련이 있다. 정보 환경의 특성은 물리적, 사회적 또는 디지털적일 수 있다. 물론 이러한 모든 자원은 디지털 문제 해결 노력에서 다소 동시에 나타나지만, 문제 해결을 위해서는 특정 종류의 자원을 적어도 정신적으로 처리해야 할 필요가 있다. 이에 적응형 문제 해결력에서는 물리적 자원, 사회적 자원, 디지털적 자원을 모두 고려한다. 이러한 적응형 문제 해결력 평가를 위하여 PIAAC에서는 일상 및 직장 생활 문제를 설명하는 시나리오 기반 과제를 제시한다. 예를 들어, “저녁 식사 준비” 과제에서는 달성해야 할 목표로 “1) 오전 8시 30분까지 아이를 학교에 데려다 주고, 2) 상점에서 저녁 식사 재료를 산 후, 3) 오전 10시까지 집에 도착할 수 있는 가장 빠른 방법을 찾는 것”을 제시하고 있다. 이러한 목표와 조건은 메모로 표시되며, 지도와 함께 상단에 시계가 표시된다. 지도상의 위치를 클릭하면 운전 시간에 대한 정보를 보여 준다. 그러나 실생활에서와 마찬가지로 길을 가다가 갑자기 상점 중 하나가 문을 닫았고 다른 계획을 세워야 하는 상황에 놓이도록 하고, 응시자는 이를 해결해야만 한다. 위 문항에서 적응형 문제 해결력의 평가 내용은 다음 <표 II-37>과 같다(OECD, 2021: 169).

<표 II-37> 적응형 문제 해결력의 평가 내용(예시)

예시 문항	평가 내용
응시자에게 동시에 두 가지 목표를 달성하도록 요청하고, 문제 요소는 접근 가능하고 눈에 띄며 시각적으로 정렬된 방식으로 제시됨.	문제의 구성
정보 환경은 풍부하지 않음. 문제 자체를 넘어서는 정보가 많이 제공되지 않음.	정보의 환경
상황의 역동성은 평균. 변경이 유도되면 응시자에게 변경 사항에 대한 프롬프트가 표시되고 세부 사항이 설명됨.	상황의 역동성
문제는 개인적 문제 맥락에 배치되며, 디지털 정보 환경과 물리적 정보 환경이 혼합된 정보 환경에 배치됨.	문제의 맥락, 정보의 환경

3. 디지털 리터러시 역량 실태 조사 방안 수립

1) 국어 사용 환경 및 현황 분석

최근 5년간 국민의 국어능력과 관련된 국어 사용 환경 및 현황을 분석한 결과는 다음과 같다. 첫째, 전 국민이 하루 중 미디어 이용에 보내는 시간의 비중이 매우 크다. 통계청 조사에 따르면, 2019년 현재 우리나라 10세 이상 국민은 여가 시간으로 4시간 47분을 사용하고 있으며 이는 2014년과 비교해 2분 감소한 것이다. 여가 시간 중 미디어 이용에 가장 많은 2시간 26분을 사용하고 있으며 이는 2014년과 비교해 5분 증가한 것이다. 즉, 우리나라 국민의 여가 시간은 다소 줄고 있으나 여가 시간 중 미디어 이용에 보내는 시간은 오히려 늘고 있다.

의무 시간(일, 학습) 중에도 각종 미디어가 이용된다는 점을 감안해 볼 때, 그리고 특히 최근 COVID-19의 팬데믹 상황에서 재택근무와 원격 수업이 일반화된 점을 감안해 볼 때, 전 국민의 미디어 이용 시간이 최대 6시간 33분(즉, 미디어 이용(2:26)+일(3:13)+학습(0:54)=6:33)에 이른다는 추정도 가능하다. 즉, 국민은 하루 시간의 1/4가량을 미디어를 사용하며 살고 있는 것이다.

<표 II -38> 전 국민(10세 이상)의 하루 시간 사용(통계청, 2020: 8)

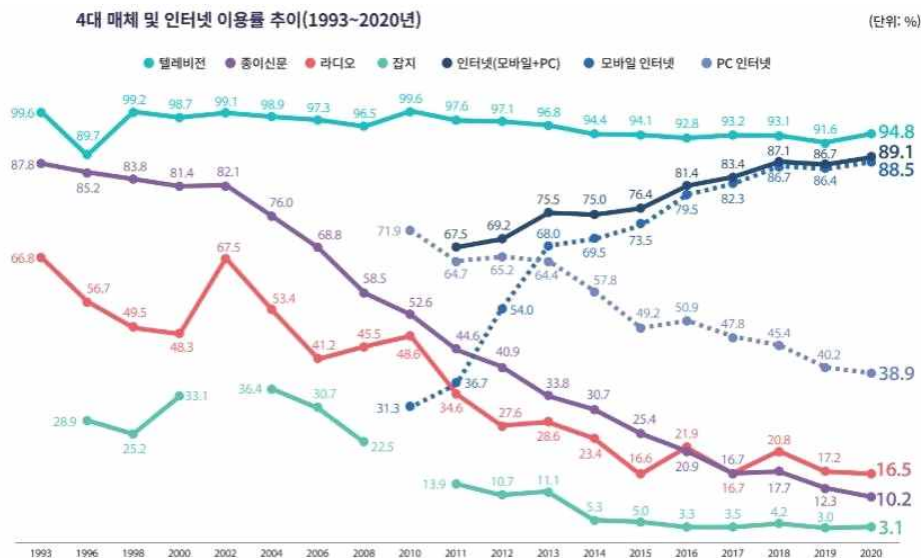
		단위(시간:분)		
		2014년	2019년	차이
☐	필수 시간	11:14	11:34	0:20
	수면	7:59	8:12	0:13
	식사 및 간식	1:56	1:55	-0:01
	기타 개인 유지 ¹	1:18	1:27	0:09
☐	의무 시간	7:57	7:38	-0:19
	일(구직 활동 포함)	3:16	3:13	-0:03
	학습	1:07	0:54	-0:13
	가사 노동 ²	1:55	1:56	0:01
	이동	1:39	1:36	-0:03
☐	여가 시간	4:49	4:47	-0:02
	교제 및 참여	1:13	0:59	-0:14
	문화 및 관광	0:04	0:04	0:00
	미디어 이용	2:21	2:26	0:05
	스포츠 및 레포츠	0:30	0:30	0:00
	기타 ³	0:42	0:49	0:07

주) 1. 개인 건강 관리, 개인 위생 및 외모 관리; 2. 가정 관리, 가족 및 가구원 돌보기; 3. 게임 및 놀이, 개인 취미 활동, 자원봉사 등

지난 30여 년간 국민이 이용하는 미디어의 종류도 매우 달라졌는데, 매스 미디어, 특히 인쇄 기반 미디어의 이용률은 급격히 줄어든 반면 인터넷 기반 미디어의 이용률은 급격히 늘어났다. 매스 미디어를 대표하는 텔레비전, 종이신문, 라디오, 잡지 등 이른바 4대 매체에 대한 지난 1주일간의 이용률 추이를 살펴보면 텔레비전을 제외한 종이신문, 라디오, 잡지의 이용률은 하락세가 지속되고 있다. 2020년 현재, 종이신문은 1993년 87.8% 대비 77.6%p 하락한 10.2%의 이용률을 보인다. 잡지는 최고 이용률을 보인 2004년 36.4% 대비, 33.1%p 하락한 3.1%의 이용률을 보이며, 라디오는 최고 이용률을 보인

2002년 67.5%와 대비하여 51.0%p 하락한 16.5%의 이용률을 보인다. 반면, 텔레비전은 1993년 이래 꾸준히 90% 이상의 이용률을 보여 왔으며 2020년 이용률은 94.8%에 달한다.

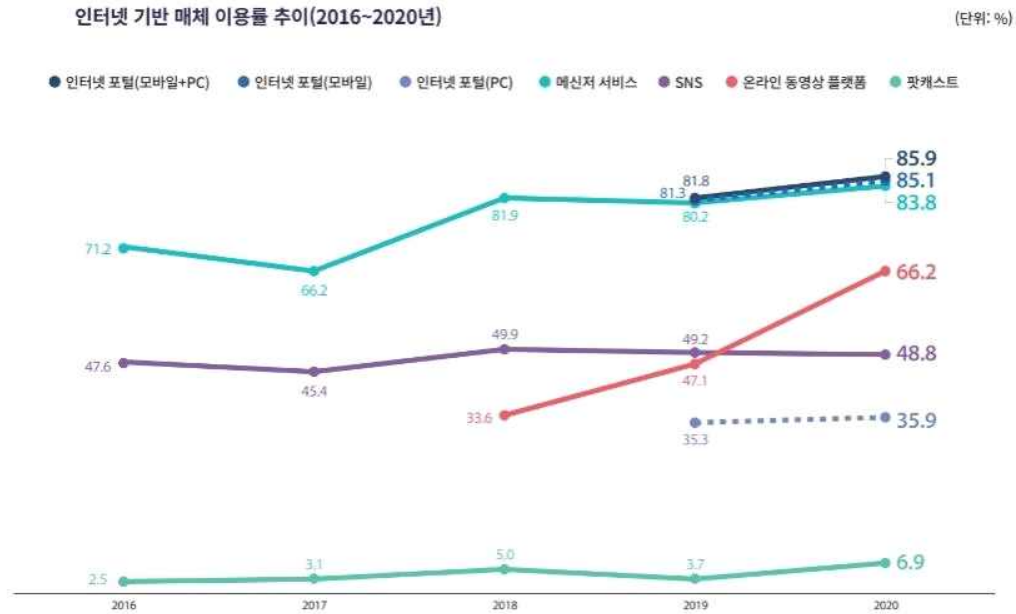
인터넷 기반 미디어 이용률을 보면 2020년 현재 전체 인터넷(모바일+PC)은 2011년 67.5% 대비 21.6%p 상승한 89.1%의 이용률을 보인다. 이는 PC 인터넷이 2010년 71.9% 대비 33.0% 하락한 38.9%의 이용률에 그쳤음에도 불구하고, 모바일 인터넷이 2010년 31.3% 대비 57.2% 상승한 88.5%의 이용률을 보였기 때문이다.



[그림 II -6] 1993~2020년 4대 매체 및 인터넷 이용률 추이(한국언론진흥재단, 2020: 57)

둘째, 인터넷 기반 미디어 중 인터넷 포털, 메신저 서비스, 온라인 동영상 플랫폼, SNS의 이용률이 특히 높은 것으로 나타났다. 2020년 현재, 인터넷 포털(모바일+PC)은 85.9%의 이용률로 여러 인터넷 기반 미디어 중 가장 높은 이용률을 보이고 있으며, 메신저 서비스가 83.8%의 이용률로 비슷하게 높은 이용률을 보이고 있다.

또한 2020년에 온라인 동영상 플랫폼 이용률의 가파른 상승이 두드러지는데, 처음 조사된 2018년 33.6% 대비 약 2배 정도 상승한 66.2%의 이용률을 보인다. 아울러 SNS 이용률은 꾸준히 일정 수준을 유지하고 있는데, 2016년 47.6%의 이용률을 보인 이래 꾸준히 45% 이상의 이용률을 보여 왔으며, 2020년을 기준으로 48.8%의 이용률을 보인다.



[그림 II-7] 2016~2020년 인터넷 기반 미디어 이용률 추이(한국언론진흥재단, 2020: 58)

셋째, 국민의 종이책 독서율은 감소한 반면, 전자책 독서율은 증가하고 있다. 2019년에 성인의 종이책 연간 독서율은 2017년 59.9% 대비 7.8%p 감소한 52.1%이고, 초·중·고교 학생의 종이책 연간 독서율은 2017년 91.7% 대비 1.0%p 감소한 90.7%로 나타났다. 즉, 국민의 종이책 독서율은 감소 추세에 있으며 이는 특히 청소년보다 성인에게서 더 분명하게 나타났다.

한편 2019년을 기준으로 성인의 전자책 연간 독서율은 2017년 14.1% 대비 2.4%p 증가한 16.5%이고, 초·중·고교 학생의 전자책 연간 독서율은 2017년 25.3% 대비 7.4%p 증가한 32.7%로 나타났다. 즉, 아직까지 종이책 독서율에 한참 미치지 못하지만 국민의 전자책 독서율은 증가 추세에 있으며 특히 기성세대인 성인보다 미래 세대인 초·중·고교 학생에게서 더 빠른 증가 추세를 보이고 있다.

2019년에 처음 조사한 오디오북 연간 독서율은 성인이 3.5%, 초·중·고교 학생이 18.7%로 나타났다. 증감 추세를 판단할 수 있는 과거의 조사 자료가 없으나, 기성세대인 성인보다 미래 세대인 초·중·고교 학생에게서 그 독서율이 훨씬 높게 나타났다는 점을 감안해 볼 때, 오디오북 또한 전자책과 마찬가지로 향후 그 독서율이 증가할 것이라 예상된다.

<표 II-39> 성인 및 초·중·고교 학생의 연간 독서율 추이(백원근 외, 2020: 10-11)

	성인			초·중·고교 학생		
	2017년	2019년	증감	2017년	2019년	증감
종이책	59.9%	52.1%	-7.8%	91.7%	90.7%	-1.0%
전자책	14.1%	16.5%	2.4%	25.3%	32.7%	7.4%
오디오북	-	3.5%	-	-	18.7%	-

주) 연간 독서율: 지난 1년간 일반 도서(교과서, 학습 참고서, 수험서, 잡지, 만화 제외)를 1권 이상 읽은 사람의 비율

2) 디지털 리터러시의 개념 및 조사 방향

(1) 디지털 리터러시의 개념

이상의 변화된 국어 사용 환경에 대한 분석 내용을 고려하면, 디지털 매체의 등장 및 발전으로 인해 변화된 국어 사용 환경에서는 국어능력을 새로운 방식으로 정의할 필요가 있다. 전통적으로 국어능력은 그 언어 양식이 구어인지 문어인지, 그리고 의사소통 활동이 표현 활동인지 이해 활동인지에 따라 듣기(구어·이해), 말하기(구어·표현), 읽기(문어·이해), 쓰기(문어·표현)의 넷으로 구분하여 네 가지 능력의 총합으로서 정의되어 왔다.

디지털 매체를 중심으로 변화된 국어 사용 환경에서는 국어능력을 넷으로 나누어 접근하는 방식이 더 이상 유효하지 않게 된다. 가령, 인터넷 문서 읽기 상황에서 독자는 문자 텍스트뿐만 아니라 이와 함께 제시된 다양한 시청각 보조 자료를 함께 접하게 되므로 전통적인 읽기만이 아니라 듣기는 물론 새로운 형태의 의사소통인 보기까지를 수행해야 한다. 프레젠테이션 상황에서 발표자는 구두 언어를 통한 발표만 수행하는 것이 아니라 시각 자료 준비 과정에서 문자 텍스트 쓰기를 함께 수행해야 한다. 이처럼 디지털 매체를 통한 국어 사용 환경에서는 듣기, 말하기, 읽기, 쓰기가 늘 혼용되는 경향이 있으므로 과거와 같은 분절적인 방식으로 국어능력을 정의하는 것이 점점 더 무의미해지고 있다.

이와 관련하여 최근 국어교육 분야에서는 디지털 매체를 중심으로 변화된 국어 사용 환경에서 필요한 국어능력을 ‘미디어 리터러시’나 ‘디지털 리터러시’라는 용어를 사용하여 새롭게 정의하고자 시도하여 왔다. 미디어 리터러시란, “미디어를 책임 있게 활용하고, 미디어가 생산하는 정보와 문화를 비판적으로 이해·성찰하며, 의미 있는 정보와 문화를 생산·공유할 수 있는 능력”으로 정의된다(정현선 외, 2016: 211). 미디어 리터러시와 유사 개념인 디지털 리터러시는 “한 개인이 자신의 목적을 실현하기 위해 디지털 도구와 기술을 활용하여 텍스트를 탐색·이해·평가·적용하고, 새로운 텍스트를 창조하며, 사회 구성원들과 원활하게 소통할 수 있는 능력”이라고 정의된 바 있다(김종윤 외, 2017: 12).

미디어 리터러시와 디지털 리터러시의 관계에 대해서는 1) 미디어 리터러시가 디지털 리터러시를 포함하는 상위 개념이라고 보는 경우, 2) 디지털 리터러시가 미디어 리터러시를 포함하는 상위 개념이라고 보는 경우, 3) 미디어 리터러시와 디지털 리터러시가 대등하게 구별되는 가운데 서로 교집합을 지닌다고 보는 경우로 다양한 논의가 있어 왔으나(노들, 옥현진, 2020: 106-110; 박종임, 2021: 11-14; 정현선 외, 2016: 214-216), 결정적인 차이는 미디어 리터러시가 대중 매체로 대표되는 ‘올드 미디어’의 수용 및 생산 능력, 특히 대중 매체에 대한 비판적 이해를 강조하던 것에서부터 비롯된 좀 더 전통적인 개념인 데 비해 디지털 리터러시는 최근 등장한 새로운 미디어의 수용 및 생산 능력에 초점을 맞춘 개념이라는 점이다(정현선 외, 2016: 214-216). 변화된 국어 사용 환경에서 필요한 국어능력을 미디어 리터러시라고 정의하든 디지털 리터러시라고 정의하든 그 결과는 크게 다르지 않을 수 있겠으나, 미디어 리터러시 논의로부터 시작해 디지털 리터러시에 대한 관심이 확장(옥현진, 오은하, 김종윤, 2018)되고 있으며 최근의 평가 연구에서 전반적으로 ‘디지털 리터러시’라는 용어를 사용하고 있다는 점에서 본 연구에서 또한 ‘디지털 리터러시’라는 용어를 사용하여 접근하고자 한다.

보다 구체적으로 디지털 리터러시 중 특히 ‘비판적 리터러시’, ‘복합양식 리터러시’, ‘기능적 리터러시’의 세 가지가 새로운 매체 환경에서 필요한 국어능력이라고 간주해 볼 수 있다. 이와 관련하여 비판적 리터러시, 비주얼 리터러시, 복합양식 리터러시, 문화적 리터러시, 시민적 리터러시, 기능적 리터러시 등 다양한 하위 요소들이 미디어 리터러시의 구성개념이 된다는 논의가 제시되어 왔다(예: 김아미, 2015; 정현선, 2014, 정현선 외, 2016; 황연주, 2001; Hobbs, 2010). 이 가운데 비판적 리터러시, 복합양식

리터러시, 기능적 리터러시의 세 가지가 2015 개정 국어과 교육과정 성취기준에 반영되어 있음을 확인할 수 있다(정현선 외, 2016: 221-222). 이렇게 볼 때, 국어능력으로서의 디지털 리터러시를 비판적 리터러시, 복합양식 리터러시, 기능적 리터러시로서 정의하는 것이 가능하다.

<표 II-40> 디지털 리터러시의 하위 구성 개념과 그 정의(정현선 외, 2016: 217-218)

디지털 리터러시의 하위 구성 개념	정의
비판적 리터러시	정보 생산의 주체와 의도를 비판적으로 파악하고, 정보의 표현 방식과 생산 방식에 따른 의미 효과를 비판적으로 이해할 수 있는 능력
비주얼 리터러시	미디어의 다양한 시각적 이미지의 의미 작용을 문화·예술의 차원에서 이해하는 능력
복합양식 리터러시	미디어의 의미를 구성하는 다양한 기호와 언어의 작용을 이해하고 활용할 수 있는 능력
문화적 리터러시	사회적·문화적·언어적 다양성 증대로 인한 리터러시의 변화를 이해하고, 사회적·문화적·언어적으로 다양한 맥락에서 리터러시를 실행하는 능력
시민적 리터러시	사회적 책임감을 가지고 리터러시를 실천하고, 적절한 정보를 갖추고 능동적으로 사회에 참여하는 능력
기능적 리터러시	기술 변화로 인한 멀티미디어 텍스트의 속성을 이해하고, 다양한 텍스트와 기술을 사용할 수 있는 능력

<표 II-41> 디지털 리터러시의 하위 구성 개념과 관련된 성취기준(정현선 외, 2016: 221-222)

디지털 리터러시의 하위 구성 개념	2015 개정 국어과 교육과정 중 관련 성취기준
비판적 리터러시	[9국02-07] 매체에 드러난 다양한 표현 방법과 의도를 평가하며 읽는다.
복합양식 리터러시	[6국02-05] 매체에 따른 다양한 읽기 방법을 이해하고 적절하게 적용하며 읽는다. [9국03-08] 영상이나 인터넷 등의 매체 특성을 고려하여 생각, 느낌, 경험을 표현한다.
기능적 리터러시	[6국01-05] 매체 자료를 활용하여 내용을 효과적으로 발표한다. [6국03-02] 목적이나 주제에 따라 알맞은 내용과 매체를 선정하여 글을 쓴다. [9국01-11] 매체 자료의 효과를 판단하며 듣는다. [9국02-08] 도서관이나 인터넷에서 관련 자료를 찾아 참고하면서 한 편의 글을 읽는다. [9국03-01] 쓰기는 주제, 목적, 독자, 매체 등을 고려한 문제 해결 과정임을 이해하고 글을 쓴다.

한편, 디지털 리터러시의 개념을 전통적인 국어능력의 외연으로만 한정하지 않고 더욱 적극적으로 외연을 확장하여 접근할 수도 있다. 디지털 리터러시를 2015 개정 교육과정 중 국어과의 교과 역량이 아니라 총론의 핵심 역량과 관련시킨다면, 의사소통 역량 및 지식정보처리 역량과 밀접한 관련이 있다고 간주할 수 있다(정현선 외, 2016: 231). 나아가 디지털 리터러시의 하위에 다음 표와 같이 '의미 이해와

전달’, ‘책임 있는 미디어 이용’, ‘감상과 향유’, ‘미디어 기술 활용’, ‘정보 검색과 선택’, ‘창작과 제작’, ‘사회·문화적 이해’, ‘비판적 분석과 평가’의 8가지 수행 목표를 설정할 수 있다(정현선 외, 2016: 232-233). 실제로 많은 논의들이 디지털 리터러시 혹은 미디어 교육의 체계를 이처럼 국어과 교육이나 교과 역량의 틀에 얽매이지 않고 제시하고 있는 상황이다(예: 강진숙 외, 2019; 김대희, 2007; 김양은, 2007; 김현진 외, 2020; 변숙자, 2021).

<표 II-42> 미디어 교육의 내용 영역 비교 및 종합(장은주, 2020: 47)

김양은(2007)	김대희(2007)	정현선 외(2016)	강진숙 외(2019)	김현진 외(2020)	
•미디어 지식	•미디어에 대한 지식	•사회·문화적 이해	•지식	-	⇒ 지식
•미디어 감상	•미디어를 통한 문화	•감상과 향유	-	-	⇒ 감상(하기)
•미디어 분석	•미디어 언어의 사용	•비판적 분석과 평가	•비평	•이해와 비평	⇒ 분석(하기)
•미디어 생산		•의미 이해와 전달 •창작과 제작	•의사소통 •구성/제작	•표현과 생산	⇒ 표현(하기)
-	-	•정보 검색과 선택 •미디어 기술 활용	•접근/활용	•접근과 활용	⇒ 활용(하기)
-	•미디어에 대한 태도	•책임 있는 미디어 이용	•참여	•윤리와 보안 •소통과 참여	⇒ 태도
-		-	-	•웰빙과 문화	

이렇듯 사회가 급격히 변하고 기술 발전이 빠르게 진행되면서, 한 사회 구성원으로서 갖추어야 할 리터러시(literacy)는 점차 다양해지고 그 수준도 높아지고 있다. 초기의 리터러시가 3Rs(Reading, Writing, Arithmetic)에서 출발하여 컴퓨터 리터러시(computer literacy), 정보 리터러시(information literacy) 등을 거쳐 현재는 디지털 리터러시(digital literacy)에 이르고 있는 것에서도 그 중요성을 알 수 있다. 주지하듯이 디지털 리터러시는 Gilster(1997)의 연구에서 출발한 용어로, ‘컴퓨터를 통해 다양한 출처로부터 찾아낸 여러 가지 형태의 정보를 이해하고 자신의 목적에 맞는 새로운 정보를 조합해냄으로써 올바르게 사용하는 능력’을 뜻한다. 즉 디지털 리터러시는 단순히 디지털 기술을 잘 다루고 정보를 적절히 처리하는 능력 이상의 것이 필요하다는 인식에서 출발한다. 이와 관련하여 노은희 외(2018)에서는 ‘디지털 환경에서 학습자가 주도적이고 가치로운 삶을 살아가기 위해 디지털 기술을 올바르게 이해·사용하여, 정보 및 그 내용물을 적절하게 탐색·활용하고, 비판적으로 분석·평가하며, 생산적으로 소통·창조하는 복합적인 역량’이라고 정의하기도 하였다.

디지털 리터러시의 주요한 특성으로는 범주상의 다차원성과 수준상의 연속성을 꼽을 수 있는데(이애화, 2015: 84), 이를 고려하여 디지털 리터러시의 개념을 4영역으로 제시하고 각 범주와 그 실천 역량을 함께 진술하면 다음과 같다(노은희, 2018: 23).

- 디지털 기술에 대한 올바른 이해와 사용
- 디지털 기술을 활용한 정보와 그 내용물의 적절한 탐색 및 활용
- 디지털 기술을 통해 파악한 정보와 그 내용물의 비판적 분석 및 평가
- 디지털 기술을 통한 정보와 그 내용물의 생산적인 소통 및 창조

이와 관련하여 지난 2018 국민의 국어능력 조사에서는 제시한 내용 중 2번째와 3번째 요소에 해당하는 ‘디지털 정보를 검색·평가·통합하는 능력’이 디지털 텍스트 읽기 능력과 밀접한 관련이 있다고 보았다. 이에 디지털 리터러시를 측정하기 위해 디지털 영역을 설정하고 디지털 문항 2개를 개발하여 평가에 활용한 바 있다. 본 연구에서는 읽기 영역뿐 아니라 듣기·말하기 영역 역시 디지털 리터러시 역량과 밀접한 관련이 있다는 판단 아래, 듣기·말하기 역량을 타당하게 조사하기 위해서는 듣기·말하기와 관련하여서도 디지털 리터러시를 조사하기 위한 문항의 개발이 요구된다고 본다. 또한 디지털 리터러시 영역의 중요성을 고려하였을 때 추후 영역별, 연차별 국어능력 실태 조사로 전환 시 디지털 리터러시 영역을 별도로 상정하는 안을 적극 고려할 필요가 있을 것이다. 이를 위해 초등학교와 중학교의 국어과 교육과정 내 듣기·말하기와 관련된 디지털 리터러시 제반 내용을 분석한 결과는 다음과 같다.

구분		해당 내용	비고(선정의 이유 등)
듣기·말하기	핵심 개념	▶ 듣기·말하기와 매체	
	일반화된 지식	의사소통의 목적, 상황, 매체 등에 따라 다양한 담화 유형이 있으며, 유형에 따라 듣기와 말하기의 방법이 다르다.	
		화자와 청자는 의사소통의 목적과 상황, 매체에 따라 적절한 전략과 방법을 사용하여 듣기·말하기 과정에서의 문제를 해결하며 소통한다.	매체에 따라 적절한 전략과 방법을 사용하는 문제 해결과 소통은 디지털 리터러시와 관련성이 높음.
	학년군별 내용 요소	[초 5-6] • 발표[매체 활용]	
		[중 1-3] • 매체 자료의 효과	
	기능	• 자료·매체 활용하기	

[그림 II-8] 초·중등학교 국어과 교육과정 내 디지털 리터러시 관련 내용
(노은희 외, 2018: 54)

국어과의 경우 ‘디지털 기술을 활용한 정보와 그 내용물의 탐색 및 활용’, ‘디지털 기술을 통해 파악한 정보와 그 내용물의 비판적 분석 및 평가’, ‘디지털 기술을 통한 정보 및 그 내용물의 생산적인 소통 및 창조’에 주로 관여한다. 특히 듣기·말하기는 매체에 대한 비판적 이해(듣기) 및 매체를 활용한 소통(말하기)과 관련한 내용이 디지털 리터러시와 밀접한 관련을 가지고 있음을 알 수 있다. 따라서 본 연구에서도 이러한 점을 반영하여 디지털 리터러시와 관련한 듣기·말하기 역량을 설정하고 조사 영역 및 요소를 설정을 고려하고자 한다.

(2) 디지털 리터러시의 기존 조사 문항 분석

가. 디지털 리터러시 관련 평가 도구

앞선 2장에서 분석한 바와 같이 OECD에서는 PISA와 PIAAC을 통해 디지털 리터러시를 평가 대상으로 다루고 있으며, 특히 PISA에서는 2015년 이후 모든 영역에서 인쇄 기반의 평가를 없애고 디지털 기반 평가를 실시하고 있다. 본 조사의 목적이 변화하는 문식 환경 속에서 국민의 국어능력을 면밀히 조사하는 데 있고, 최근 여러 조사 및 평가 사례들로부터 디지털 리터러시를 중시하고 있음을 고려하였을 때, 듣기·말하기 영역의 조사 설계 시 국민들의 디지털 리터러시를 측정할 수 있도록 해야 함은 물론이다.

이에 앞선 디지털 리터러시 실태 조사 사례 내용을 간략히 정리하고, 타 시험에서 디지털 리터러시를 다루고 있는 조사 문항의 사례를 보이겠다. 먼저 디지털 리터러시 조사 사례에 나타난 특징은 다음과 같다. 첫째, 디지털 리터러시를 측정하는 조사들은 국가 주도로 대상 인구에 대한 이해를 통해 관련 교육 및 훈련을 제공하기 위한 목적으로 일정 주기로 이루어지고 있다. 둘째, 디지털 리터러시의 조사 영역은 읽기를 중심으로 컴퓨터와 웹 활용과 관련이 있다. 셋째, 디지털 리터러시의 조사 대상은 초점 대상에 따라 청소년부터 성인까지 상이하다. 넷째, 디지털 리터러시의 조사는 대부분 컴퓨터를 통해 이루어지며, 가정 방문 또는 학교에서의 웹 접속 등의 네트워크 기반으로 한다. 이와 관련하여 다음은 디지털 리터러시를 다루는 평가의 내용을 정리한 것이다.

<표 II-43> 디지털 리터러시 관련 조사

평가명	조사 영역 및 요소	조사 대상	조사 방법
국제성인역량조사 (PIAAC)	· 영역: 문식성, 수리력, 컴퓨터 기반 문제 해결력 · 문식성: 단어 및 문장 이해, 해석, 평가 · 컴퓨터 기반 문제 해결력: 적절한 목표 및 계획 수립, 정보 접근 및 활용	24개국 ³⁾ , 16~65세, 국가별 5,000명(총 166,000명)	컴퓨터 기반 가정 방문 및 컴퓨터 기반 면담(10년 주기)
국제학업취도평가 (PISA)	· 영역: 읽기, 수학, 과학, 디지털 읽기 소양(2009, 2012) · 읽기: 과정(텍스트 과정, 과제 관리, 유창하게 읽기), 텍스트(출처, 조직과 탐색, 체재, 유형), 상황 · 디지털 텍스트 독해 평가 범주: 접근 및 확인, 통합 및 해석, 성찰 및 평가	OECD 회원국 만 15세 학생	2015년 이후 컴퓨터 기반 평가(객관식 및 구성형 문항)

온라인 국제읽기발달연구 (ePIRLS, 2016년~)	영역: 읽기 목적(문식 경험, 정보 습득 및 활용), 인지 유형(정보 확인, 추론, 해석 및 통합, 점검 및 평가)	4학년 학생(5년 주기)	웹사이트 탐색에 대한 선다형 및 구성형 문항
CBAL	- 시나리오를 통한 디지털 리터러시 성취도 평가 - ETS 연구 프로젝트 - 학습 상황에서의 인지 능력 평가	K-12 학생	비공개
DIL	- 영국 개방 대학에서 디지털 및 정보 리터러시 평가를 위한 기준 제공 4개의 수준으로 구분: 접속(0 수준), 디지털 리터러시 기초 실행 단계(1 수준), 상호 작용하는 디지털 리터러시 실행 단계(2 수준), 개성적이고 협력적인 디지털 리터러시 실행 단계(3 수준), 전문적인 디지털 정체성 단계(마스터 단계)	대학생	해당 사항 없음. (실제 평가 도구로 구현되지 않음.)
ICILS	- 각국 학생들의 컴퓨터 정보 리터러시 평가 및 비교 2013년부터 시행	중학교 2학년 학생	전국 단위 2단계 총화 표집(5년 주기, 우리나라의 경우 약 2900명 표집)
ORCA	- 교실 기반 평가 도구 - 중학생들의 디지털 텍스트 읽기 능력 평가 - 개발 진행 중	중학교 1학년 학생	미국 북동부의 2개 주에서 1200명 표집(편의 표집)
PARCC	- 대학 진학 및 직업 준비도 평가 - 미국 학생들의 성취 향상과 21세기 역량 제고를 위해 발표된 CCSS의 실제적 평가 방안을 모색하는 차세대 평가	3-8학년 학생 및 고등학생	평가를 실행하는 미국 주의 학생들이 정해진 기간에 평가에 참여 (2016년 기준 10개 주 참여)

PARCC	[수행 수준 기술의 기준]			
	• 텍스트 복잡성 수준	• 표현		
	• 읽은 내용을 출처로 하는 근거의 질	• 언어 지식과 관습		
	• 반응에 나타난 읽기 이해의 정확성			
SBA	[읽기]	• 문학 텍스트	• 정보 텍스트	
	[쓰기]	• 구성/목적	• 근거/평가	• 관습
	[말하기/듣기]	• 듣기		
	[연구]	• 정보의 해석과 통합	• 정보와 출처 분석	• 근거 사용

나. PIAAC 평가 문항

전술한바, PIAAC은 독서 리터러시와 문제 해결 능력을 핵심 정보 처리 역량으로 규정하고 이 같은 역량을 컴퓨터를 통해 측정하고 있다. PIAAC의 독서 리터러시는 매체에 따른 다양한 유형의 텍스트를 포괄한다. 이때 매체의 형식에 디지털을 포함하여 디지털 독서 리터러시를 측정할 수 있는 틀을 마련하였고 디지털 매체의 특징 중 하나인 문서를 오갈 수 있는 하이퍼링크 기능이 있는 멀티 텍스트를 포맷에 포함하였다(박주현, 2018: 66). PIAAC의 독서 리터러시 평가 틀은 다음과 같다.

<표 II -45> PIAAC의 독서 리터러시 평가 틀(박주현, 2018: 66)

구분	독서 리터러시			
내용	매체	• 인쇄 기반 • 디지털		
	포맷	• 연속/산문 텍스트(서사, 논증, 기술(Description)) • 비연속/문서 텍스트(표, 목록, 그래프) • 혼합(Mixed) 텍스트(산문과 문서 요소들의 조합) • 멀티(Multiple) 텍스트(독립적으로 생성된 요소들의 병렬 또는 연결)		
인지 전략	• 접근 및 확인	• 통합 및 해석(텍스트의 일부를 관련시키기)	• 평가 및 숙고	
맥락	• 개인적	• 직무 관련	• 지역 사회	• 교육

위의 평가 틀을 기반으로 PIAAC에서 공개하고 있는 디지털 독서 리터러시 평가 문항을 분석하여 보고하고자 한다. PIAAC에서는 디지털 독서 리터러시와 관련한 본 검사 문항의 전체 내용은 공개하지 않으나 몇몇의 문항들에 대한 구조 및 내용은 제공하고 있다. 박주현(2018)의 분석 내용을 바탕으로 디지털 독서 리터러시 평가 문항의 구조와 내용을 간략히 정리하면 다음과 같다.

<표 II -46> PIAAC의 디지털 독서 리터러시 평가 문항

문항 수준	문항 제목	내용
1 이하	선거 결과	-득표수가 가장 적은 후보자 찾기 -노동조합 선거 결과, 세 명의 선거 후보자 이름과 득표수가 적힌 표를 보고 내용을 파악해야 함.
1	복제 의약품	-국내 시장 점유율이 10% 이상인 나라의 개수 알아내기 -신문 기사와 시장 점유율을 나타낸 표를 통해 두 자료 간 연관성을 파악하고 과제를 수행해야 함.
2	건강 달리기 대회	-주최자의 연락처를 찾을 수 있는 링크 클릭하기 -웹사이트와 제시된 몇 개의 링크를 통해 과제를 수행하기 위한 적합한 링크를 파악해야 함.
3	도서관 자료 검색	-도서 검색 목록 중 특정 도서의 저자 이름을 찾기 -도서관 웹사이트에 제시된 도서 검색 목록에서 책 제목과 저자의 이름을 살피면서 많은 양의 정보 속에서 적절한 정보를 찾아내야 함.
4	도서관 자료 검색	-도서 검색 결과 목록에서 특정 내용의 책을 찾기 -책 제목과 내용을 소개하는 방대한 정보를 살펴 과제에서 요구하는 적합한 책을 찾아내야 함.

이상의 디지털 독서 리터러시 평가 문항에 더하여 PIAAC의 ‘디지털 정보 리터러시’와 밀접한 ‘디지털 문제 해결 능력’에도 주목할 수 있다. PIAAC에서 디지털 문제 해결 능력은 정보를 획득, 평가하고 타인과 의사소통하며 실제적인 과업을 수행하기 위해 디지털 기술, 의사소통 도구 및 네트워크를 활용하는 능력을 뜻한다(OECD, 2013b: 59, 박주현, 2018: 67에서 재인용.). 이러한 디지털 문제 해결 능력은 크게 과제/문제 진술, 기술(technologies), 인지 차원으로 나뉘며(OECD, 2012: 48, 박주현, 2018: 67에서 재인용.), 디지털 문제 해결 능력의 수준은 ICT 활용 측면과 정보 리터러시의 두 측면에서 설명된다. 다음은 디지털 정보 리터러시이자 디지털 문제 해결 능력의 평가를 위한 PIAAC의 평가 틀이다.

<표 II -47> PIAAC의 디지털 문제 해결 능력 평가 틀(박주현, 2018: 67)

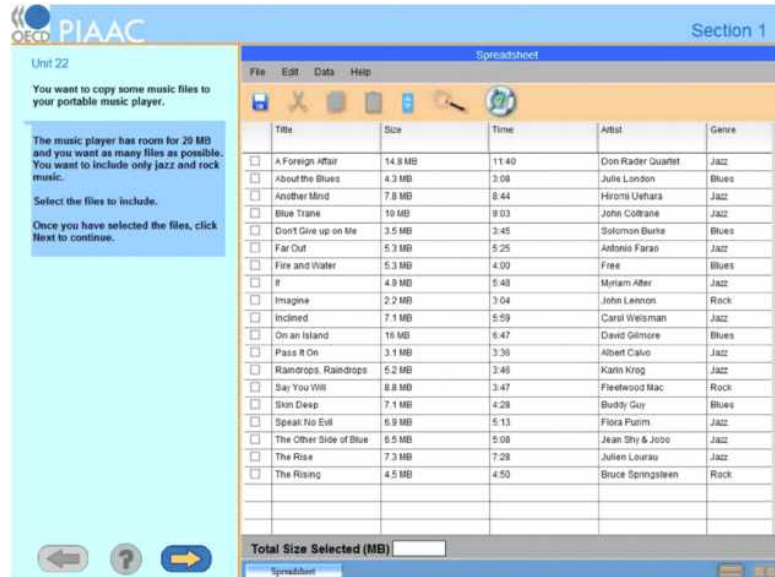
구분	디지털 문제 해결 능력	
내용	기술	•하드웨어 기기 •소프트웨어 응용 프로그램 •명령어와 기능 •구현(텍스트, 그래픽, 비디오 등)
	문항 특성	•내재적 복잡성(단계 수, 해결에 필요한 대안들, 계산 및 변환의 복잡성, 제약 조건 수) •문항 진술의 명료성(상세하게 기술된 정도)
인지 전략	•목표를 설정하고 진행 과정을 관찰하기 •정보를 획득하고 평가하기	•계획하기 •정보 활용하기
맥락	•개인적	•직무 관련 •지역 사회

이를 바탕으로 디지털 문제 해결 능력에 대한 PIAAC의 문항을 정리하면 다음과 같다.

<표 II -48> PIAAC의 디지털 문제 해결 능력 평가 문항

문항 수준	문항 제목	내용
1	파티 초대	-주어진 기준에 따라 이메일을 폴더에 나누어 분류하기 -받은 편지함에 파티 초대에 관한 답장을 포함하여 5개의 이메일이 있고, 피험자는 파티에 참석할 수 있는 사람과 그렇지 않은 사람을 구분하기 위해 답장을 분류해야 함.
2	자전거 동호회 회원 명단	-스프레드시트에서 몇 가지 기준에 따라 여러 개의 열에 담긴 대량의 정보를 처리하여 관련 항목을 적절하게 찾기 -스프레드시트에서 필요한 정보를 추출하고 이메일 보내기 -자전거 동호회에서 두 가지 조건을 만족하는 회원을 표시한 명단을 요청하는 문서 작성 페이지와 문제 해결에 필요한 정보가 제공됨. 피험자는 정렬 기능을 사용하여 필요한 정보를 추출해야 함.
	음악 파일 다운	-휴대용 음악 플레이어에 다운로드할 파일을 선택하기 -음악 장르 측면 및 용량에서의 기준을 충족해야 함. -파일 크기와 음악 장르별로 스프레드시트를 정렬할 수 있음.
3	회의실 예약	-예약 시스템을 사용하여 회의실 예약 신청을 처리하기 -예약 신청 가운데 한 건은 처리될 수 없음을 파악하고 처리 불가 이메일을 보내야 함. -회의실 예약 신청 이메일이 담긴 받은 편지함 화면과 회의실 예약 시스템 화면을 오갈 수 있음.
4	직업 탐색	-5개의 웹사이트에서 직업을 검색한 결과를 보고 등록을 요구하거나 비용을 지불하지 않는 사이트들을 북마크하기 -링크, '뒤로', '앞으로' 화살표, 홈 아이콘을 사용하여 웹페이지를 오갈 수 있음. -특정 웹페이지를 북마크하고 해당 북마크를 보거나 변경할 수 있음.

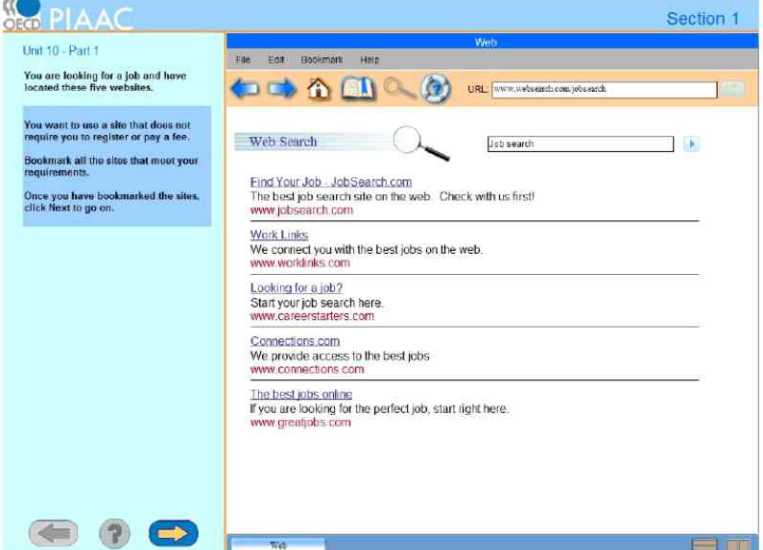
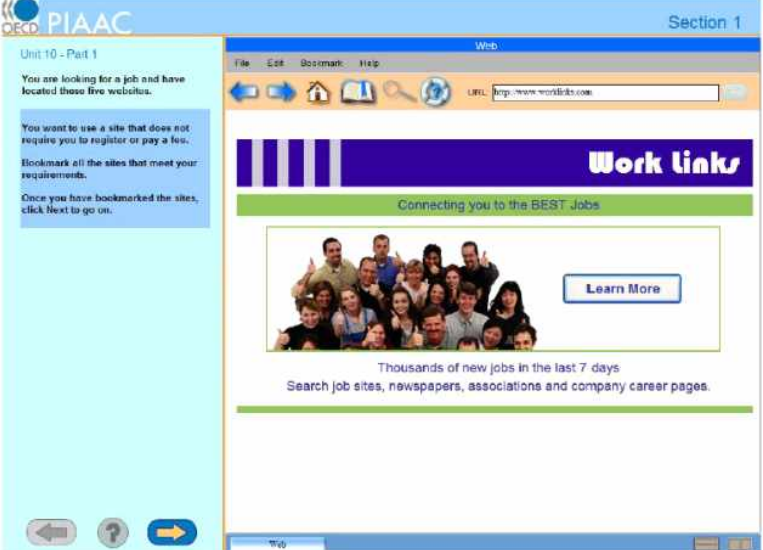
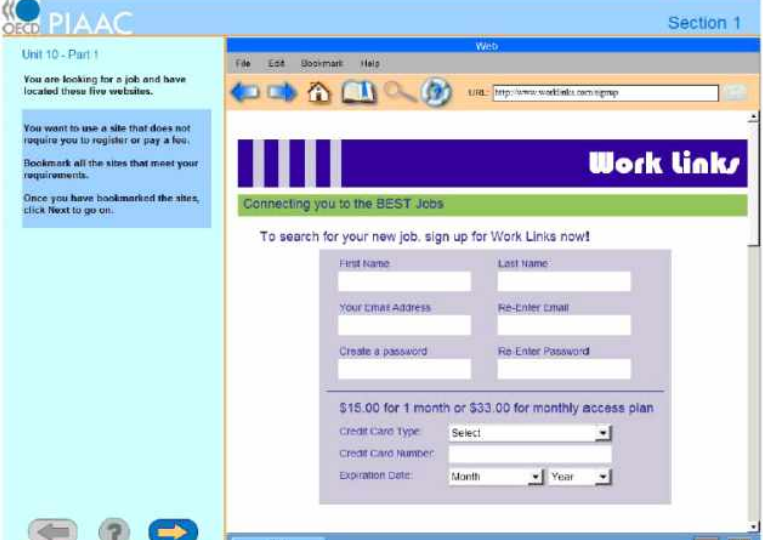
위의 문항 중 2수준의 '음악 파일 다운' 문항이 실제 구현된 화면은 다음 그림과 같다. 이 소프트웨어에는 자동 합산 기능이 포함되어 있어 파일이 선택되거나 취소될 때 총 파일 크기를 갱신함으로써 작업을 용이하게 한다. 응답자는 파일을 선택하면서 진행 상황을 모니터링하고 문제의 제시된 제약 조건을 만족하는지 여부를 확인하기 위해 지정된 기준을 확인해야 한다(박주현, 2018: 77).



[그림 II -9] 디지털 문제 해결 능력 관련 PIAAC 문항 사례(1)

더하여, 4수준의 ‘직업 탐색’ 관련 문항의 실제 모습은 다음과 같다. 피험자는 ‘직업 검색 결과 제공 화면’을 보고 해당 목록에 제시된 5개의 웹사이트 페이지에 각각 접속하여 내용을 확인한 후 지시 사항에의 부합 여부를 판단해야 한다. 아래 그림 중 ‘Work Links’는 5개의 웹페이지 중 하나이다. 피험자는 해당 사이트들이 등록을 요구하거나 수수료를 지불하지 않는 사이트에 부합하는지를 확인하기 위해 ‘Learn More’ 버튼을 눌러 세부적인 내용을 확인해야 한다.

<표 II -49> 디지털 문제 해결 능력 관련 PIAAC 문항 사례(2)

1	직업 검색 결과 제공 화면	 The screenshot shows a web browser window titled 'Web' with the URL 'www.websearch.com/jobsearch'. The page displays search results for job websites, including 'Find Your Job - JobSearch.com', 'Work Links', 'Looking for a job?', 'Connections.com', and 'The best jobs online'. The left sidebar contains instructions for finding and bookmarking job websites.
2	Work Links 화면	 The screenshot shows a web browser window titled 'Web' with the URL 'http://www.worklinks.com'. The page features a large image of a group of people and a 'Learn More' button. The text below the image states: 'Thousands of new jobs in the last 7 days. Search job sites, newspapers, associations and company career pages.'
3	Learn More 클릭 화면	 The screenshot shows a web browser window titled 'Web' with the URL 'http://www.worklinks.com/signup'. The page displays a sign-up form for 'Work Links' with fields for First Name, Last Name, Email Address, Password, and Credit Card information. The text above the form says: 'To search for your new job, sign up for Work Links now!'

다. PISA 평가 문항

PISA는 2015부터 컴퓨터 매체를 통한 독서 리터러시만을 측정하고 있다. 이는 PISA 2009에서 인쇄 매체를 기반으로 한 독서 리터러시의 측정 방식에 추가로 컴퓨터 매체에 기반한 독서 리터러시를 측정 한 결과, 인쇄 매체 독서 리터러시와 디지털 독서 리터러시의 상관관계가 0.83으로 매우 높게 나타났기 때문이다(OECD, 2011: 74). PISA의 독서 리터러시 평가 틀은 다음과 같다.

<표 II -50> PISA 2018의 독서 리터러시 평가 틀(박주현, 2021: 142 참고)

구분	범주	종류
과정	유창하게 읽기	
	텍스트 과정	• 정보 찾기 - 텍스트 내 정보에 접근하고 검색하기 - 관련 텍스트를 찾고 선택하기 • 이해하기 - 문자적 의미를 마음에 떠올리기 - 통합하고 추론하기 • 평가하고 성찰하기 - 질과 신뢰성을 평가하기 - 내용과 형식을 추론하기 - 갈등을 찾고 처리하기
	과제 관리	목표 설정 및 계획 수립, 점검, 조정
텍스트	출처	단일, 다중
	조직과 탐색	정적, 동적
	형식	연속, 비연속, 혼합
	유형	기술, 서사, 설명, 논증, 지시, 상호 작용, (복합)
상황	개인적, 공적, 교육적, 직업적, (복합)	

위의 표에서처럼 PISA 2018의 독서 리터러시 평가 틀은 ‘과정’, ‘텍스트’, ‘상황’의 세 차원으로 구분되는데, ‘텍스트 과정’에서는 디지털 독서를 반영하여 독서의 구성 요소에 디지털을 반영함으로써 디지털 정보 리터러시의 모형과 유사한 양상을 띤다(박주현, 2021: 142). 특히 아래 그림과 같은 다중 텍스트 문항은 저자가 다르거나 출판, 제목 등이 다른 다중 출처 텍스트를 활용한 문항으로 여러 텍스트와 웹사이트 등을 이동하면서 내용을 동적으로 파악해야 한다는 점에서 디지털 매체 환경에서의 독서 능력을 평가하기 위해 PISA 2018에 새로 도입된 평가 유형이다. 아울러 디지털 독서를 위해서는 탐색 역량이 필요한데 ‘조직과 탐색(항해)’은 이를 위해 등장한 요소이며, 디지털 텍스트를 탐색하고 특정 구절에 접근하기 위해 스크롤바, 탭, 텍스트 크기 및 위치 조정 도구, 하이퍼링크 등의 도구가 제공된다(김남희, 2012: 62-63).



[그림 II -10] 디지털 독서 리터러시 관련 PISA 문항 사례

또 ‘텍스트 유형’과 관련하여서 PISA 2015까지는 Werlich(1976)의 분류 체계에 따라 텍스트를 기술, 서사, 설명, 논증, 지시의 5가지 유형으로 구분한 반면 PISA 2018에서는 디지털 미디어 텍스트를 반영하여 텍스트 유형에 ‘상호 작용’을 포함하였다는 점에 주목할 필요가 있다. 이 같은 유형은 어떤 일을 하도록 요청, 회의를 조직, 친구와 사회적 약속을 하는 것과 같은 특정한 목적을 달성하는 평가 문항으로 구현된다. 이상의 내용을 바탕으로 PISA의 디지털 독서 리터러시 관련 평가 문항을 정리하여 보이면 다음과 같다.

<표 II -51> PISA의 디지털 독서 리터러시 평가 문항

문항 수준	문항 번호	내용
1	E005Q01	-화면에 제시된 내용을 읽고 물음에 답하기
2	E005Q02	-멀티 텍스트 포맷으로 접속과 검색 전략 요구 -링크를 통해 페이지를 이동하여 해당 페이지의 내용을 읽고 물음에 답하기 -1수준 문항에 비해 지시 사항에 따라 링크를 클릭하는 정도의 요구가 추가됨.
	E005Q03	-특정 날짜에 게시한 링크로 이동했을 때 연결된 웹사이트의 주요 기능 파악하기 -웹사이트의 내용이 비교적 추상적이며 익숙하지 않은 단어들이 제시됨.
4	E005Q04	-탭과 링크를 통해 여러 웹페이지를 오가며 각 웹페이지에 제시된 상세 내용을 파악하고 지시 사항에 따라 적합한 하나를 선택하여 이메일 보내기 -여러 텍스트의 종류가 결합된 혼합 유형, 서답형, 여러 인지 전략 요구

이 중 4수준의 문항의 질문은 ‘마이카의 1월 1일자 블로그를 읽으시오. ‘iwanttohelp’ 사이트로 가서

마이카를 위한 봉사활동 기회를 찾으시오. 찾은 봉사활동 기회를 마이카에게 전달하기 위하여 ‘봉사활동 기회 세부 정보’에 있는 이메일 버튼을 이용하시오. 왜 그 기회가 그녀에게 적합한지를 이메일에 설명하시오. 그런 다음 ‘보내기’ 버튼을 클릭하여 이메일을 보내시오.’와 같다. 그리고 이 문항이 실제 구현된 화면을 보이면 다음 그림과 같다. 피험자는 제시된 링크를 클릭하여 iwanttohelp 홈페이지로 이동해야 하며, 해당 홈페이지의 내용을 확인한 후 화면 오른쪽 상단에 있는 ‘Latest Opportunities’를 클릭하여 또 다른 웹페이지로 이동해야 한다. 이때 전체 웹페이지를 보기 위해서는 스크롤바를 사용해야 한다. 해당 웹페이지에는 4개의 봉사활동이 제시되어 있고, 이는 상세 내용을 확인할 수 있는 링크로 연결되어 있다. 피험자들은 탭과 링크를 활용하여 웹페이지를 수시로 오가면서 상세 내용을 파악해야 한다. 이러한 과정을 통해 물음에 적합한 하나의 봉사활동을 선택하고 ‘친구에게 이메일 보내기’ 버튼을 클릭하여 친구에게 봉사활동 기회에 관한 정보를 보내야 한다. 최종적으로 보낸 이 이메일 내용이 문항의 채점 대상이 된다.



[그림 II -11] 디지털 독서 리터러시 관련 PISA 문항 사례

이상의 평가 문항 분석 결과를 종합하면, PISA의 독서 리터러시 평가 문항은 독서의 목적에 따른 다양한 상황에서 디지털 기기와 웹브라우저라는 디지털 도구를 통해 디지털 텍스트에 대한 인지적 독서 과정 및 능력을 측정하고 있음을 알 수 있다. 그리고 이러한 평가 문항을 통해 디지털 리터러시의 핵심 요소로 정보 및 자료 공유 기술의 활용과 같은 정보의 소통 능력, 정보의 재구성 또는 새로운 정보를 창출할 수 있는 능력, 디지털 시대를 살아가는 사람들에게 요구되는 개인 정보 보호, 저작권, 네티켓 등의 윤리적 요소, 더 나아가 문제 해결을 위한 사고 과정 및 실천적 측면, 학습이나 협업 과정에 적용될 수 있는 능력 등이 강조되고 있음을 확인할 수 있다.

3) 디지털 리터러시 영역 조사를 위한 시사점

이상의 국내외 자국어능력 제반 실태 조사 사례 및 듣기·말하기 실태 조사 사례, 디지털 리터러시 조사 사례 및 조사 문항 등을 다각도로 검토하여 디지털 리터러시 영역의 조사 방향 수립을 위한 다음과 같은 시사점을 얻을 수 있다.

첫째, 여러 조사에서 디지털 리터러시를 측정하기 위해 어떠한 내용 요소를 고려하고 있는지 주목하여 이를 영역별 문항에 포함할 필요가 있다. 디지털 리터러시 조사 사례에서는 점점 디지털 리터러시가 적극적으로 고려되는 양상으로 변모하여 왔음을 확인할 수 있었다. PISA의 경우, 읽기 소양의 정의 자체에 디지털 리터러시를 반영하거나, 평가 틀에 디지털 매체를 고려한 새로운 형태의 읽기를 반영하는 요소들을 새롭게 추가하는 등 이전 주기에서 컴퓨터 기반 평가(Computer Based Assessment, CBA)와 ‘역동적인 텍스트(dynamic text)’의 개념을 도입하였던 것에서 한 발 더 나아가고 있음을 볼 수 있다. PIAAC에서도 마찬가지로 ‘적응형 문제 해결력(Adaptive Problem Solving)’을 신설함으로써, 이전 주기의 ‘컴퓨터 기반 문제 해결력’에서 보다 더 나아간 모습을 보여 준다. 이는 컴퓨터 기반 문제 해결의 상황에서 여러 문제를 동시에 해결하는 능력을 평가할 수 있도록 하기 위함으로, 영역별 조사 시 이 같은 조사 요소와 문항들을 적극 고려할 필요가 있을 것이다.

둘째, 영역별, 연차별 정례화 과정에서 1개 연도에 디지털 리터러시 영역을 별도로 상정하여 조사할 필요가 있다. PISA, PIAAC을 포함하여 많은 평가 사례를 검토한 결과 여러 평가에서 이미 디지털 리터러시 능력을 주요한 평가 대상으로 설정하고, 실제적인 디지털 리터러시를 평가하고자 하는 노력을 엿볼 수 있다. 주지하듯 본 조사의 목적이 변화하는 문식 환경 속에서 국민의 국어능력을 면밀히 조사하는 데 있으며, 이처럼 최근 여러 자국어능력 및 문식력 조사에서 디지털 리터러시를 적극적으로 반영하고 있음을 고려한다면, 연차별, 영역별 실태 조사 방안 설계 시 국민들의 디지털 리터러시를 별도의 영역으로 측정할 수 있는 방안을 적극 고려할 필요가 있다.

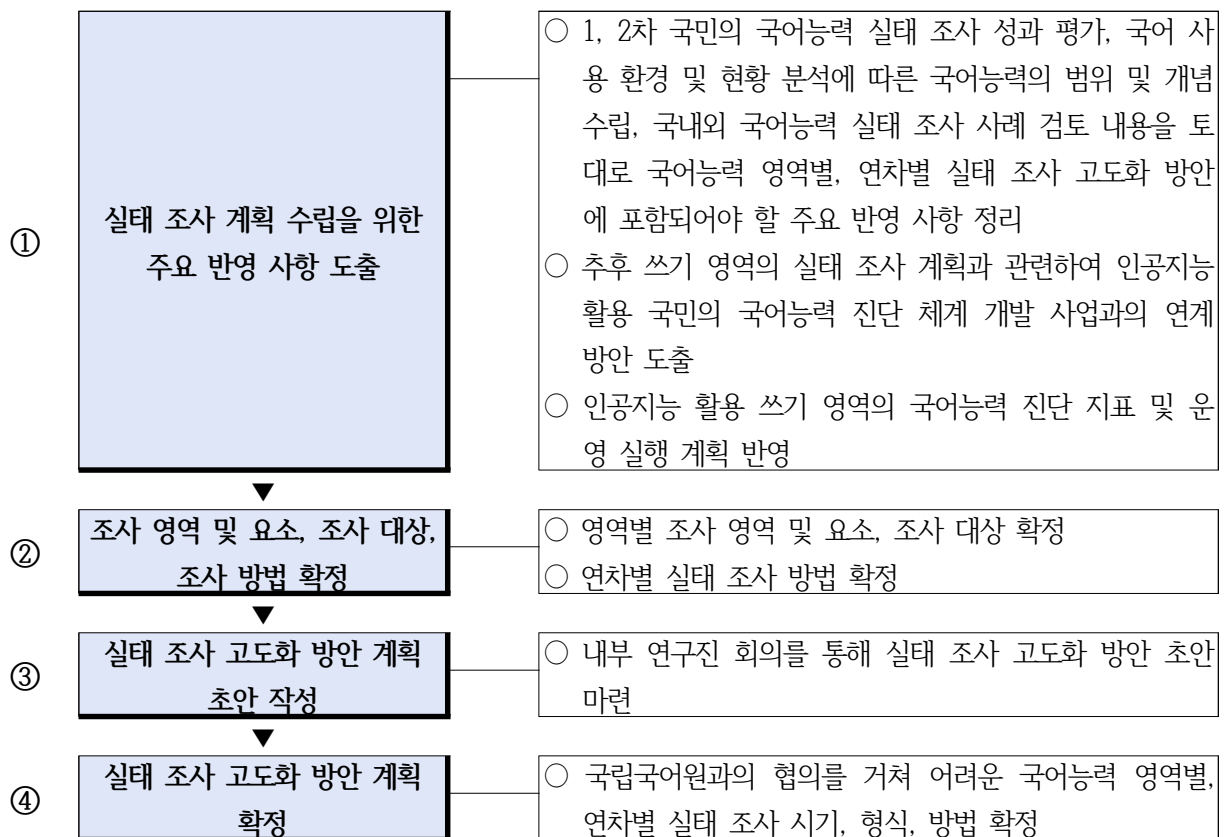
셋째, 디지털 리터러시 영역 조사 시 이에 특징적인 배경 변인 설문을 구성하여 이를 바탕으로 한 다각적인 결과 해석을 도모할 필요가 있다. 선행된 조사들 중 일부 조사에서는 배경 변인을 다양하게 설정하여 이를 설문의 형태로 조사 시행 시 함께 수집하고, 배경 설문 조사 결과를 전체 조사 결과의 해석에 활용하고 있다. 국내에서는 ‘2008년 국민의 기초 문해력 조사’나 ‘2018 국민의 국어능력 실태 조사’가 대표적이며, 국외에서는 ‘OECD 국제 성인 문해력 조사(IALS)’, ‘OECD 국제 성인 문해 및 생활능력조사(ALL)’, ‘국제학업성취도평가(PISA)’, ‘국제성인역량조사(PIAAC)’ 등이 있다. 특히 PISA와 PIAAC에서는 디지털 리터러시나 ICT 활용과 관련한 여러 배경 변인들을 설정하여 이러한 변인들과 언어능력과의 관계를 포착하고자 하였으므로 이러한 점들을 고려할 필요가 있다.

4. 국어능력 영역별·연차별 실태 조사 고도화 방안 계획

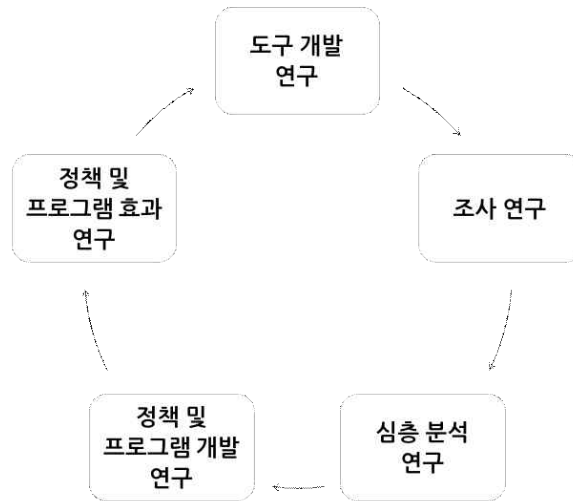
1) 5년 주기 실태 조사 사업의 연차별 정례화 방향

기존의 2013, 2018 실태 조사에서는 3년 차-5년 차 사이의 환류 정책을 시행하는 데 한계가 있었으며, 4개 영역을 통합한 대면 조사로 진행하여 조사 시간이 길고 조사 대상의 피로도가 증가하여 정확도가 떨어진다는 문제가 존재하였다. 이뿐 아니라 변화하는 국어 환경을 적시에 파악하여 반영하기 어렵다는 점에서 본 연구에서는 영역별 연차 조사 안을 제안하고자 한다. 이에 본 연구에서는 다음과 같은 과정에 따라 국어능력 실태 조사 고도화 방안을 수립하였다.

<표 II -52> 국어능력 실태 조사 고도화 방안 수립 과정



특히 기존의 1차 및 2차 실태 조사의 시행 체계 내에서는 5년 주기로 실태 조사가 진행되더라도 ‘1년 차: 조사 도구 개발 → 2년 차: 조사 시행 및 결과 분석’의 과정만 진행될 뿐 3년 차-5년 차 사이의 환류 정책을 시행하는 데 한계가 있었다. 그러나 연차별·영역별 분할 실태 조사를 시행하면 주기 내 5년 동안에 각 영역의 실태 조사 결과에 대한 심층적인 분석이 가능해짐은 물론 다음과 같은 구체적인 환류 정책을 수립하는 데 효과적일 것이라 판단된다.



[그림 II-12] 국어능력 실태 조사의 연차별 정례화

이상의 과정과 방향을 전제로 앞선 1, 2차 실태 조사의 성과를 분석하여 영역별, 연차별 실태 조사를 위한 시사점을 정리하면 다음과 같다.

<표 II-53> 1, 2차 실태 조사 성과 분석을 바탕으로 한 3차 실태 조사 시사점

	1차, 2차 실태 조사의 성과 분석	➡	3차 실태 조사를 위한 시사점
조사 도구의 측면	5개 영역 통합형 조사 도구 설계 - 비동등 집단 설계 방식으로 1시간 내외의 조사 도구 설계 - 피조사자들의 집중도가 저해되는 문제	➡	연차별 영역 분할형 조사 도구 설계
조사 인원의 측면	- 조사 도구 및 시간을 고려한 한정된 조사 인원 할당 - 관련 예산의 불충분	➡	- 조사 도구의 축소 및 예산 증대를 통한 조사 인원 확대 - 국민 관심도 증대
조사 방식의 측면	1:1 방문 면접형 조사 - 지면 및 디지털 검사지 병행 사용 - 조사 환경 및 집중도 차이 발생	➡	- 동일 공간 내방 방식으로 전환 필요 - 관련 예산 증대
결과 해석의 측면	5개 영역 통합형 조사 도구로 인해 표피적 수준에서의 해석 - 관련 환류 정책 수립의 한계	➡	- 연차별 영역 분할형 조사 도구 설계 - 영역별 하위 요소의 세분화 및 심층적 분석 - 세밀화된 환류 정책 수립
연차별 정례화의 측면	- 조사 도구 개발 및 시행에만 한정된 연구 시행 - 환류 정책 수립 및 시행의 미비	➡	- 주기 내 5년 동안 지속적인 영역별 조사 시행 - 영역별 심층 분석을 통한 환류 정책의 다각화

2) 연차별 조사 및 연구 계획

이상의 과정에 따라 본 연구에서는 다음 두 개의 방안을 국민의 국어능력 연차별 조사 및 연구 계획으로 제안한다.

(1) 1안

<표 II -54> 국어능력 실태 조사 고도화 방안(1안)

	듣기·말하기 영역	읽기, 쓰기 영역	규범, 디지털 리터러시 영역	정책 제언
1년 차 (’22년)	예비 조사 → 도구 개발			■ 연차별 영역 분할 조사 계획 수립
2년 차 (’23년)	본 조사 → 결과 분석 * 특화된 배경 설문	예비 조사 → 도구 개발		■ 듣기·말하기 영역 정책 제언
3년 차 (’24년)		본 조사 → 결과 분석 * 특화된 배경 설문	예비 조사 → 도구 개발	■ 읽기, 쓰기 영역 정책 제언
4년 차 (’25년)			본 조사 → 결과 분석 * 특화된 배경 설문	■ 규범, 디지털 리터러시 영역 정책 제언
5년 차 (’26년)				■ 조사 결과 종합 및 심층 분석

1안은 조사하고자 하는 영역의 ‘예비 조사 및 도구 개발’과 ‘본 조사’가 2년에 걸쳐 연속적으로 이루어지는 방식이다. 따라서 2023년에는 ‘듣기·말하기’ 영역의 본 조사가 이루어지는 동시에 5년이라는 조사 주기상 ‘읽기, 쓰기’ 영역의 예비 조사 및 도구 개발이 동시에 진행된다. 그리고 2024년에는 읽기, 쓰기 영역의 본 조사와 기존의 문법 영역으로 시행되었던 조사를 규범 영역으로 그 범위를 제한하여 규범, 디지털 리터러시 영역의 예비 조사 및 도구 개발을 시행한다. 이어서 2025년에는 규범 영역과 디지털 리터러시 영역의 본 조사를 실시하며, 5년 차인 2026년에는 4년 차까지의 조사 결과를 종합하고 심층적으로 분석하여 5년 주기의 국민의 국어능력 영역별, 연차별 조사를 마무리한다.

특히 본 연구에서는 상술한바, 국어능력 실태 조사를 위한 영역으로 기존의 국어능력 조사 영역인 ‘듣기·말하기’, ‘읽기, 쓰기’, ‘규범’ 영역을 준수하되, ‘디지털 리터러시 영역’을 별도의 조사 영역으로 제안한다. 앞서 본 연구에서는 1, 2차 국민의 국어능력 실태 조사 성과 평가, 국어 사용 환경 및 현황 분석, 국내외 국어능력 실태 조사 사례 검토를 통해 디지털 리터러시가 변화하는 국어 사용 환경에 필요한 국어능력임을 확인하였으며, 이미 국외 자국어능력 조사를 포함하여 2018 국민의 국어능력 실태 조사 등 최근 여러 조사 및 평가 사례들에서는 디지털 리터러시를 중시하고 있음을 알 수 있다. 이에 디지털 리

터러시 관련 요소를 기존의 듣기, 말하기, 읽기, 쓰기, 규범 영역 내에서도 언어 자료의 실제성 측면에서 반영하는 동시에 4년 차에 디지털 리터러시 영역을 별도로 상정하여 전문화할 필요가 있다. 그리고 이때의 디지털 리터러시 영역 조사는 단순히 언어 자료를 웹페이지 형태로 제시하는 수준을 넘어서서 국제성 인역량조사인 PIAAC 수준의 문항 개발 및 조사 시행 방식에 준하여 이루어져야 할 것이다. 아울러 이때에는 시나리오 기반의 CBT가 시행되어야 한다는 점에서 조사 인원의 감축도 고려할 필요가 있다. 또한 디지털 리터러시 영역을 독립적으로 상정한 첫 평가이므로 시범적 성격을 전제하고 접근하여야 할 것이다.

아울러 1, 2차 조사에서는 4개 영역 통합 조사를 실시하다 보니 해당 영역에 적합한 배경 설문 조사를 실시하지 못하였다는 점에서 이에 대한 보완책을 마련할 필요가 있다. 1차 조사인 2013년 조사에서는 인구 통계학적 변인만을 배경 설문의 변인으로 삼은 데 비하여, 2차 조사인 2018년 조사에서는 ‘독서량’, ‘ICT 친숙도’, ‘최종 전공’, ‘국어 관련 학습 활동’ 등 인구 통계학적 변인 외에 여러 심층 변인을 추가하여 배경 설문을 실시하였다는 점에서 의의가 있다. 다음 <표 II-55>는 2018년 조사의 심층 변인과, 해당 변인을 조사하는 배경 설문의 문항을 보인 것이다.

<표 II-55> 2018년 조사의 심층 변인 및 배경 설문의 내용

변인	배경 설문 내용
전공	귀하의 최종 전공은 무엇입니까? ① 어문 계열 ② 상경 계열 ③ 사회 계열 ④ 법정 계열 ⑤ 의학/약학 계열 ⑥ 이공 계열 ⑦ 자연 계열 ⑧ 사범/교육 계열 ⑨ 예술/체능 계열 ⑩ 기타 ()
독서량	귀하께서는 한 달 평균 몇 권의 책(종이책, 전자책)을 읽습니까? ① 전혀 읽지 않음 ② 1~2권 정도 ③ 3~4권 정도 ④ 5~6권 정도 ⑤ 7권 이상
인터넷 이용량 (ICT 친숙도)	귀하께서는 하루 평균 몇 시간 정도 인터넷을 이용하십니까?(모바일 포함) ① 전혀 이용하지 않음 ② 1시간 미만 ③ 1시간 이상~2시간 미만 ④ 2시간 이상~3시간 미만 ⑤ 3시간 이상~4시간 미만 ⑥ 5시간 이상
국어 관련 학습 활동 및 분야	현재 참여 중이거나 최근 1년간 참여했던 국어 관련 학습 활동(교양 글쓰기 강좌 수강 등)에 대해 질문드립니다. (1) 귀하가 참여한 국어 관련 학습 활동 형식을 모두 골라 주십시오. (복수 응답) ① 문화센터 국어 관련 강좌 수강 ② 학교 교육과정 ③ 온라인 학습 사이트 ④ 개인 교습 ⑤ 학원 수강 ⑥ 문화 학교 국어 관련 강좌 수강 ⑦ 기타() ⑧ 참여한 적 없음 (2) (앞 문항 ①~⑦ 응답자만) 귀하의 국어 관련 학습 활동과 관련된 분야를 모두 골라 주십시오. (복수 응답) ① 말하기 ② 듣기 ③ 읽기 ④ 쓰기 ⑤ 문법 ⑥ 기타()

하지만, 이 역시 읽기 영역과 관련된 변인들의 비중이 높고, 각 변인들을 다소 포괄적인 문항으로 묻고 있어 제한적이라고 볼 수 있다. 이는 2018년 조사에서 심층 변인을 설정할 때 배경 설문을 실시하고 있는 국내외 조사들('PIAAC', 'PISA', '2008년 기초문해력 조사' 등)을 참고하였는데, 이들이 주로 읽기와 관련된 조사들이었던 것이 영향을 준 것으로 보인다. 따라서 앞으로 각 영역별로 조사를 시행하게 된다면 심층 변인 또한 각 영역의 특성을 반영하여 설정될 필요가 있으며, 이때 배경 설문을 실시하고 있는 국내외 조사들뿐만 아니라, 국어능력의 각 영역별 능력에 관한 연구들과, 영역별 능력에 영향을 주는 변인들에 관한 국내외의 다양한 연구들을 참고하여야 할 것이다. 이와 함께 전문가 협의회를 실시하여 영역별 전문가와의 협의를 바탕으로 최종적으로 영역별 심층 변인들을 개발하여야 한다. 또한, 4개 영역의 통합적인 변인을 조사하는 방식이 아니기 때문에, 각 변인들에 대한 설문을 보다 구체적으로 설계할 필요가 있다. 이처럼 각 영역별로 심층 변인을 매년 개발하고, 이를 바탕으로 각 영역별 조사 결과를 해석한다면 영역별로 전문적이고 심층적인 해석이 가능할 것으로 기대된다.

따라서 1안의 2, 3, 4년 차의 '특화된 배경 설문'에서는 각각의 영역에 특화된 배경 설문을 구성하고 설문과 영역별 조사 결과와의 상관관계를 분석하는 안을 제안하였다. 마지막으로 5년 차의 '조사 결과 종합 및 심층 분석'은 다음과 같이 한국교육과정평가원의 수능 및 학업성취도 평가의 후속 연구 방식을 참조하여 수행하는 방식을 고려해 볼 수 있다.

『학업성취도 평가 및 대학수학능력시험 결과 자료 연구 활성화 지원 연구』 공모 안내

교육과학기술부와 한국교육과정평가원 공동으로 국가단위 평가 자료의 과학적 분석에 기반한 교육정책 수립을 위해 '학업성취도 평가 및 대학수학능력시험 결과 자료 연구 활성화 지원 연구' 공모를 다음과 같이 실시하오니 연구자들의 많은 참여를 바랍니다.

□ 목적

- 국가단위의 평가 자료를 활용한 연구를 적극 지원하여 객관적 데이터에 기반한 교육정책 수립을 위한 기반 조성
- 평가 결과에 근거하여 교육정책 지원의 실효성을 평가하는 책무성 체제의 기반 조성
- 대규모 평가 자료의 분석 결과를 토대로 교수학습 및 교육과정 개선하는 환류체제 구축

□ 일정

- 평가자료 활용을 위한 워크숍
 - 일시 : 5월 30일(수) 14:00
 - 장소 : 한국교육과정평가원 대회의실
- 연구 계획서 제출 및 심사 결과 발표
 - 연구계획서 온라인 접수 : 6월 4일(월) ~ 6월 11일(월) 18:00
 - 연구계획서 심사 : 6월 12일(화) ~ 6월 28일(목)
 - 연구계획서 심사 결과 발표 및 자료 탑재 : 6월 29일(금)
- 연구 수행 기간 : 6월 30일(토) ~ 11월 2일(금)
- 최종 연구물 제출 : 11월 2일(금)
- 심포지엄 개최

□ 데이터 제공

- 데이터는 2011년 대비 데이터 제공 범위가 확대될 예정임
- 제공 방법 : 웹사이트(<http://naeacsat.kice.re.kr>)에 로그인 하여 승인된 데이터에 한해 다운로드
 - 제공 요청된 모든 자료는 연구계획서에 제시된 연구 목적에 부합하는 가를 심사 후 승인된 자료에 한하여 제공
 - 자료 제공이 승인된 연구자는 자료활용동의서에 동의한 후 해당 자료 다운로드

□ 지원 대상

- 국내 대학(부설연구소 포함) 교원
- 초·중·고 교사
- 박사 과정 학생
- 국내외에 거주하는 정책 연구 수행에 적합한 기관 및 개인 등

성과물 ID	FRE 2021-10	성과물 분류코드	수시연구보고서	주제	교육평가
분류체계 대분류	교육	분류체계 중분류	초중고교육		
성과물 한글명	국가수준 학업성취도 평가 자료를 활용한 연구 활성화 지원 연구 (FRE 2021-10)				
성과물 영문명					
총서명	연구보고 FRE 2021-10				

[그림 II-13] 한국교육과정평가원의 학업성취도 후속 연구 방식 결과

이처럼 한국교육과정평가원에서는 대규모 평가 자료를 분석하여 교수·학습과 교육과정을 개선하기 위한 환류 체제를 구축할 수 있도록 연구 공모 과제를 지속적으로 진행하고 있다. 이와 같은 후속 연구 공모와 관련하여 한국교육과정평가원과 교육과학기술부가 주최한 심포지엄의 일정과 내용을 제시하면 다음과 같다.

국가수준 학업성취도 평가와 대학수학능력시험 자료 활용 분석 심포지엄				
• 일시 : 2012년 12월 7일(금) 09:30~17:10 • 장소 : 한국교원단체총연합회관 컨벤션센터 • 주최 : 한국교육과정평가원·교육과학기술부				
일시	발표 및 내용			비고
09:30~10:10	등록			
10:10~10:30	개회식			다산홀
인사말 성대제(한국교육과정평가원장)				
분과발표				
10:40~12:40	제 1 주제 교육정책과 학업성취	[발표 1] 학교제정운영 특성과 교육성과 : 단위학교 교수학습비를 중심으로 박소영, 신하영(숙명여대)	다산홀	사회 : 최인봉(KICE)
		[발표 2] 지역 및 학교수준 분석을 통한 고교-대학 연계 방안 모색 김원철, 권영신, 장정화(성균관대)		
		[발표 3] 재수생들의 수능성적 변화에 관한 탐색적 연구 김성식(서울교대), 황지현(한국방송통신대), 박미희, 배정현(서울대)		
		[토론] 이명애(KICE), 이수정(단국대), 김준엽(홍익대)		
10:40~12:40	제 2 주제 성취 모형	[발표 1] 학생의 사회경제적 수준에 따른 교육맥락 변인 효과 및 성취 특성 분석 : PISA, NAEA 연계 연구 정혜경, 조지민(KICE)	단체홀	사회 : 김원수(KICE)
		[발표 2] PISA 2003, 2006 및 2009 최상위 5% 집단의 수학·과학성취도와 학습방법 및 정서적 변인에 대한 구조방정식 모형 서혜애, 김동화(부산대), 손경우(경상대), 이봉우(단국대), 김미정(비트리아대)		
		[발표 3] 학생 배경변수의 잠재프로파일별 따른 학업성취수준의 차이 : 2009년 및 2010년 국가수준 학업성취도 데이터 분석 박철영(광주교과고), 열시창(전남대), 오선우(신창초)		
		[토론] 신택수(명지대), 이봉주(경북대), 이기중(국민대)		
12:40~14:30	점심식사			남강홀
14:30~16:30	제 3 주제 장외적 특성과 학업성취	[발표 1] 일반계 고등학생의 학습환경과 학습경험이 학업성취도와 학교 생활만족도에 미치는 영향 유영란, 김유나(이화여대)	다산홀 A	사회 : 박도영(KICE)
		[발표 2] 학교급에 따른 학생의 수업참여와 학업성취에 대한 흥미와 유용성, 지각된 유용감의 역할 송주연, 강아, 김성일(고려대)		
		[발표 3] 개인공부시간, 사교육시간이 학업적 효능감과 행동조절능력을 매개로 학업성취에 미치는 영향 : 학교급에 따른 다집단 분석 황혜영(노스캐롤라이나대), 조은별, 박수원(서울대)		
		[토론] 도승이(성균관대), 이은주(경북대), 박찬호(KICE)		
14:30~16:30	제 4 주제 교과학습과 학업성취	[발표 1] 국어 학습 관련 정서적 요소와 국어 성취도 간의 관계 연구 : 자신감, 흥미, 가치 인식과 수업 참여, 성취도 간의 관계를 중심으로 김영란(KICE), 장봉기(버지니아대)	다산홀B	사회 : 송민영(KICE)
		[발표 2] 국가수준 영어 학업성취도 성적을 통해서 본 영어공교육 강화를 위한 방안 제시 : 2004, 2006, 2008, 2011년도 초·중·고등학생들의 학습동기 및 자기주도 학습요인을 중심으로 김희경(캘리포니아주립대), 윤지환(KICE)		
		[발표 3] 대단위 평가를 활용한 우리나라 학생들의 국어 성취수준에 따른 학업 특성 분석 연구 : PISA 2009와 NAEA 2009와 2010 결과를 중심으로 최숙기(한국교원대)		
		[토론] 설현수(중앙대), 송미영(KICE), 김미경(KICE), 정혜승(경인교대)		
14:30~16:30	제 5 주제 방과후학교 효과	[발표 1] 방과후학교의 연속적 처치(sequential treatment) 효과 분석 : 고1-고2 시기 학업성취도를 중심으로 박현정, 이준호(서울대)	제1연수실	사회 : 양승호(KICE)
		[발표 2] 방과후학교 및 EBS강의가 고등학생의 학업성취도와 학습태도에 미치는 영향 백순근, 김해지, 박경인(서울대)		
		[발표 3] 2009년 학업성취도와 PISA 2009를 연계한 방과후활동 효과 분석 : 사회적 포섭 개념을 활용하여 임미령, 임정국, 윤재현, 이현(동국대학교)		
		[토론] 박인용(KICE), 김희경(KICE), 정혜원(충남대)		
14:30~16:00	제 6 주제 수직적도 기반의 학업성장 평가	[발표 1] 학업성취도 평가의 수직적도 구축 및 학업성장 지표 개발 장경아, 이재희(KICE), 강태훈(성신여대)	제2연수실	사회 : 김인숙(KICE)
		[발표 2] 초, 중, 고 학업성취도 평가들의 수직연계성 분석 : 국어, 수학 교과의 성취기준을 중심으로 박선화(KICE), 남가영(아주대)		
		[토론] 임현경(KEDI), 김화경(상명대)		
14:30~17:10	제 7 주제 다문화·탈북 가정 학생의 학업성취 특성 분석	[발표 1] 학업성취도 평가 결과에 기반한 다문화·탈북 가정 학생의 학업성취 특성 신진아, 박성욱(KICE), 최숙기(한국교원대)	단체홀	사회 : 김종훈(서울하늘초)
		[발표 2] 다문화·탈북 가정 학생의 국어와 학업성취도 평가 결과 분석 김영란(KICE)		
		[발표 3] 다문화·탈북 가정 학생의 2011년 중3 사회와 학업성취도 평가 결과 이정우, 서민원(KICE)		
		[발표 4] 차별기능문학 분석을 통해 살펴본 다문화 가정 학생의 영어 학업성취 특성 이영주(KICE)		
		[토론] 강태훈(성신여대), 손원숙(경북대), 권순희(전주교대), 설규주(경인교대), 김경현(한국교원대)		
17:10	폐회			

[그림 II -14] 학업성취도 및 대학수학능력시험 자료 활용 분석 심포지엄

위의 그림에서 확인할 수 있듯이 평가 결과의 변인별 성취 특성 분석을 비롯하여 이전 평가 결과와의 비교 분석, 평가 결과의 교육적 환류 방향 모색 등 다양한 주제의 연구 및 발표가 이루어졌음을 알 수 있다. 이러한 심포지엄을 포함하여 후속 연구 결과 최종적으로 발간된 보고서의 목차를 제시하면 다음과 같다.

학업성취도 평가와 대학수학능력시험 결과 자료를 활용한 연구 활성화 지원



차례

I. 서론	1	1. 연구의 필요성 및 목적	3
2. 연구 내용	5	가. 대수능과 학업성취도 성적 데이터 클리닝 및 변수 확인	5
나. 학업성취도 평가와 연계된 설문 자료 검토 및 레이아웃 작성	6	다. 대수능 및 학업성취도 성적 데이터 파일 레이아웃 확인 및 검토	6
라. 연구 자료를 활용한 연구 계획서 공모 및 심사	6	마. 연구 결과에 대한 심포지엄 개최	6
3. 연구 방법	7	가. 문헌 연구	7
나. 통계 분석	7	다. 전문가 협의회	7
라. 워크숍 개최	8	마. 심포지엄 개최	8
II. 2012년 평가 자료 활용 연구공모 활성화 지원 개선	9	1. 자료 공개 범위의 확대	11
가. 고등학교 2학년 학교 향상도 자료	12	나. 2009년(고1)-2010년(고2) 학업성취도 연계 자료	13
다. 성취도 고1(2009)-PISA(2009) 연계 자료	13	라. 초·중학교 학업성취도 표집 자료(최도점수 제공)	14
2. 공개 자료 처리 및 생성	14	가. 고등학교 2학년 학교 향상도 자료	15
나. 2009년(고1)-2010년(고2) 성취도 연계 자료	16	다. 성취도 고1(2009)-PISA(2009) 연계 자료	18
III. 평가 자료 활용 연구공모 및 심포지엄 개최	21	라. 자료 제공 시기의 개선	130
1. 연구공모의 개요	23	마. 기타 개선 사항	131
가. 목적	23	V. 결론 및 제언	133
나. 연구과제 유형과 지원예산	23	1. 결론	135
다. 연구공모 절차	25	가. 2012년 평가 자료 활용 연구공모 활성화 지원 개선	135
2. 홍보 및 워크숍 개최	28	나. 평가 자료 활용 연구공모 및 심포지엄 개최	137
가. 공모 홍보	28	다. 평가 자료 활용 연구 활성화 개선 방안	140
나. 학업성취도 평가 및 대학수학능력시험 결과 자료 연구 활성화 지원 연구 워크숍	28	2. 정책 제언	141
3. 연구과제 공모 및 심사	30	가. 연구공모 방식 및 절차 개선을 위한 제언	142
가. 연구과제 공모	30	나. 연구 공모 활성화를 위한 중·장기적 개선을 위한 제언	147
나. 연구과제 심사	30	* 참고문헌	151
4. 심포지엄 개최	60	* Abstract	155
가. 심포지엄 실행 계획	60		
나. 심포지엄 세부 일정	60		
다. 심포지엄 결과	64		
5. 평가자료 활용 분석 연구 결과	66		
가. 심포지엄 세션별 연구개요	66		
나. 심포지엄 발표 연구	72		
다. 심포지엄 발표 연구의 정책 제언	85		
IV. 평가 자료 활용 연구 활성화 개선 방안	101		
1. 연구공모 방식 및 절차 개선 방안	103		
가. 연구비 지원	103		
나. 연구공모 방식	106		
다. 공모 심사 강화	112		
2. 평가 자료 활용 연구 활성화를 위한 중·장기적 개선 방안	120		
가. 자료 공개 범위의 확대	120		
나. 타 기관의 자료와의 연계	122		
다. 자료 제공 대상의 확대	128		

[그림 II-15] 한국교육과정평가원의 학업성취도 후속 연구 결과 목차

이처럼 위의 과정을 참조하여 국민의 국어능력 실태 조사의 경우 ‘양적 데이터 및 실제 응답 자료(쓰기, 말하기) 공유 → 연구 주제 공모 → 연구 결과에 대한 연구 보고 대회 개최 → 별도의 홈페이지 결과 공유 + 국어원 혹은 관련 기관 발간 소식지 등과 연계’와 같은 방식에 더하여 국어 정책 통계 포털, 홍보 방안과 연계하는 방식 역시 적극 고려할 수 있을 것이다. 예컨대, 가장 가깝게는 다음과 같이 국립국어원에서 시범 운영 중인 국어 정책 자료 홈페이지를 이용할 수 있다. 국어 정책 자료 홈페이지는 국어 정책과 관련된 자료 정보를 객관적이고 구체적인 수치로 제공하여 국어 정책의 현황을 쉽게 이해하도록 하는 목적에서 제공되고 있으며, 2011년부터 구축된 국내외 국어 정책 자료가 모여 있어 언어 및 국어 연구와 조사에 이용할 수 있도록 한다. 따라서 연차별, 영역별 국민의 국어능력 조사 결과 자료를 공유하고, 이를 기반으로 한 분석 결과 활용 연구 및 정책 등 역시 홍보하고 그 결과물을 게시할 수 있다.

[그림 II -16] 국립국어원 국어 정책 자료 홈페이지

이 외에도 대한민국 정책브리핑(www.korea.kr), 국가정책연구포털(www.nkis.re.kr), 국가통계포털(kosis.kr) 등의 국가적인 사이트와 뉴스 포털 등을 활용하여 5년 차의 ‘조사 결과 종합 및 심층 분석’과 관련한 내용을 다각도로 홍보할 수 있을 것이다. 참고로 정책브리핑에서는 매년 국가 수준 학업성취도 평가 결과를 공유하고 있으며, 국가정책연구포털에서는 교과별 학업성취도 평가 결과에 관한 연구 보고서를 탑재하고 있다. 국가통계포털에서도 한국어능력시험 등급 등에 국어 제반 통계 자료를 게시하고 있다는 점에서 이들과 연계하여 연구를 수행하고 홍보하는 방안을 고려해 볼 수 있다.

(2) 2안

앞선 1안의 경우 2년 차, 3년 차의 연구 수행 과정에서 특정 영역의 도구 개발과 또 다른 영역의 본 조사 결과 분석이 동시에 진행되는 일정이라는 점에서 연구진의 구성이 다양화될 수밖에 없는 한계가 존재한다. 따라서 아래의 2안과 같이 하나의 연차 연구 동안 하나의 영역의 도구 개발과 본 조사 시행 및 결과 분석이 이루어지는 방안도 고려할 필요가 있다.

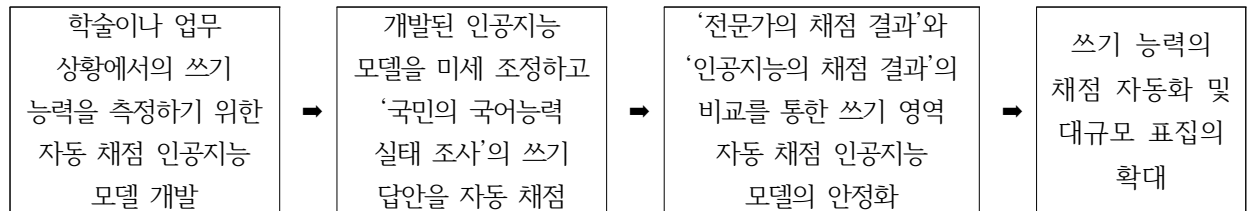
<표 II -56> 국어능력 실태 조사 고도화 방안(2안)

	듣기·말하기 영역	읽기, 쓰기 영역	규범, 디지털 리터러시 영역	정책 제언
1년 차 (’22년)	예비 조사 → 도구 개발			■ 연차별 영역 분할 조사 계획 수립
2년 차 (’23년)	본 조사 → 결과 분석 * 특화된 배경 설문			■ 듣기·말하기 영역 정 책 제언
3년 차 (’24년)		예비 조사 → 도구 개발 본 조사 → 결과 분석 * 특화된 배경 설문		■ 읽기, 쓰기 영역 정책 제언
4년 차 (’25년)			예비 조사 → 도구 개발 본 조사 → 결과 분석 * 특화된 배경 설문	■ 규범, 디지털 리터러 시 영역 정책 제언
5년 차 (’26년)				■ 조사 결과 종합 및 심 층 분석

(3) 인공지능 활용 연계 방향

본 연구에서는 국어능력 영역별, 연차별 실태 조사 고도화 방안 계획을 수립하는 과정에서 ‘인공지능 활용 국어능력 진단 체계 개발’ 과업과의 연계 방향을 제안하고자 한다. 해당 사업이 아직은 초기 단계라는 점에서 구체적인 연계 계획을 수립하는 데에는 난점이 있으므로 본 연구에서는 중장기적인 계획의 특성을 전제하고 ‘쓰기’ 영역에서의 활용 방향을 다음과 같이 제안해 보고자 한다.

<표 II -57> 인공지능 사업과 국어능력 실태 조사의 연계안



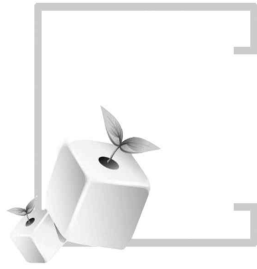
‘인공지능 활용 국어능력 진단 체계 개발’ 과업은 학술이나 업무 상황에서 기초가 되는 쓰기 능력을 진단하는 데 목적을 두고 있기 때문에, 해당 연구에서 제시하는 쓰기 과제는 본 연구의 쓰기 과제보다 내용, 조직, 표현의 수준이 높은 편이다. 즉, 본 연구는 전 국민을 대상으로 한 조사이기 때문에 위 과업에 비하여 쓰기 과제의 수준이 낮고 답안의 길이 자체도 짧다. 이에, 해당 과업에서 중장기적으로 개발하고자 하는 자동 채점 인공지능 모델이 구축된다면, 해당 모델을 본 연구에서 조사하고 있는 ‘쓰기’ 영역의 답안에 적용할 때 모델을 곧바로 사용하기보다는, 본 연구의 목적에 맞게 미세 조정(Fine-tuning, Sudharsan Ravichandiran, 전희원, 정승환, 김형준 역, 2021: 75-76)을 할 필요가 있다.

이를 위하여, 우선 본 연구의 쓰기 영역 답안에 대한 전문가 채점 데이터가 확보되어야 한다. 그리고 쓰기 영역 답안 데이터의 일부와 해당 답안의 전문가 채점 결과를 함께 모델에 학습시킨 후, 학습시키지 않고 남아 있는 일부 데이터를 대상으로 학습된 모델이 자동 채점을 수행하도록 한다. 이후, 모델이 자동 채점한 결과를 전문가의 채점 결과와 비교하여 자동 채점을 안정화하는 과정이 후행되어야 한다.

만약 모델이 자동 채점한 결과를 전문가의 채점 결과와 비교하였을 때 그 정확도가 낮은 경우, ‘인공지능 활용 국어능력 진단 체계 개발’의 인공지능 모델이 중장기적으로 어떻게 개발될지 정확하게 알 수는 없지만, 해당 모델을 개발할 때 설정한 여러 가지 요소들이 학술, 직무 작문의 채점에는 용이하게 작용하였으나 본 연구의 작문 채점과는 큰 관련이 없기에 정확도가 낮게 나온 것으로 보이므로, 이러한 요소들의 기준을 조정하여 본 연구의 쓰기 영역에 보다 적합하도록 자동 채점 인공지능 모델을 안정화할 수 있을 것으로 보인다.

이를 통하여 장기적으로 쓰기 능력의 채점이 자동화된다면, 대규모 조사를 채점하는 것에 따른 비용과 인력 문제를 해결할 수 있게 되고, 나아가 이를 바탕으로 더 많은 국민을 대상으로 한 조사도 가능하게 될 것으로 보인다.

한편, 현재 인공지능 사업이 ‘쓰기’ 영역에 국한되어 시행되고 있으므로 추후 ‘듣기·말하기’ 영역과 관련하여서도 인공지능 사업을 추진할 필요가 있을 것이며, 인공지능 사업과의 본격적인 연계는 국민의 국어능력 2-3년 차(1안) 혹은 3년 차(2안)인 읽기, 쓰기 영역에서 인공지능의 쓰기 채점표, 문항 유형 등을 고려하여 문항을 개발하고 선정할 수 있도록 연계할 필요가 있다. 아울러 인공지능과의 연계는 선다형 문항에는 적용되지 않으며, 말하기, 쓰기와 같은 표현형 조사 문항에 적용하는 방식으로 계획을 구체화해 나가야 할 것이다.



제 3 장

조사 도구의 개발



1. 조사 도구 개발의 절차

본 연구에서는 조사 도구의 개발 절차는 예정된 계획 및 일정에 따라 이루어졌다. 주요 과제는 영역별 이론 구축, 조사 도구 구성, 1차 예비 조사의 설계 및 시행, 1차 예비 조사의 채점 및 결과 분석, 예비 조사 결과를 바탕으로 한 조사 도구 및 조사 방식의 보완, 2차 예비 조사 시행, 2차 예비 조사의 채점 및 결과 분석, 조사 도구 정교화 및 관련 정책 제언이며 세부 내용은 아래의 <표 Ⅲ-1>을 참고할 수 있다.

<표 Ⅲ-1> 본 연구의 조사 도구 개발 절차

주요 과제	세부 내용
이론 구축	- 국어의 세부 영역 전문가들이 참여하여 영역별 국어능력을 개념화하고, 국어능력의 하위 요소를 탐색함. - 국내외 자국어능력 조사 및 국어능력 관련 선행 연구를 조사하여 국어능력을 재개념화하고 듣기·말하기 영역 조사 도구의 설계 방향을 설정함.
조사 도구의 구성	‘듣기·말하기’의 세부 조사 영역 및 요소 설정 - 가교 문항의 검토 - 문항 형식 결정 - 검사지 구성
1차 예비 조사 시행	‘성별, 연령대, 지역 규모, 학력, 직업군’의 5개 변인 설정 - 인구 분포 통계를 활용한 2단계 층화 후 집락 추출, 집락에서 개인 할당 추출 방식을 적용하여 조사 대상의 대표성 확보 - 서울, 인천, 경기 동 지역(중소도시), 경기 읍면 지역 4개의 층으로 구분 - 조사원 선발 및 교육 240명을 대상으로 예비 조사 실시 - 조사 검증
1차 예비 조사의 채점 및 결과 분석	- 예비 조사 결과의 채점 - 영역별 점수 분포를 분석하여 조사 도구의 적합성 분석 - 조사 시행 방식의 타당성 분석 - 조사 대상 표본 추출 방식의 타당성 분석 - 행렬 표집 설계 검사지 간 동등화, 수직 척도화 분석 → 신규 문항의 적절성, 가교 문항의 적절성 평가
조사 도구 및 조사 방식의 보완	1차 예비 조사 결과를 바탕으로 영역별 삭제 및 수정 문항의 선정 기준 마련 - 조사 시행 방식의 타당도 조사 대상의 대표성 확인

	▪ 2차 예비 조사를 위한 영역별 문항 삭제 및 수정 ▪ 연구진 평가회 및 전문가 검토
▽	▽
2차 예비 조사 시행	▪ ‘성별, 연령대, 지역 규모, 학력, 직업군’의 5개 변인 설정 ▪ 인구 분포 통계를 활용한 2단계 층화 후 집락추출, 집락에서 개인 할당 추출 방식을 적용 ▪ 서울, 인천, 경기 동 지역, 경기 읍면 지역 4개의 층으로 구분 ▪ 조사원 선발 및 재교육 ▪ 240명을 대상으로 예비 조사 실시 ▪ 조사 검증
▽	▽
2차 예비 조사의 채점 및 결과 분석	▪ 1, 2차 예비 조사 결과를 바탕으로 영역별 문항 삭제 및 수정 ▪ 조사 시행 방식의 타당도 조사 대상의 대표성 확인 ▪ 연구진 평가회 및 전문가 자문단 검토 ▪ 본 조사를 위한 조사 도구의 최종 선정
▽	▽
조사 도구 정교화 및 관련 정책 제언	▪ 각 조사 영역별 세부 결과 분석 및 문항의 정교화 ▪ 연구원 평가회 및 전문가 자문단 검토 ▪ 관련 정책 수립을 위한 전반적, 영역별 제언

2. 영역별 조사 도구의 구성

1) 듣기 영역

(1) 세부 조사 영역 및 요소 설정

본 연구에서는 2018년 조사와의 일관성을 유지하기 위해서 듣기와 말하기 기본 조사 틀을 준용하되, 현대 사회에서 요구되는 역량의 개념을 고려하여 실제 삶과의 연관성 및 협력적 소통에 대한 필요성을 반영하고자 하였다. 또한 매체 환경의 변화에 따라 강조되고 있는 디지털 리터러시와 관련한 내용을 반영함으로써 듣기와 말하기 영역의 조사 타당성 및 실제성을 제고하고자 하였다. 이를 위해 기존의 사실적, 추론적, 비판적 듣기에 더하여 공감적 듣기를 중영역 평가 항목으로 신설하고, 상대의 입장이나 처지를 파악하며 들을 수 있는 역량을 갖추고 있는지를 조사하고자 하였다.

그동안의 국어교육에서는 인지적 관점에 바탕을 두고 구어 담화를 정확하게 이해하는 것을 중요한 듣기 능력으로 간주하였다(정상섭, 2005). 그러나 최근에는 대인 관계를 위한 의사소통 능력이 중시되며, 상대와의 관계 형성 및 원활한 상호 작용을 위해서 상대의 처지를 이해하고 감정을 공유하려는 ‘공감’의 자세에 대한 필요성이 강조되고 있다(박재현, 2016). 따라서 기존의 인지적 차원의 듣기 능력에 대한 점검 외에 정서적, 대인관계적 측면의 능력을 측정하기 위해 공감적 듣기의 기본적인 역량과 관련한 내용을 추가하였다.

듣기·말하기 능력은 음성 언어를 통한 의사소통이라는 점을 기반으로 구어 담화를 잘 듣고 이해하고, 구어를 통해 자신의 생각을 효과적으로 표현하는 것이라고 정의되어 왔다. 그러나 현대 사회에서는 듣기와 말하기의 대상이 되는 매체가 단순히 구어에 머무르지 않고, 다양한 디지털 매체를 통해 듣고 말하는 행위가 이루어지고 있다. 더불어 이러한 환경의 변화에 따른 새로운 듣기·말하기 능력이 요구되고 있다. 언어 사용자들의 일상생활 의사소통의 상당한 부분이 디지털 매체를 통해 이루어지고 있는 만큼, 현대 사회 구성원의 구어 소통 능력을 타당하게 평가하기 위해서는 디지털 매체를 통한, 혹은 디지털 매체에 대한 듣기·말하기 능력을 측정할 수 있는 문항이 필요하다. 따라서 본 연구에서는 디지털 리터러시⁴⁾와 관련한 문항을 설계함으로써 디지털 리터러시 역량을 함께 조사할 수 있도록 구성하였고, 비판적 듣기의 하위 영역인 형식의 조사 문항에 매체 사용 효과의 적절성을 조사하는 문항을 설계하였다. 이는 앞서 언급한 바와 같이 현대 사회 의사소통의 변화 상황에 대한 고려 및 사회 구성원에게 요구되는 의사소통 역량을 반영하여 설정한 것으로, 보다 실제적인 듣기 능력 조사가 이루어질 수 있도록 한 것이다. 구체적인 세부 조사 영역 및 요소는 <표 III-2>와 같다.

4) ‘디지털 리터러시(digital literacy)’의 개념 및 관련 논의는 II장 3절 2) 참조.

<표 Ⅲ-2> 2018, 2022 세부 조사 영역 및 요소

2018			2022		
중영역	하위 영역	문항 수	중영역	하위 영역	문항 수
가. 사실적 듣기	1. 지시 이행 능력	1	가. 사실적 듣기	1. 지시의 이행 능력	1
	2. 세부 정보의 파악	2		2. 세부 정보의 파악	2
	3. 핵심 정보의 파악	1		3. 핵심 정보의 파악	2
나. 추론적 듣기	1. 내용의 추리	2	나. 추론적 듣기	1. 내용의 추리	1
	2. 심리 또는 태도의 추리	1		2. 심리 또는 태도의 추리	2
	3. 적용	1		3. 적용	2
다. 비판적 듣기	1. 내용의 평가	2	다. 비판적 듣기	1. 내용의 평가	2
	2. 형식의 평가	1		2. 형식의 평가	1
			라. 공감적 듣기	1. 상대의 입장이나 처지 파악	2
합계		10	합계		15

담화 유형 역시 2018년 조사와의 검사 도구 동등화를 고려하여 설정하였으며, 국립국어원(2012: 62)의 기초 연구에서의 ‘가능한 한 국민들이 실제 생활에서 흔히 접할 수 있는 담화 유형을 대상으로 듣기 능력을 평가한다.’는 내용을 준용하였다. 다만 추가된 ‘공감적 듣기’와 관련한 담화 유형을 설정하기 위해 ‘설명’이나 ‘설득’과 다른 ‘친교’적 담화 목적에 ‘대화’를 추가하였으며, 디지털 리터러시 관련 역량을 고려하여 다양한 매체에 대한 고려가 담화 유형에 반영될 수 있도록 하였다. 또한 2018년 틀에서 설명과 설득에 중첩되는 담화 유형으로 ‘대화’만을 제시한 것을 수정하여, ‘강연’을 추가하고 대화와 강연은 설득, 설명이 모두 고려될 수 있는 담화 유형임을 분명히 하였다. 마지막으로 담화의 목적이나 상황이 어느 한쪽에 치우치지 않고 고르게 안배될 수 있도록 고려하였다. 2018년도 본 조사와 2022년도에 시행될 예비 조사의 구체적인 듣기 영역 담화 유형은 <표 Ⅲ-3>, <표 Ⅲ-4>와 같다.

<표 III-3> 2018년 본 조사 듣기 영역 담화 유형

문항 번호	담화 유형 능력	설명적			설득적				
		지시	뉴스	면담	대화		강연	뉴스	토론
1	가-1			○					
2	나-2			○					
3	가-3	○							
4	가-2		○						
5	가-2		○						
6	다-1								○
7	다-1								○
8	가-1						○		
9	나-1			○					
10	나-1		○						
11	나-1							○	
12	다-2					○			
13	나-3						○		
	계	1	3	3	1		2	1	2

<표 III-4> 2022년 예비 조사 듣기 영역 담화 유형⁵⁾

문항 번호	담화 유형 능력	설명			설명, 설득				설득		친교
		지시	뉴스	면담	대화		강연		뉴스	토론	대화
					설명	설득	설명	설득			
1	가-1	○■									
2	가-2			■			○				
3	가-2							○	■		
4	가-3		○					■			
5	가-3				■					○	
6	나-1			○						■	
7	나-2					■					○
8	나-2			○						■	
9	나-3		■					○			
10	다-1						■			○	
11	다-2					○		■			
12	라-1			■							○
13	라-1				○						■
계		2	2	4	1	2	2	4	1	4	3

5) 예비 조사를 위해 각 기준별로 2배수의 문항을 개발하였다. ○은 (가)형, ■는 (나)형 검사지 구성을 위해 개발한 문항이다.

(2) 가교 문항 검토

2022년 예비 조사 및 2023년 본 조사를 위한 가교 문항으로는 2018년도 [L-5], [L-10] 두 문항을 선정하였다. 해당 가교 문항은 가교 문항이 조사하고자 하는 피험자의 능력 속성을 대표할 수 있어야 하며 전체 문항의 곤란도 분포를 반영해야 하고, 높은 변별도를 지녀야 한다는 원칙에 의거하여 선정된 것이다. 2018년 조사 당시 듣기 영역의 곤란도 평균은 0.65였는데, 가교 문항 각각의 곤란도는 0.49, 0.68로 곤란도가 적절한 것으로 나타났고, 변별도 역시 두 문항 각각 0.32, 0.33으로 변별력이 있는 문항으로 분석되었다. [L-5]는 라디오 방송의 토론을 듣고 토론 참여자가 상대의 주장을 반박하는 질문으로 적절한 것을 묻는 문항이고, [L-10]은 강연을 들은 청중의 반응으로 적절하지 않은 것을 묻는 문항이다. [L-5]는 담화 유형상으로는 토론에, 목적상으로는 설득적 담화에 속하며, 조사 영역은 ‘비판적 듣기’의 ‘내용의 평가’에 속한다. [L-10]은 담화 유형상으로는 강연에, 목적상으로는 설득적 담화에 속하며, 조사 영역은 ‘추론적 듣기’의 ‘적용’에 속한다.

담화 지문의 내용은 각각 ‘세는 나이를 만 나이로 통일해야 한다’는 논제에 대한 토론과 ‘4차 산업혁명 시대의 특징과 이를 위한 준비’에 대한 강연의 내용이다. 내용적인 측면에서도 현재 유효하며, 2018년도의 내용과 달라진 점은 없으므로 2022년 예비 조사 및 2023년 본 조사에도 활용될 수 있음을 확인하였다.

듣기 1번 문항 (가교)				
2022 국민의 국어능력 실태 조사	예비 조사	L-1	검사지	가, 나
	비판적 듣기 (내용의 평가)	설득적 담화		가교 문항
조사 문항				

자료 출처																	
해설																	
본 조사 결과	<table> <tr> <td rowspan="2">본 조사 결과 (고전 검사 이론)</td><td>변별도</td><td>0.320</td><td>변별력이 있는 수준으로 나타남.</td></tr> <tr> <td>곤란도</td><td>0.497</td><td>곤란도가 적절한 수준으로 나타남.</td></tr> <tr> <td rowspan="2">본 조사 결과 (문항 반응 이론)</td><td>변별도</td><td>0.968</td><td>변별도가 적절한 수준으로 나타남.</td></tr> <tr> <td>곤란도</td><td>0.851</td><td>곤란도가 어려운 수준으로 나타남.</td></tr> </table>	본 조사 결과 (고전 검사 이론)	변별도	0.320	변별력이 있는 수준으로 나타남.	곤란도	0.497	곤란도가 적절한 수준으로 나타남.	본 조사 결과 (문항 반응 이론)	변별도	0.968	변별도가 적절한 수준으로 나타남.	곤란도	0.851	곤란도가 어려운 수준으로 나타남.		
본 조사 결과 (고전 검사 이론)	변별도		0.320	변별력이 있는 수준으로 나타남.													
	곤란도	0.497	곤란도가 적절한 수준으로 나타남.														
본 조사 결과 (문항 반응 이론)	변별도	0.968	변별도가 적절한 수준으로 나타남.														
	곤란도	0.851	곤란도가 어려운 수준으로 나타남.														

듣기 2번 문항 (가교)				
2022 국민의 국어능력 실태 조사	예비 조사	L-2	검사지	가, 나
	추론적 듣기(적용)	설득적 담화		가교 문항
조사 문항				
자료 출처				
해설				

본 조사 결과	본 조사 결과 (고전 검사 이론)	변별도	0.337	변별력이 있는 수준으로 나타남.
		곤란도	0.685	곤란도가 적절한 수준으로 나타남.
	본 조사 결과 (문항 반응 이론)	변별도	1.247	변별도가 적절한 수준으로 나타남.
		곤란도	-1.433	곤란도가 쉬운 수준으로 나타남.

(3) 문항 형식

듣기 영역의 경우 피험자가 듣기 지문을 들으며 문제를 풀되, 발문과 선지는 읽을 수 있도록 검사 화면에 제시하였다. 각 문항을 풀기에 앞서 피험자가 듣게 될 듣기 지문의 담화 유형을 알 수 있도록 다음과 같은 발문을 화면상에 제시하며 동시에 음성으로 들을 수 있도록 하였다.

[L-1] ※ 다음은 희망은행의 자동 응답 전화입니다. 잘 듣고 물음에 답하세요.

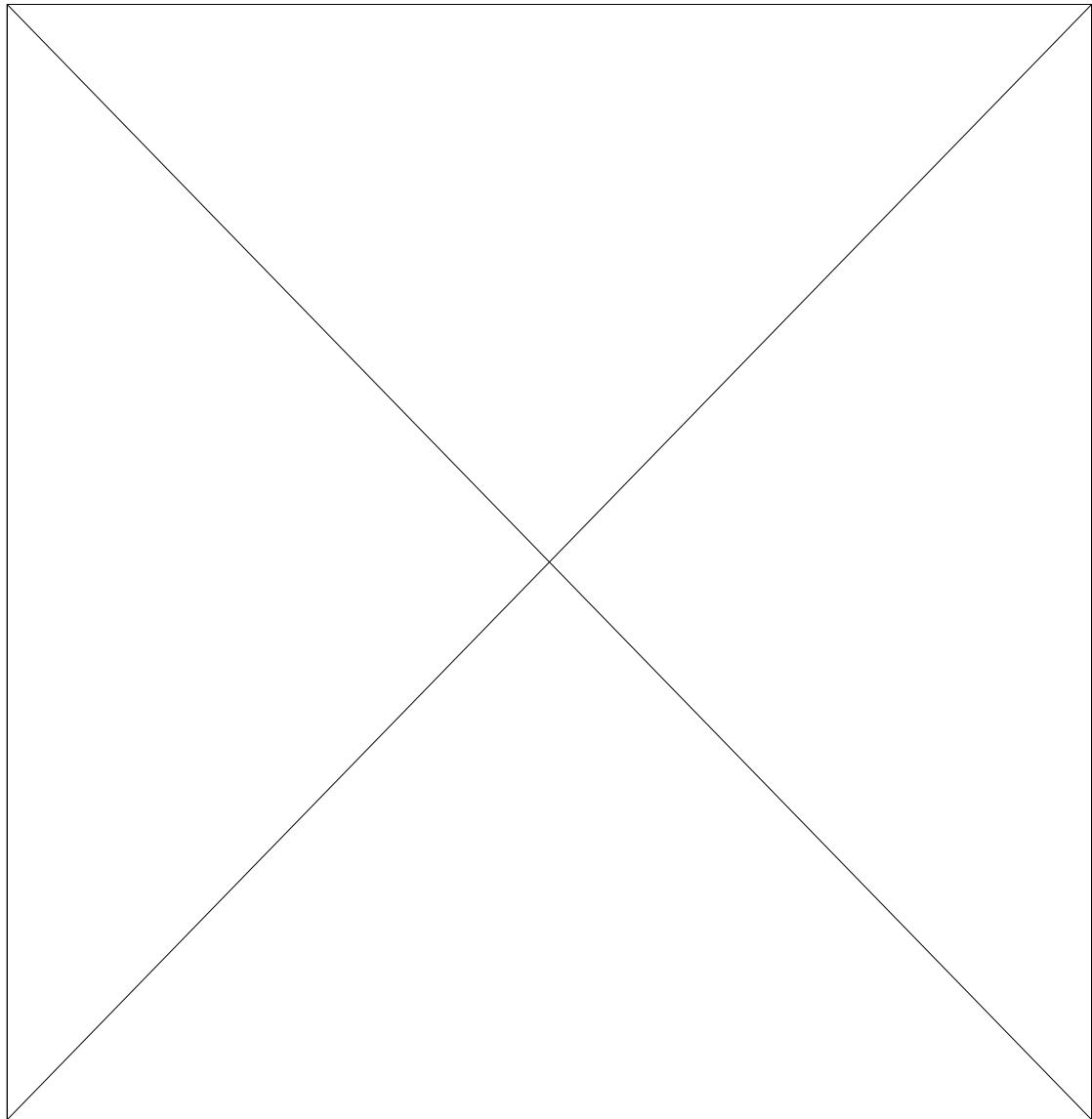
[L-9] ※ 다음은 뉴스의 일부입니다. 잘 듣고 물음에 답하세요.

[L-13] ※ 다음은 의사와 환자 간의 대화입니다. 잘 듣고 물음에 답하세요.

또한 가급적 긍정 발문의 비율이 부정 발문의 비율보다 높거나 같도록 발문을 구성하였다. 이는 긍정 발문을 제시함으로써 피험자가 문항을 직관적으로 이해할 수 있게 하기 위함이다. 한편 문항 개발 시 선지를 구성할 때는 가독성과 명확성을 중시하였다. 가독성을 위하여 ①~④ 선지가 가급적 길이순으로 배열될 수 있도록 설정하였고, 선지가 명확하게 제시될 수 있도록 지나치게 길어지는 것을 지양하였다. 1차 예비 조사 분석 결과 이후에는 정답률이 지나치게 높거나 변별도가 낮게 나온 문항의 경우 선지의 내용 및 제시 방식에 문제가 있을 수 있음을 고려하였다. 연구진 내부 회의 및 전문가 검토를 통해 지나치게 길이가 길거나 한 번에 이해하기 어려운 선지, 또 모호함이 있는 선지는 그 길이를 짧게 하거나 의미가 간명하게 전달될 수 있도록 표현을 수정하였다. 그리고 일부 선지는 매력도를 높이는 방향으로 내용을 수정하였다.

또한 제안서에서 문항 및 선지 구현 방식의 다양화를 제시한바, [L-13]과 같이 피험자가 영상 특히 현대인의 매체 생활에 많은 영향을 주는 유튜브 영상을 시청하면서⁶⁾ 문항을 풀 수 있도록 새로운 유형을 개발하였다.

6) 아래에 제시한 문항 예시는, 문항의 형태를 보여 주기 위해 유튜브 영상을 캡처하여 구성한 것이고 실제 문항에서는 유튜브 동영상을 시청하고 응답하는 형태로 구성되어 있다.



2) 듣기·말하기 영역

(1) 세부 조사 영역 및 요소 설정

① 조사 영역

2023년 국민의 국어능력 실태 조사는 조사 도구 정교화의 차원에서 구어 의사소통의 실제성을 강화하기 위해 듣기 영역과 말하기 영역을 연계한 듣기·말하기 영역을 신설하여 시행할 계획이다. 이는 듣기 영역과 말하기 영역을 별개의 문항으로 구분해 시행하였던 2018년도 조사와 변별되는 지점으로, 듣기, 말하기가 역동적이고 상호교섭적으로 이루어지는 일상 언어 생활을 조사에 반영하였다는 점에서 의미가 있다. 듣기·말하기 조사 영역은 2018년도 조사에서 말하기 영역으로 설계되었던 ‘설명하는 말하기’, ‘설득하는 말하기’ 조사 영역을 보완하고, 언어의 통합성 및 구어 담화의 역동성을 반영하기 위하여 일상

대화에 나타나는 ‘친교적 대화(듣기·말하기)’ 상황에 초점 맞추어 설계되었다. 구체적인 듣기·말하기 조사 영역은 아래와 같다.

<표 III-5> 2023년도 듣기·말하기(신설) 조사 영역

2023				
중영역	하위 영역	담화 유형	대화 상황	문항 수
가. 친교 표현하기	1. 배려하며 말하기	일대일 대화	개인적/공적	1
	2. 상호교섭하기	일대일 대화	개인적	1
합계				2

기본적으로 듣기·말하기 조사 영역은 2018년도 국민의 국어능력 실태 조사의 말하기 영역 평가 설계 기준을 준용해 ‘실제성’, ‘듣기·말하기 목적’, ‘담화 유형’, ‘대화 상황’과 같은 항목을 고려하고 있다. 이번 예비 조사는 일반 국민의 국어능력 실태 조사를 위해 기초 자료를 수집하는 첫 단계로, 물리적 시간과 비용이 투입되는 만큼 조사의 효율성이 담보되어야 한다. 다시 말해 듣기·말하기와 관련한 국민의 제반 언어생활 중에서 몇 가지 영역이나 장면을 선별해 조사를 시행해야 하므로 아래 네 가지 항목에 기초하여 조사 영역을 설계하는 것이 필요하다.

- 실제성
- 듣기·말하기 목적
- 담화 유형
- 대화 상황

상황 제시형으로 수행되었던 기존 2018년도 말하기 영역 조사를 고도화한다는 기초 아래 듣기·말하기 영역은 담화 맥락을 ‘듣기’ 자료로 구체화하여 제시된 의사소통 상황에 적절히 대처하고 응답할 수 있도록 구성되었다. “말하기, 듣기에서의 구어 사용 능력(Wilkinson, 1965)” 혹은 “통합적 구두 의사소통 능력(이창덕, 2008)”으로 간주되는 언식성(oracy)의 개념에서도 알 수 있듯 구어 의사소통 상황은 듣기와 말하기의 통합 양식을 전제하고 있다. 또한 “소리는 인간을 실제 한가운데 그리고 동시성 속에 놓이게 하여 지금 여기의 경험을 전달하고 상황을 담아(Ong, 1967: 이동후, 2016 재인용)” 낼 수 있기 때문에 본 연구에서는 듣기와 말하기의 통합형 문항을 개발하여 조사의 실제성을 확보하고자 하였다. 즉 피험자로 하여금 구어 의사소통에서의 청자와 담화 맥락을 텍스트 형태로 제시하여 상황을 ‘가정’하게 하는 것이 아니라 피험자에게 ‘실제’ 음성을 들려주는 방법으로 말하기의 상황을 현실감 있게 제시해 주고자 하였다.

한편 듣기·말하기의 목적, 담화 유형, 대화 상황은 각각 ‘친교’, ‘(일대일) 대화’, ‘개인적 상황/공적 상황’으로 설정하였다. 일반적으로 말하기의 목적은 친교, 정서 표현, 설명, 설득으로 구분되는데 2018년도 실태 조사에서는 발화 순서 교대 등과 같은 변인 통제의 어려움을 들어 친교적 목적의 말하기를 조사 영역에서 제외한 바 있다. 결과적으로 2018년도 조사에서는 화자와 청자의 상호 작용을 중심으로 이루어지는 친교 목적의 의사소통을 다루지 못하였으며, 본 연구는 이러한 점을 보강하고자 듣기·말하기 통합 문항에 ‘친교 표현하기’ 영역을 새롭게 설정하였다. ‘친교’란 “언어 사용자의 접촉에 초점을 맞추어 화자가 청자에게 자신의 생각과 감정을 우호적으로 나타내는 등 듣기·말하기의 고유한 상호작용적 특성(서현석,

2007)”)이 잘 드러나는 의사소통 목적을 가리킨다. 그리고 일상 속 ‘대화’는 이러한 친교 적 목적에 가장 적합한 담화 유형에 해당한다. 이러한 점을 고려해 본 연구는 ‘친교 표현하기’를 일대일 개인적/공적 대화의 상황으로 구성하였다.

아울러 친교 목적의 의사소통에서 ‘배려’와 ‘상호교섭’이 중요한 요소로 다루어지고 있는 점을 감안하여 이를 하위 조사 영역으로 설정하였다. ‘배려하며 말하기’는 상대의 말을 경청하고 상대의 입장을 헤아려 공손하게 말하는 청자 지향적 말하기의 하나이다. 이는 관계를 이루며 살아가는 사회 속에서 반드시 요구되어야 할 의사소통 능력 중 하나로, 언어생활의 기본적인 예절이자 도리라고도 볼 수 있다. 그리고 ‘상호교섭하기’는 화자와 청자가 협력적, 역동적으로 의미를 구성해 가는 과정을 뜻하는 말로, 구어 의사소통의 기본적인 속성에 해당한다. 상호교섭적 관점은 통시적 흐름 속에서 화자와 청자의 역할이 교차한다는 생각에 바탕하여 화자와 청자의 분리된 역할 개념이 아닌 ‘참여자’로서 두 역할을 동시에 수행하는 언어 사용자의 존재를 상정한다(Barnland, 1970; 서영진, 2012 재인용). 요지는 두 하위 조사 영역 모두 일방향적 말하기가 아닌 듣기 활동이 함께 수행되는 쌍방향적 의사소통을 전제하고 있다는 점이다. 이에 본 연구는 상황 맥락에 맞게 상대의 입장과 처지를 고려하거나 상대의 요청이나 반응에 적절히 대응하는 말하기가 수행될 수 있는 조사 문항을 개발하고자 하였다.

이상의 논의에 따라 설계한 듣기·말하기 영역 예비 조사 문항 정보는 아래와 같다.

<표 III-6> 2022년 제1차 예비 조사 듣기·말하기 영역의 문항 정보

문항 번호	문항 유형	친교 표현하기		배점
		배려하며 말하기	상호교섭하기	
LS-1	가	○		30
LS-1	나	●		30
LS-2	가		○	30
LS-2	나		○	30
합계		2	2	120

※ ○: 개인적 상황, ●: 공적 상황

② 조사 형식

앞서 ‘① 조사 영역’에서 밝힌 바와 같이 듣기·말하기 조사 영역은 말하기 영역과 마찬가지로 수행형(주관식) 문항으로 구성되어 있다. 우선 각 문항은 문자 언어를 사용한 상황 제시형이 아닌 음성 언어를 활용한 상황 제시형으로 개발되었다. 구어 의사소통의 상황 맥락을 사실적이고 현실감 있게 제시하기 위하여 대화 상황을 하나의 문장으로 기술하는 것이 아니라 가상의 인물과 상황을 설정하여 ‘대화 형식’으로 맥락을 파악할 수 있게 구성하였으며, 듣기 조사 영역에서의 음성 자료처럼 두 인물의 대화가 이어지는 가운데 피험자가 대화 상황에 적합한 응답을 할 수 있도록 하는 과제를 부여하였다. 이는 음성 자료로 제시한 대화 상황에 피험자가 ‘반응’할 수 있게 설계한 것으로, 2018년도 조사에서 지적됐던 ‘친교 표현하기’ 문항 구현의 현실적 한계를 보완하기 위한 하나의 방법으로 제안되었다.

결과적으로 듣기·말하기 영역 또한 후술하게 될 말하기 영역과 동일한 방식으로 조사와 채점이 이루어

지게 되었다. ‘친교 표현하기’에서 나타나는 말 차례 교환을 음성 자료로 구현하였기 때문에 대화의 마지막 발화를 피험자가 ‘혼자 말하기’ 방식으로 응답할 수 있도록 구성하였고, 녹음된 응답은 다수의 채점자가 교차 채점하는 방식으로 진행되었다.⁷⁾ 아울러 채점 기준은 문항별로 내용, 조직, 표현, 전달이라는 조사 항목을 일관되게 적용하였다. 다만 ‘내용’ 조사 항목의 경우 ‘친교 표현하기’의 하위 조사 영역에 따라 서로 다른 채점 요소를 설정하였다. 문항 유형(가, 나)에 따라 구체적인 기술이 다르게 나타나지만 공통적으로 ‘배려하며 말하기’는 상대의 입장이거나 처지를 고려하여 말하는 청자 지향적 말하기에, ‘상호교섭하기’는 상대에게 적절하게 반응하는 말하기에 초점 맞추어져 있다. 구체적인 조사 요소는 다음과 같다.

<표 III-7> 친교 표현하기(배려하며 말하기) 조사 요소 ①

조사 항목			채점 요소
내용 구성	내용 선정	가	• 문제에서 요구하는 응답을 한다. (배려) • 상대의 부담을 최소화하고, 합리적인 대안을 제시한다.
		나	• 문제에서 요구하는 응답을 한다. (사과) • 잘못된 내용에 대해 언급하고, 상대의 기분을 헤아려 말한다.
	조직	가	• 응답의 흐름이 배려를 나타내는 데 적합하다. • 발화의 시작과 마무리가 자연스럽다.
		나	• 응답의 흐름이 사과하는 데 적합하다. • 발화의 시작과 마무리가 자연스럽다.
표현 및 전달	표현		• 적절한 어휘를 사용한다. • 어법에 맞게 말한다.
	전달		• 발음, 속도, 어조 등이 자연스럽다(유창성). • 발음이 명료하다(정확성).

7) ‘혼자 말하기’ 수행 방식과 채점 방식에 대해서는 말하기 영역에서 자세히 논하고자 한다.

<표 III-8> 친교 표현하기(상호교섭하기) 조사 요소 ②

조사 항목			채점 요소
내용 구성	내용 선정	가	• 문제에 부합하는 적절한 응답을 한다. • 문제에 대한 설명과 해결을 위한 반응이 잘 드러난다(적극성).
		나	• 과제에서 요구하는 응답을 한다. (거절) • 거절의 이유를 맥락에 맞게 설명한다.
	조직		• 응답이 이전 발화의 흐름에 잘 대응된다. • 발화의 시작과 마무리가 자연스럽다.
표현 및 전달	표현		• 적절한 어휘를 사용한다. • 어법에 맞게 말한다.
	전달		• 발음, 속도, 어조 등이 자연스럽다(유창성). • 발음이 명료하다(정확성).

(2) 문항 형식

듣기·말하기 영역은 언어의 통합성 및 구어 담화의 역동성을 고려하여 듣기와 말하기를 연계하여 문항을 구성하고자 하였다. 다만 시행상의 어려움 등을 감안하여, 문항의 구조를 단순화하고 문항 자체의 난도가 듣기나 말하기 능력의 측정에 미치는 영향을 최소화하기 위해, 일상적 대화 상황에서 청자와 화자의 역할 수행을 통해 듣기·말하기 능력을 평가할 수 있는 문항으로 설계하였다.

‘LS-1. 배려하며 말하기’는 문제 상황이 발생하는 대화를 들려주고, 대화의 마지막에 이어질 말을 응답하도록 하였다. ‘LS-2. 상호교섭하기’는 상대방에게 요청하는 대화를 들려주고 요청에 대해 적절히 응답하도록 하였다. 녹음된 대화를 들려주고 피험자가 대화 참여자의 입장에서 응답을 하도록 구성함으로써 대화 상황의 실제성을 높이고자 하였다. 구체적인 듣기·말하기 문항의 형식은 <표 III-9>와 같다.

<표 III-9> 듣기·말하기 문항의 형식

LS-1 (배려)	[발문] 다음은 (대화 참여자들) 간의 대화입니다. 잘 듣고 (대화 참여자1)의 입장에서 (대화 참여자 2)를 (조사영역) 해 주세요. [조건] 준비 시간 5초, 응답 시간 20초 [듣기 자료] 대화 참여자들의 대화(음성 녹음)
LS-2 (상호 교섭)	[발문] 다음은 (대화 참여자들) 간의 대화입니다. 잘 듣고 (문제상황)에 대해 (조사영역) 해 주세요. [조건] 준비 시간 5초, 응답 시간 30초 [상황] 대화 참여자들의 대화(음성 녹음)

3) 말하기 영역

(1) 세부 조사 영역 및 요소 설정

① 조사 영역

2023년 국민의 국어능력 실태 조사의 말하기 영역 평가는 2018년에 수행한 국민의 국어능력 평가와의 연계 속에서 이루어져야 한다. 기본적으로 2023년도 말하기 조사 문항은 2018년도 말하기 영역의 조사 영역에 준용하여 개발하되 디지털 기술의 발달과 같은 언어 사용을 둘러싼 매체 환경의 변화를 반영하여 조사 영역과 요소를 수정·보완하였다. 2023년 국민의 국어능력 실태 조사의 말하기 영역은 2018년 조사 영역에 기반을 두고 있지만 사회적·시대적 요구에 맞게 조사 영역을 일부 조정하였으며 2022년도 예비 조사와 2018년도 조사의 조사 영역을 비교한 구체적인 내용을 제시하면 아래와 같다.

<표 III-10> 2018년도 조사 영역과 2022년도 조사 영역의 비교

2018			2022		
중영역	하위 영역	문항 수	중영역	하위 영역	문항 수
가. 설득하는 말하기	1. 설득(주장하는 말)	1	가. 설득하는 말하기	1. 설득(주장하는 말)	2
	2. 설득(요청하는 말)	1			
나. 설명하는 말하기	1. 설명(경험 또는 개념을 설명하는 말)	1	나. 설명하는 말하기	1. 설명(경험 또는 개념을 설명하는 말)	1
				2. 설명(프레젠테이션)	2
합계		3	합계		5

<표 III-10>에 제시된 바와 같이 2022년도 예비 조사 말하기 조사 영역은 기존 2018년도 조사 영역 ‘가-2. 설득(요청하는 말)’을 제하고 ‘나-1. 설명(경험 또는 개념을 설명하는 말)’에 ‘설명(프레젠테이션)’을 추가하는 형태로 수정되었다. 그리고 ‘설득하는 말하기’의 전형에 가장 가까운 ‘주장하는 말’ 조사 영역을 동일하게 이어갔다. 단, (3) 문항 형식에 언급해 두었듯 상황 및 조건을 줄글로 제시하는 것이 아니라 대화 형식의 대본을 제시하여 말하기의 실제성을 확보하고자 하였다. 그리고 ‘나. 설명하는 말하기’ 영역에 ‘프레젠테이션’을 추가 구성하였다. 전통적인 말하기 능력에 더하여 매체를 이용한 말하기 능력이 요구되고 있는 지금, 디지털 리터러시 관련 역량을 측정하기 위한 하나의 방법으로 발표용 PPT 자료를 활용한 설명하는 말하기를 제안하였다. 아울러 피험자의 조사 응시 피로도를 줄이기 위해 기존의 문항 수(3개)를 그대로 유지하고자 하였으며 2022년도 예비 조사는 그 1.5배수인 5개의 문항으로 구성하였다.

<표 III-11> 2022년도 예비 조사 말하기 조사 영역

2022				
중영역	하위 영역	담화 유형	대화 상황	문항 수
가. 설득하는 말하기	1. 설득(주장하는 말)	토론	공적	2
나. 설명하는 말하기	1. 설명(경험 또는 개념을 설명하는 말)	일대일 대화	개인적	1
	2. 설명(프레젠테이션)	발표	공적	2
합계				5

조사 영역별 담화 유형은 각각 대화, 토론, 발표(강연) 세 가지로 설정하였다. 담화 유형은 말하기의 목적과 실제 언어생활에서 차지하는 비중, 사회적 중요성 등을 두루 고려한 결과이다. 일상생활 속에서 특정한 청자를 대상으로 청자의 나이나 지위 등을 고려해 말하는 일대일 대화 상황이 빈번하게 나타난다는 점, 생각과 주장을 논리적으로 표현하는 능력에 대한 사회적 기대와 요구가 높다는 점, 학업 및 직무 수행 과정에서 매체를 활용해 여러 정보를 효과적으로 전달하는 능력이 중요해지고 있다는 점에서 ‘설득하는 말하기’, ‘설명하는 말하기’ 능력을 측정하는 데 대화, 토론, 발표 세 담화 유형이 적합하다고 판단하였다. 그리고 ‘경험 또는 개념을 설명하는 말’은 개인적 상황으로, ‘프레젠테이션’과 ‘주장하는 말’은 공적 상황으로 구성하여 말하기 상황을 균형 있게 안배하였다. 이상의 내용을 정리하면 다음과 같다.

<표 III-12> 2022년 제1차 예비 조사 말하기 영역의 문항 정보

문항 번호	문항 유형	설명적 말하기		설득적 말하기	배점
		경험 또는 개념을 설명하는 말	프레젠테이션	주장하는 말	
S-1	가교	○			50
S-2	가		●		50
S-2	나		●		50
S-3	가			●	50
S-3	나			●	50
합계		1	2	2	250

※ ○: 개인적 상황, ●: 공적 상황

② 조사 형식

말하기 평가는 수행 능력을 중심으로 이루어져야 하며 과제 기반의 평가로 설계됨으로써 피험자의 실제적 말하기 능력을 측정할 수 있어야 한다(김평원, 2007). 이에 따라 본 연구는 피험자의 실제 말하기

수행을 녹음하고, 녹음한 결과를 다수의 채점자가 교차 채점하는 방식으로 피험자의 말하기 능력을 조사하였다. 또한 특정한 말하기 상황과 주제를 제시함으로써 피험자가 상황을 이해하고 과제를 해결하기 위한 말하기를 수행할 수 있도록 문항을 개발하였다. 이와 함께 말하기 조사의 쟁점들을 고려하여 신뢰할 만하고 유용한 말하기 조사가 이루어질 수 있도록 조사 형식을 설정하였다.⁸⁾

말하기 조사는 피험자의 수와 구성에 따라 ‘혼자 말하기’ 방식과 ‘함께 말하기’ 방식으로 이루어질 수 있다. 함께 말하기 방식의 경우 의사소통 참여자의 실제 상호 작용 양상을 확인할 수 있다는 장점이 있지만 시행상의 어려움이 있어 말하기 평가는 대체로 ‘혼자 말하기’ 방식으로 이루어진다(민병곤, 정재미, 박재현, 2014). 혼자 말하기 방식은 조사 상황에서 한 명의 피험자의 말하기 수행에 온전히 집중하면서 말하기와 관련한 피험자의 제반 능력을 측정할 수 있다는 장점이 있지만 상호 작용 능력을 파악하는 데에는 분명 한계를 갖고 있는 것이 사실이다. 하지만 이는 구체적인 청자를 상정하고 말하기 맥락을 충분히 제공하는 것을 통해 보완될 수 있다. 이러한 이유로 2013년도 국민의 국어능력 평가(1차 조사)를 포함해 2018년도 실태 조사(2차 조사)에서도 말하기 영역의 조사는 가상의 청자를 상정한 ‘혼자 말하기’ 형식을 취하고 있다(김종철 외, 2013; 민병곤 외, 2018).

한편, 말하기 조사의 주요 쟁점 중 하나는 조사의 신뢰도와 관련되어 있다. 실제 수행 능력을 평가의 대상으로 삼아야 하는 말하기 평가의 특성상, 평가 결과를 수치화하기 어려운 측면이 있고 채점 시에 채점자의 주관의 개입될 수 있는 난점이 있어 채점자 내 신뢰도와 채점자 간 신뢰도를 완벽하게 보장하기 어려운 면이 있다(민병곤, 박재현, 2015). 이에 따라 신뢰로운 조사가 이루어지기 위해서는 명확한 채점 기준을 설정하고 채점자의 주관의 개입에 영향을 주지 않을 수 있도록 채점 방식을 설계해야 한다. 일반적으로 말하기 평가의 채점은 피험자의 말하기를 종합적으로 평가하는 ‘총체적 채점’ 방식과 독립된 평가 요소에 따라 피험자의 말하기 능력을 분석적으로 평가하는 ‘분석적 채점’ 방식으로 이루어진다. 총체적 채점 방식은 종합적인 평가가 이루어질 수 있고 채점이 비교적 용이하다는 장점이 있지만 채점자의 주관의 개입할 수 있다는 단점이 있다. 분석적 채점 방식은 채점 기준에 따라 채점을 진행하므로 신뢰로운 채점이 가능하다는 장점이 있지만 채점자의 인지적 부담이 크고 개별 기준에 따른 채점이 말하기 수행 능력 담보하지 못할 수 있다는 단점이 존재한다. 하지만 이번 2023년도 국민의 국어능력 실태 조사가 대규모로 진행되는 영역별(화법) 분할 조사라는 점에서 일관된 채점 기준을 적용하여 신뢰성 있는 결과를 도출하여야 하기 때문에 보다 객관적인 접근이 가능한 분석적 채점 방식을 선택·적용하였다.

말하기 영역 채점 기준은 듣기·말하기 영역과 마찬가지로 담화 유형별로 서로 다르게 설정하였다. 이는 설득적 말하기와 설명적 말하기의 말하기 특성이 서로 다른바, 유형별로 상이한 채점 기준을 적용해야 한다는(조재윤, 2007) 2018년도 조사의 논의를 따른 것이다. 이와 관련된 세부적인 조사 요소는 <표 Ⅲ-13>, <표 Ⅲ-14>와 같다.

8) 조사 형식 설계 시 박재현(2015)에서 제시한 화법 평가의 쟁점을 중점적으로 고려하였다.

<표 III-13> 설명적 말하기(프레젠테이션) 조사 요소 ①

조사 항목		채점 요소
내용 구성	내용 선정	- 슬라이드의 내용을 충분하게 설명한다. - 복잡한 내용을 간명하게 정리하여 말한다.
	조직	- 발표의 흐름이 자연스럽다. - 발화의 시작과 마무리가 효과적이다.
표현 및 전달	표현	- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.
	전달	- 발음, 속도, 어조 등이 자연스럽다(유창성). - 발음이 명료하다(정확성).

<표 III-14> 설득적 말하기(주장하는 말) 조사 요소 ②

조사 항목		채점 요소
내용 구성	내용 선정	- 문제에서 요구하는 응답을 한다. - 문제 상황에 대한 설명과 주장의 근거가 잘 드러난다.
	조직	- 내용의 흐름이 논증에 적합하다. - 발화의 시작과 마무리가 유기적이다.
표현 및 전달	표현	- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.
	전달	- 발음, 속도, 어조 등이 자연스럽다(유창성). - 적절한 시간을 지켜 말한다.

(2) 가교 문항 검토

2022년도 예비 조사에서는 2018년도 조사와의 동등화를 위해 2018년도 말하기 실태 조사의 분석 결과에 근거하여 곤란도와 변별도가 적절하고 타당하다고 판별된 문항을 가교 문항으로 설정하였다. 가교 문항은 2023년도에서 사용할 전체 문항 수의 20%이며, 변별도와 곤란도를 고려하여 최종적으로 2018년도 말하기 영역 2번 문항을 가교 문항으로 채택하였다.

가교 문항은 ‘설명적 말하기(경험 또는 개념을 설명하는 말)’ 조사 영역에 해당하는 문항으로 일대일 대화 상황에서 이루어지는 개인적 말하기로 설계되었다. 이 문항은 응급 상황 혹은 위기 상황에서 대처하는 방법을 어린 학생에게 설명하는 문항이다. 피험자의 피로도를 고려해 문항 구조를 단순화하고자 한 문항 구성 의도에 부합할 뿐 아니라 하위 영역(경험 또는 개념을 설명하는 말)이 설명적 말하기의 가장 전형적인 형태를 띠고 있다는 점에서 가교 문항으로 적합하다고 보았다. 수치상으로도 2018년도 말하기 문항의 전체 평균 곤란도가 0.604, 위 가교 문항의 곤란도가 0.604로 동일하게 나타나 보통 수준의 적절한 난도를 보였으며, 조사 항목별로 도출한 변별도 또한 0.761~0.866 사이에 분포하여 비교적 양호한 수준으로 나타났다.

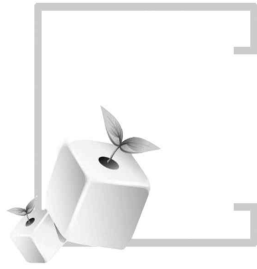
말하기 1번 문항 (가교)												
2022 국민의 국어능력 실태 조사			예비 조사		S-1		검사지		가, 나			
			개념 또는 경험을 설명하는 말		일대일 대화 (사적)				가교 문항			
조사 문항												
자료 출처												
해설			평가 항목		채점 요소	채점 척도(점수)					점수 비중	
						매우 미흡	미흡	보통	우수	매우 우수		
			내용 구성	내용 선정	- 질문에서 요구하는 대답을 한다. - 청자가 이해하기 쉽게 구체적으로 설명한다.	3	6	9	12	15	30 %	60 %
				조직	- 내용의 흐름이 설명에 적합하다. - 시작과 마무리가 자연스럽다.	3	6	9	12	15	30 %	
			표현 및 전달	표현	- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.	2	4	6	8	10	20 %	40 %
				전달	- 발음, 속도, 어조가 적절하다. - 적절한 시간을 지켜 말한다.	2	4	6	8	10	20 %	
본 조사 결과			본 조사 결과 (고전 검사 이론)	변별도	내용 선정	0.762	모든 영역 변별도가 양호한 수준으로 나타남. 그중 조직 영역의 변별도는 매우 양호한 수준을 보임.					
					조직	0.866						
					표현	0.761						
					전달	0.777						
				곤란도	내용 선정	0.593	모든 영역에서 곤란도가 보통 수준으로 나타남.					
					조직	0.561						
					표현	0.639						
					전달	0.597						
			본 조사 결과 (문항 반응 이론)	변별도	내용 선정	0.897	모든 영역에서 변별도가 적절한 수준 이상으로 나타남. 그중 전달 영역의 변별도는 높은 수준을 보임.					
					조직	1.077						
					표현	0.740						
					전달	1.523						
				곤란도	내용 선정	-0.066	모든 영역에서 곤란도가 보통 수준으로 나타남.					
					조직	0.068						
					표현	0.119						
					전달	0.016						

(3) 문항 형식

가교 문항 ‘S-1. 설명하는 말하기(경험 또는 개념을 설명하는 말)’는 2018년도 조사에서 시행한 문항으로 구체적인 ‘청자’와 설명이 필요한 ‘상황’을 글상자에 제시하는 형식을 취하였다. ‘S-2. 설명하는 말하기(프레젠테이션)’는 발표에 활용할 PPT 슬라이드를 문항 하단에 배치하고, 해당 자료를 활용하여 말할 내용을 구성할 수 있게 하였다. 마지막 ‘S-3. 설득하는 말하기(주장하는 말)’에서는 논제를 단순 기술하지 않고 토론회 사회자의 발언을 대본 형식을 빌려 제시함으로써 말하기 상황의 사실성을 높이고자 하였다. 구체적인 말하기 문항의 형식은 <표 III-15>와 같다.

<표 III-15> 말하기 문항의 형식

S-1 (설명)	[발문] 아래의 상황 중 하나를 골라 (청자)에게 (조사 영역)해 주세요. [상황] ~한 상황에서 ~하는 방법 [조건] 2분 정도 자연스럽게 자신의 의견을 말해 주십시오. 준비 시간은 1분입니다.
S-2 (설명)	[발문] (청자)를 대상으로 (담화 유형)하는 상황을 가정하고, (주제)에 대한 (매체)에 따라 (조사 영역)해 주세요. [상황] 대중 강연 [조건] 준비 시간 2분, 발표 시간 2분 내외 [자료] PPT 슬라이드
S-3 (설득)	[발문] (논제)를 주제로 한 토론회에 청중으로 참여하였습니다. 찬성과 반대 입장 중 하나를 선택하고 그에 대한 근거를 들어 (조사 영역)해 주세요. [조건] 준비 시간 1분, 발표 시간 2분 내외 토론회 사회자 발언



제 4 장

1차 예비 조사 시행 및 결과 분석



1. 1차 예비 조사 설계

1) 1차 예비 조사를 위한 검사지 구성

(1) 영역별 검사지 구성

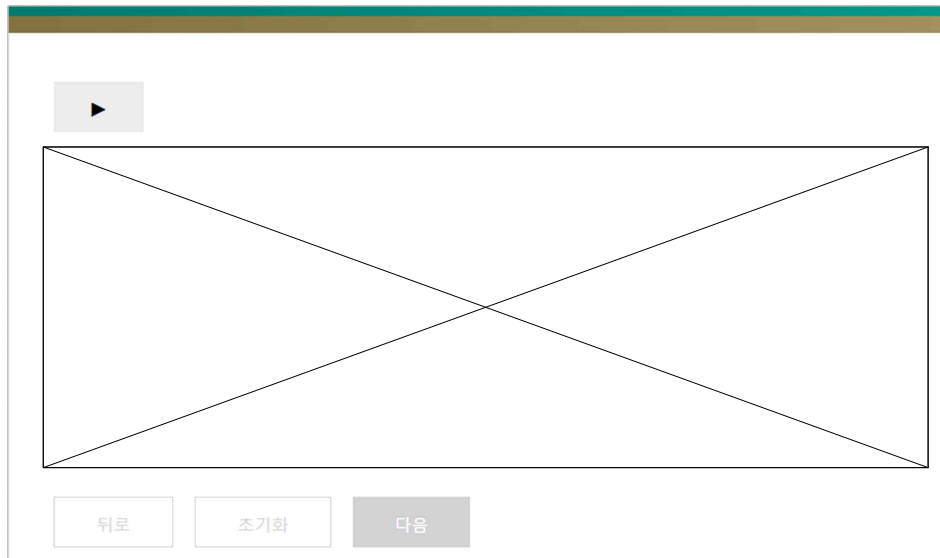
검사지 영역은 듣기, 말하기, 듣기·말하기 총 3가지 영역이며, 각 영역은 (가)형과 (나)형으로 구분된다. 영역별 2가지 유형((가)형, (나)형)을 조합하여 아래와 같이 검사지를 총 8세트로 구성하였다.

<표 IV-1> 영역별 검사지 구성

검사지 set (Booklet)	듣기 가 13문항 (신규13)	듣기 나 13문항 (신규13)	말하기가 2문항 (신규2)	말하кина 2문항 (신규2)	듣말 가 2문항 (신규2)	듣말 나 2문항 (신규2)	듣기 말하기 공통 영역 3문항 (듣기 가교2, 말하 기 가교1)	문항 수
검사지1	포함		포함		포함		포함	20
검사지2	포함		포함			포함	포함	20
검사지3	포함			포함	포함		포함	20
검사지4	포함			포함		포함	포함	20
검사지5		포함	포함		포함		포함	20
검사지6		포함	포함			포함	포함	20
검사지7		포함		포함	포함		포함	20
검사지8		포함		포함		포함	포함	20
검사지 문항 구성				검사 시간 및 인원				
• 듣기 문항: 총 15문항(신규 13문항, 가교 2문항) • 말하기 문항: 총 3문항(신규 2문항, 가교 1문항) • 듣기·말하기 문항: 총 2문항(신규 2문항)				• 검사 시간: 총 50분 * 듣기 15문항 × 2분 = 30분 ** 말하기 3문항 × 5분 = 15분 *** 듣기·말하기 2문항 × 2.5분 = 5분 • 검사 인원: 240명(검사지 8종 × 30명 = 240명)				

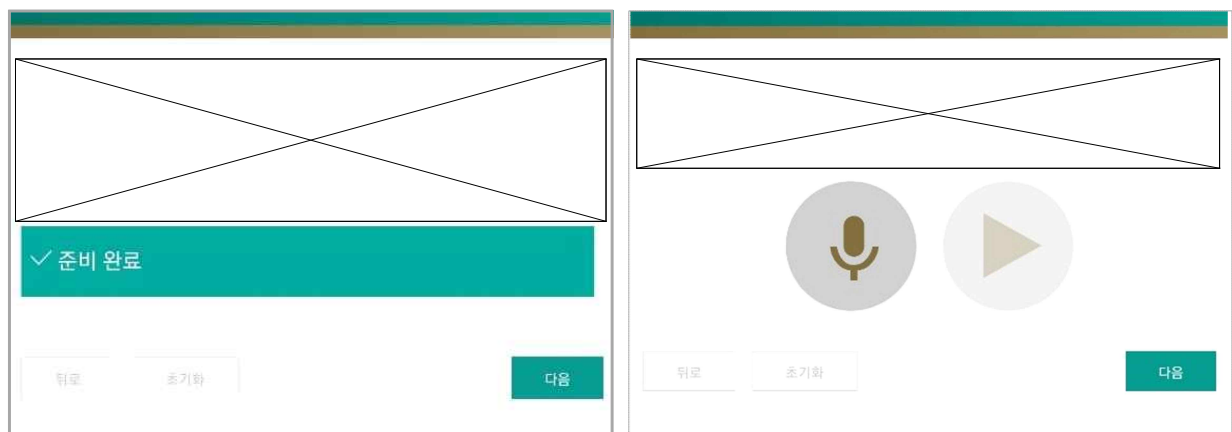
(2) 영역별 검사 문항의 실제

영역별 검사 문항의 경우 ‘문두’ 및 ‘선지’는 화면에 읽을 수 있는 형태로 제공하였고 ‘지문’과 ‘문두’ 까지 음성 녹음 자료로 제공하였다. 아래 그림은 듣기 영역의 예시이며 조사원이 지문 플레이를 누르면 문두와 함께 지문이 재생되고 피조사자는 문두 및 선지를 읽은 후 답안을 제출한다.



[그림 IV-1] TAPI 화면 예시 - 듣기 영역

듣기·말하기와 말하기 영역도 듣기 영역과 마찬가지로 ‘문두’ 및 ‘선지’는 화면에 읽을 수 있는 형태로 제공하였고 ‘지문’과 ‘문두’ 모두 음성 녹음 자료로 제공하였다. 조사원이 지문 플레이를 누르면 문두와 함께 지문이 재생되고 피조사자는 말하기를 위한 준비 시간을 가진 뒤 녹음 버튼을 눌러 말하기 답안을 제출한다.



[그림 IV-2] TAPI 화면 예시 - 듣기·말하기, 말하기 영역 공통

2) 1차 예비 조사의 설계

(1) 모집단 및 표본 수

1차 예비 조사의 모집단은 서울, 경기, 인천에 거주하는 만 20세~69세의 남녀로 정의하였다. 본 조사

의 목적은 개인 차원의 행동과 태도를 측정하는 것이므로, 조사 기본 단위(elementary unit: 조사의 대상이 되는 가장 최소의 요소)는 개인으로 정의된다.

1차 예비 조사의 목표 표본 수는 조사 목적과 조사 비용, 기간 등을 고려해 총 240명으로 결정하였다. 240명 표본의 오차 한계는 단순확률표본을 기준으로 95% 신뢰 수준에서 $\pm 6.3\%$ 로 계산된다.

(2) 표본 수 배분

1차 예비 조사는 본 조사에서 사용할 문항의 2배수 분량을 조사하기 때문에, 모든 응답자에게 전체 문항을 조사한다면 응답자의 피로감이 높아질 가능성이 매우 크다. 이로 인해 각 영역별로 (가)형과 (나)형을 구분하였고, 검사지 유형별 표본 수는 동일하게 배분하였다.

(3) 모집단 층화와 표본 수 배분

층화추출은 모집단을 서로 겹치지 않은 그룹들, 즉 층(strata)들로 나누고 각 층에서 표본을 뽑는 방법이다. 표본의 크기가 같은 경우, 층화추출은 단순 확률 추출을 통해 얻은 표본보다 추정 오차의 한계를 줄일 수 있고, 특히 층 내의 관측 값들이 동질적인 경우에 더욱 효과적인 추출법으로 알려져 있다. 이 밖에도 모집단 조사 단위들을 서로 편리한 그룹으로 묶음으로써 조사 시 비용을 절감할 수 있고, 모집단 내의 하위 그룹들에 대한 모수 추정 값을 따로 구할 수 있다는 장점이 있다.

각 층을 나누는 기준은 연구 목적에 따라 여러 가지를 고려할 수 있지만, 광범위한 지역을 대상으로 이루어지는 조사에서는 행정 구역이나 지리적 위치를 층화의 1차적 기준으로 삼는 경우가 많다. 이 연구에서는 전체 모집단을 서울, 인천, 경기도 읍면 지역, 경기도 동 지역이라는 4개의 층으로 구분하고, 각 층에서 표본을 추출하는 층화추출을 진행했다.

다음으로 4개 권역으로 구분된 각 층별로 몇 명의 표본을 조사해야 하는지를 결정하였다. 이번 예비 조사는 모집단 전체의 국어능력을 추정하려는 목적보다는 다양한 인구학적 특성을 지닌 표본을 추출해 국어능력 조사지의 타당성과 신뢰도를 검토하기 위한 목적에서 진행됐다. 이러한 조사 목적을 감안해 각 층별로 최대한 같은 크기의 표본 수를 배분하고자 했다.

(4) 집락(읍·면·동) 추출과 응답자 추출

각 층별 표본 수를 결정한 후에는 각 층을 통계적으로 대표할 수 있도록 표본을 추출하였다. 이때에는 2단계 집락추출법(cluster sampling)을 사용하였다. 통계적 개념에서의 집락(cluster)은 서로 인접해 있는 개별 조사 단위들이 모여 있는 집단을 의미한다. 그리고 집락추출은 먼저 조사를 진행할 집락을 선정한 후에 그 집락에서 표본을 추출해서 조사를 진행하는 방식이다. 이때 집락에 속한 기본 단위를 전수 조사한다면 1단계 집락추출이 되지만, 일부 표본을 다시 추출해서 조사한다면 2단계 집락추출이 되는 것이다.

이 조사에서는 행정 구역상의 읍·면·동을 집락으로 간주하고, 2단계 집락추출을 진행하였다. 간단히 정리하자면 각 층에서 통계적인 방법을 통해 읍·면·동을 먼저 추출한 후에, 그 지역에서 설문 조사에 참여할 개인을 추출한 것이다.

각 층 내에서는 크기비례확률추출(Sampling with probabilities proportional to size)을 통해 읍·면·동을 선정하였다. 크기비례확률추출은 규모가 큰 집락일수록 표본으로 선정될 확률이 비례해서 높아지

게 된다. 특히 집락들의 규모가 심하게 차이가 나거나 대규모 집락이 전체 모집단에 미치는 영향이 클 경우에는 각 집락의 크기에 비례하여 뽑는 방법인 크기비례확률추출법이 널리 사용되고 있다. 이렇게 크기비례확률추출을 통해 집락을 추출할 경우에는 각 집락에서 동일한 수의 표본을 추출하는 것을 권장한다. 이렇게 하면 각 표본별로 추출 확률이 같아지는 자체가중표본(self-weighting sample)을 얻을 수 있다. 자체가중표본을 활용하면 결과 분석 과정에서 보다 간편하고 경제적인 작업이 가능하다. 각 층별로는 추가적인 가중치 보정 없는 단순 평균값으로도 결과를 산출할 수 있기 때문이다. 이 조사에서는 여러 가지 여건을 감안해 하나의 집락에서 조사되는 표본 수를 10명으로 결정하였다. 따라서 목표 표본 수 240명을 조사하기 위한 집락 수는 총 24개로 설계하였다. 이러한 과정을 거쳐 최종적으로 추출한 집락과 집락별 표본 수는 다음과 같다.

<표 IV-2> 1차 예비 조사 집락(읍·면·동) 추출 결과

층 구분	집락(읍·면·동)	목표 표본 수	조사 완료 표본 수
서울	서울특별시 동대문구 용신동	10	10
	서울특별시 도봉구 도봉제2동	10	10
	서울특별시 마포구 서강동	10	10
	서울특별시 노원구 월계2동	10	10
	서울특별시 서초구 서초2동	10	10
	서울특별시 강동구 상일제1동	10	10
인천	인천광역시 미추홀구 도화1동	10	10
	인천광역시 연수구 연수1동	10	10
	인천광역시 남동구 서창2동	10	10
	인천광역시 부평구 삼산1동	10	10
	인천광역시 서구 검암경서동	10	10
	인천광역시 미추홀구 송의1,3동	10	10
경기 동 지역	경기도 장안구 정자1동	10	10
	경기도 분당구 판교동	10	10
	경기도 부천시 성곡동	10	10
	경기도 덕양구 화정1동	10	10
	경기도 시흥시 정왕1동	10	10
	경기도 파주시 금촌1동	10	10
경기 읍면 지역	경기도 평택시 진위면	10	10
	경기도 남양주시 진건읍	10	10
	경기도 파주시 조리읍	10	10
	경기도 화성시 우정읍	10	10
	경기도 광주시 오포읍	10	10
	경기도 포천시 일동면	10	10

(5) 조사 대상자(개인) 추출 방법

추출된 집락(읍·면·동)에서 최종적으로 조사대상자(개인)를 선정하는 과정에서는 할당추출법을 사용했

다. 할당추출은 모집단을 구성하는 하부 집단(예를 들어 성별이나 연령대)의 구성비에 맞게 표본을 추출하는 방식이며, 사전 할당한 조건에 맞는 응답자를 찾아서 표집하는 방법이다. 비확률추출법이기는 하지만 조사를 신속하고 효율적으로 수행할 수 있다는 장점이 있으며, 확률추출을 적용하기 어려운 여건에서 모집단의 주요한 특성과 유사한 표본을 추출하기 위한 대안으로 여러 조사에서 널리 활용되고 있다.

1차 예비 조사에서는 개인의 국어능력과 밀접한 연관성을 지녔다고 볼 수 있는 성별과 연령대, 직업군, 학력을 기준으로 한 할당추출을 진행하였다. 성별과 연령대 변인은 집단별로 균등하게 할당했으며, 직업과 학력 변인은 2021년 인구주택총조사자료를 기준으로 비례 할당했다. 이미 모집단을 층화하는 단계에서 지역 변인을 고려했기 때문에 1차 예비 조사 표본 추출 시 고려한 변인은 ‘지역, 성별, 연령대, 학력, 직업’의 5개로 구성된다. 앞에서 설명한 지역 배분 기준을 제외하고, 나머지 각 변인 내 세부 구분은 다음과 같다.

① 성별 변인

성별 변인은 남성과 여성으로 구분하였다.

② 연령대 변인

연령대 변인은 만 20세에서 만 69세로 설정하였다. 이 범주에서 1세 구간으로는 표본을 할당하기 어렵기 때문에, 20대(만 20-29세), 30대(만 30-39세), 40대(만 40-49세), 50대(만 50-59세), 60대(만 60-69세)의 5개 구간으로 연령을 묶고 각 구간별로 표본을 할당했다.

③ 학력 변인

학력 변인은 ‘①무학, ②초등(국민)학교 졸업, ③중학교 중퇴, ④중학교 졸업, ⑤고등학교 중퇴, ⑥고등학교 재학, ⑦고등학교 졸업, ⑧대학교 재학, ⑨대학교 졸업, ⑩대학원 재학, ⑪대학원 졸업’의 11가지 항목으로 조사했다. 그러나 이러한 11가지 학력 수준별로 표본을 추출하는 것은 불가능하기 때문에, ‘중졸 이하(①~⑥), 고등학교 졸업(⑦), 대재 이상(⑧~⑪)’의 세 요소로 구분하여 표본을 추출했다.

④ 직업 변인

직업 변인은 한국 표준 직업 분류(KSCO)를 참고하여 13개 항목으로 나누어 조사했다. 표본 할당을 위한 직업군은 ‘정신노동(white collar)’, ‘육체노동(blue collar)’, 학생, 주부, 무직자를 포함한 ‘기타’의 세 집단으로 나누었다. 이때, 각 직업군에 속하는 세부 직업들은 <표 IV-3>과 같다.

<표 IV-3> 직업 구분

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

(6) 표본 구성

집락 추출 및 조사 영역 배분, 응답자 추출 과정을 통해 표집된 표본은 다음과 같이 구성된다. 평가 영역 내에서 설문지 유형은 무작위로 순환해서 조사를 진행했기 때문에 유형별 응답자 특성도 비교적 유사한 경향을 보이고 있다.

① 성×연령대×지역별 표본 구성

4개의 지역(서울, 인천, 경기 동부, 경기 읍면부), 성별(남자, 여자), 연령대(20대, 30대, 40대, 50대, 60대)로 총 40개의 구간을 설정하여 구간별로 동일 크기의 표본을 할당하였다. 하지만 실사 과정에서 직업과 학력, 검사지 영역, 유형까지 고려하여 조사함에 따라 표본을 선정하는 과정에서 어려움이 있어 구간별로 차등을 두어 조사하였다. 그럼에도 전체 성별로는 120명씩, 연령대별로는 24명씩, 지역별로는 60명씩 동일하게 조사를 완료하였다.

<표 IV-4> 1차 예비 조사 성×연령대×지역별 목표 표본

구 분	남	남	남	남	남	여	여	여	여	여	합계
	20대	30대	40대	50대	60대	20대	30대	40대	50대	60대	
합계	24	24	24	24	24	24	24	24	24	24	240
서울	6	6	6	6	6	6	6	6	6	6	60
인천	6	6	6	6	6	6	6	6	6	6	60
경기 동부	6	6	6	6	6	6	6	6	6	6	60
경기 읍면부	6	6	6	6	6	6	6	6	6	6	60

<표 IV-5> 1차 예비 조사 성×연령대×지역별 조사 완료 표본

구 분	남	남	남	남	남	여	여	여	여	여	합계
	20대	30대	40대	50대	60대	20대	30대	40대	50대	60대	
합계	24	24	24	24	24	24	24	24	24	24	240
서울	8	8	4	4	8	8	4	8	4	4	60
인천	8	8	4	4	8	8	4	8	4	4	60
경기 동부	4	4	8	8	4	4	8	4	8	8	60
경기 읍면부	4	4	8	8	4	4	8	4	8	8	60

② 지역×직업별 표본 구성

직업 변인은 2021년 인구주택총조사자료를 기준으로 4개 지역별로 비례 할당했다. 지역×직업별 목표 표본과 조사 완료 표본은 아래 <표 IV-6>, <표 IV-7>과 같으며, 구간별로 일부 차이를 허용하여 조사하였다.

<표 IV-6> 1차 예비 조사 지역×직업별 목표 표본

구분	정신노동	육체노동	기타	합계
합계	73	94	73	240
서울특별시	22	19	19	60
인천광역시	16	25	19	60
경기 동부	20	22	18	60
경기 읍면부	15	28	17	60

<표 IV-7> 1차 예비 조사 지역×직업별 조사 완료 표본

구분	정신노동	육체노동	기타	합계
합계	75	95	70	240
서울특별시	22	20	18	60
인천광역시	18	22	20	60
경기 동부	19	24	17	60
경기 읍면부	16	29	15	60

③ 지역×학력별 표본 구성

학력 변인도 마찬가지로 2021년 인구주택총조사자료를 기준으로 4개 지역별로 비례 할당했다. 지역×학력별 목표 표본과 조사 완료 표본은 아래 <표 IV-8>, <표 IV-9>와 같으며, 이 역시 구간별로 일부 차이를 허용하여 조사하였다.

<표 IV-8> 1차 예비 조사 지역×학력별 목표 표본

구분	중졸 이하	고졸	대재 이상	합계
합계	21	86	133	240
서울	4	16	40	60
인천	6	24	30	60
경기 동부	4	20	36	60
경기 읍면부	7	26	27	60

<표 IV-9> 1차 예비 조사 지역×학력별 조사 완료 표본

구분	중졸 이하	고졸	대재 이상	합계
합계	20	89	131	240
서울	2	18	40	60
인천	4	24	32	60
경기 동부	6	20	34	60
경기 읍면부	8	27	25	60

④ 검사지 유형별 표본 구성

8종의 검사지 유형별로 지역, 성별, 연령대, 직업, 학력이 고르게 조사될 수 있도록 표본을 할당하였으며, 검사지 유형별 표본 역시 구간별로 일부 차이를 허용하여 조사하였다. 목표 표본과 조사 완료 표본은 아래 <표 IV-10>, <표 IV-11>과 같다.

<표 IV-10> 1차 예비 조사 검사지 유형별 목표 표본

구분	검사지1	검사지2	검사지3	검사지4	검사지5	검사지6	검사지7	검사지8	합계
합계	30	30	30	30	30	30	30	30	240
지역	서울	7	8	7	8	7	8	7	60
	인천	8	7	8	7	8	7	8	60
	경기 동부	7	8	8	7	7	8	8	60
	경기 읍면부	8	7	7	8	8	7	7	60
성별	남자	15	15	15	15	15	15	15	120
	여자	15	15	15	15	15	15	15	120
연령대	20대	6	6	6	6	6	6	6	48
	30대	6	6	6	6	6	6	6	48
	40대	6	6	6	6	6	6	6	48
	50대	6	6	6	6	6	6	6	48
	60대	6	6	6	6	6	6	6	48
직업	정신노동	10	9	9	10	9	8	9	73
	육체노동	11	12	12	11	12	12	12	94
	기타	9	9	9	9	9	10	9	73
학력	중졸 이하	2	3	3	3	2	3	2	21
	고졸	11	10	11	11	11	10	11	86
	대재 이상	17	17	16	16	17	17	16	133

<표 IV-11> 1차 예비 조사 검사지 유형별 조사 완료 표본

구분		검사지1	검사지2	검사지3	검사지4	검사지5	검사지6	검사지7	검사지8	합계
합계		30	30	30	30	30	30	30	30	240
지역	서울	7	8	7	8	7	8	7	8	60
	인천	8	7	8	7	8	7	8	7	60
	경기 동부	7	8	7	7	8	8	8	7	60
	경기 읍면부	8	7	8	8	7	7	7	8	60
성별	남자	15	15	15	15	15	15	15	15	120
	여자	15	15	15	15	15	15	15	15	120
연령대	20대	6	6	6	6	6	6	6	6	48
	30대	6	6	6	6	6	6	6	6	48
	40대	6	6	6	6	6	6	6	6	48
	50대	6	6	6	6	6	6	6	6	48
	60대	6	6	6	6	6	6	6	6	48
직업	정신노동	10	9	7	13	8	13	8	7	75
	육체노동	10	14	16	9	18	6	9	13	95
	기타	10	7	7	8	4	11	13	10	70
학력	중졸 이하	2	3	4	2	3	2	2	2	20
	고졸	12	10	13	9	13	8	11	13	89
	대재 이상	16	17	13	19	14	20	17	15	131

2. 1차 예비 조사 진행

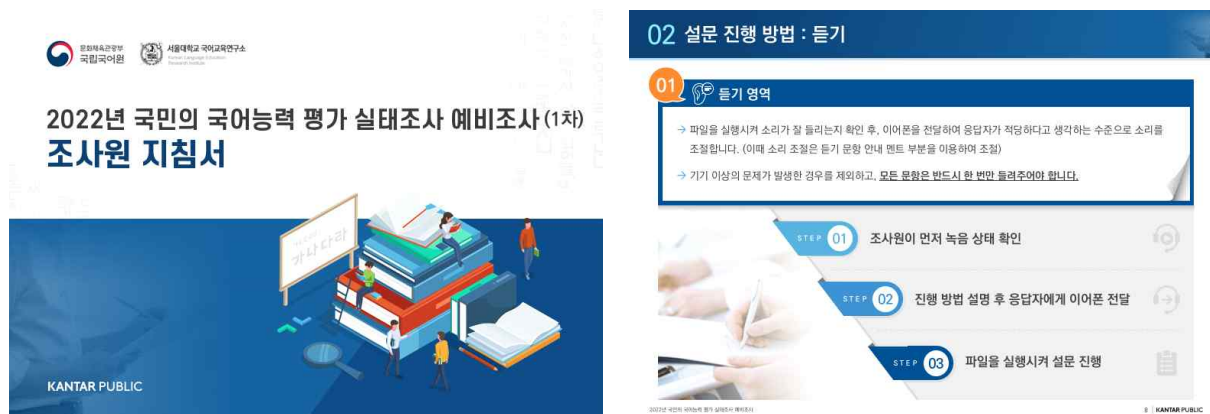
1) 조사 방법

본 조사는 태블릿 PC를 활용한 1:1 대인면접조사(TAPI: Tablet Assisted Personal Interview)로 진행됐으며, 태블릿 PC의 디지털 검사지를 통해 자료를 수집했다. 디지털 검사지 구동을 위한 소프트웨어는 안드로이드 OS 기반의 Nfield CAPI 프로그램을 이용했다. Nfield CAPI는 NIPO software에서 만든 조사 솔루션으로 오프라인 상태에서 조사가 진행되기 때문에 인터넷망의 접속이나 속도 문제로 인해 조사가 방해받는 현상을 막을 수 있다. 대신 현장에서 수집된 자료는 1차적으로 태블릿 PC에 저장되고, 나중에 인터넷 연결과 프로그램 동기화 과정을 통해 중앙 서버로 데이터가 전송되는 방식으로 작동한다.

2) 조사원 선발 및 교육

이번 조사는 응답자들이 일반 국민들이라는 점, 설문 조사라기보다는 시험과 같은 성격을 띠고 있다는 점, 응답자에게 말하기 등을 요구하는 점 때문에 조사의 곤란도가 매우 높다고 할 수 있다. 이를 고려하여 2017년과 2018년에 국어능력 실태 조사에 참여한 경험이 있는 조사원과 TAPI 조사 경험이 있는 경력 10년 이상의 조사원을 우선 선발하였으며, 총 17명을 선발하여 1차 예비 조사에 투입하였다. 조사원 17명이 240건의 설문을 진행하여 조사원 1인당 진행 표본 수는 약 14명이었다.

조사원 교육은 여의도에 위치한 KANTAR PUBLIC 교육장에서 2022년 10월 4일에 2시간가량 진행되었으며, 담당 연구원이 직접 교육을 진행했다. 조사원 교육은 인쇄된 설문지를 이용해 문항 구조와 개념을 이해하고, TAPI 시스템을 시연하는 방식으로 진행했다.



[그림 IV-3] 1차 예비 조사 조사원 지침서

<표 IV-12> 영역별 조사 진행 방법 및 지침

영역	조사 진행 방법 및 숙지 사항
듣기	● 파일을 실행시켜 소리가 잘 들리는지 확인 후, 이어폰을 전달하여 응답자에게 적당한 수준으로 소리 크기를 조절함 (소리 크기는 듣기 문항 안내 멘트 부분을 이용하여 조절) ● 기기 이상의 문제가 발생한 경우를 제외하고, 모든 문항을 반드시 한 번만 들려주어야 한다는 원칙 엄수
말하기	● 녹음이 잘 될 수 있도록 조용한 주변 환경이 유지될 수 있도록 함 ● 각 문항별로 준비 시간 1~2분과 제공 및 녹음 테스트 ● 2분간 응답자의 응답 내용 녹음 (응답자 ID와 문항 번호를 말하고 응답 시작) ● 녹음 버튼을 누르면 녹음 시작, 동일한 버튼을 한 번 더 누르면 녹음 저장(만약 녹음 종료 후 녹음 버튼을 또 누르면 기존의 녹음이 삭제되므로 주의) ● 녹음 완료 후 녹음 상태 확인 및 저장 후 다음 말하기 문항 진행 ● 혹시 모를 상황에 대비하여 개인 휴대전화로 동시에 녹음 진행
듣기 말하기	● 파일을 실행시켜 소리가 잘 들리는지 확인 후, 이어폰을 전달하여 응답자에게 적당한 수준으로 소리 크기를 조절함 (소리 크기는 듣기 문항 안내 멘트 부분을 이용하여 조절) ● 기기 이상의 문제가 발생한 경우를 제외하고, 모든 문항을 반드시 한 번만 들려주어야 한다는 원칙 엄수 ● 녹음이 잘 될 수 있도록 조용한 주변 환경이 유지될 수 있도록 함 ● 각 문항별로 준비 시간 5초 제공 및 녹음 테스트 ● 20~30초간 응답자의 응답 내용 녹음 (응답자 ID와 문항 번호를 말하고 응답 시작) ● 녹음 버튼을 누르면 녹음 시작, 동일한 버튼을 한 번 더 누르면 녹음 저장(만약 녹음 종료 후 녹음 버튼을 또 누르면 기존의 녹음이 삭제되므로 주의) ● 녹음 완료 후 녹음 상태 확인 및 저장 후 다음 말하기 문항 진행 ● 혹시 모를 상황에 대비하여 개인 휴대전화로 동시에 녹음 진행

3) 1차 예비 조사 진행 일정

<표 IV-13> 1차 예비 조사 진행 일정

단계	일정
표본 설계 및 설문지 설계	~ 9월 중순
TAPI 시스템 구축	9월 20일 ~ 9월 30일
조사원 교육	10월 4일
TAPI 시스템 최종 점검 및 보완	10월 4일
현장 조사 진행	10월 5일 ~ 10월 27일
사후 검증	10월 20일 ~ 10월 27일
통계 처리	10월 28일 ~ 10월 31일
결과 보고	10월 31일

4) 조사 검증

각 조사원이 조사한 표본 중 20% 이상 검증하는 것을 목표로 표본을 무작위로 추출해 전화 검증을 진행했다. 전화 검증은 별도의 검증 요원이 응답자에게 직접 전화를 걸어 조사 참여 여부, 주요 인구특성 등을 재확인하는 과정으로 진행했다.

1차 예비 조사 완료 표본 240건 중 171건에 대해 검증을 시도하여 통화 시도율은 71%였으며, 이 중 검증을 완료한 표본은 56건으로 검증률은 23%다. 세부 검증 결과는 다음과 같다.

<표 IV-14> 1차 예비 조사 검증 결과

구분	표본 수
전체	240
A. 검증 설문지 총수	171
B. 검증 완성 설문지	56
1. 인터뷰했음	56
C. 검증 미완성 설문지 수	115
2. 인터뷰한 적 없음	-
3. 인터뷰했으나 검증 질문의 응답 틀림	-
4. 응답 거절(끊음, 통화 곤란 등)	2
5. 응답자 부재중	1
6. 통화 중	-
7. 전화 안 받음 / 10. 자동응답전화	109
8. 번호 틀림	-
9. 결번	-
11. 기타 (나중에, 회의 중)	3
통화 시도율(B/A)	71%
검증률(B/전체 표본)	23%

일반적인 검증 절차와는 별도로 말하기 녹음 파일 전수를 다시 청취해 녹음 상태가 불량하거나 누락된 경우가 있는지 추가로 확인하는 과정도 거쳤다. 녹음 파일 1,200개 전체에 대한 검수 과정을 거쳤으며, 이 중 4건의 녹음 파일 오류가 발견되어 재조사를 진행하였다.

5) 결과물 제출

‘듣기’ 영역의 조사 결과는 SPSS 데이터와 엑셀(EXCEL) 데이터, 주요 인구학적 변인과의 교차분석 통계표 형태로 제출하였다. ‘말하기’와 ‘듣기·말하기’ 영역은 응답자의 말하기 내용을 녹음한 디지털 음성파일(MP3)로 제출하였다.

3. 1차 예비 조사 관련 특이 사항

1) TAPI 활용 관련

2017년 조사와 마찬가지로 이번 예비 조사에서도 TAPI를 활용하여 조사하여 조사 방식이 편리하고 효율적이었다. 이번 예비 조사의 경우 듣기와 말하기 영역만으로 구성되어 있어 종이 설문지는 이용하지 않고 100% 태블릿 PC로만 조사되었다.

‘말하기’ 영역의 경우 하나의 TAPI 시스템 내에서 조사 진행과 녹음이 동시에 진행되면서 응답자 ID와 녹음 파일을 매칭하는 과정이 정확하고 간편하였다. 그러나 일부 조사원의 조작 실수로 녹음 버튼을 한 번 더 누르게 되면, 기존 파일이 삭제되고 새로운 파일이 생성되며 응답 결과가 소실되는 문제가 발견되었다. 이에 조사원의 휴대폰 녹음 기능을 이용해 말하기 조사 과정을 동시 녹음하는 방식으로 응답 파일 유실에 대비했다. 그리고 실사 관리자가 사후 검증 차원에서 말하기 음성 파일 1,200개를 전수 재청취했는데, 이를 위해서는 해당 응답자의 음성 파일을 일일이 다운로드 받아 재생해야 하는 번거로움이 있었다. 이로 인해 조사 종료 후 데이터 점검 과정에 많은 시간이 소요되었는데, 대규모의 본 조사를 진행하기 위해서는 예산을 추가로 투입하여 사후 데이터 점검 과정까지 고려한 시스템 도입이 필요해 보인다.

2) 실사 과정상의 문제점

1차 예비 조사는 기본적으로 5개 변인별로 표본을 할당하고, 여기서 다시 검사지 유형별로 표본을 할당하였다. 표본 수가 240명이기 때문에 많은 변인의 할당을 고려하며 실사를 진행하기에 어려움이 따랐다. 각 영역과 세트별로 표본 수를 맞추면서, 인구학적 변인까지 고려하며 표본을 선정하는 과정에서 오는 어려움이다. 또한 녹음 자체에 대한 거부감으로 인해 조사를 거절한 경우도 많았다.

응답자의 부담감과 피로감도 컸는데 전 영역이 듣기와 말하기, 듣기·말하기 영역으로만 구성되어 있고 특히 후반부에 말하기 영역을 제시하였기 때문에 더 그러했다. 이로 인하여 말하기 영역의 응답 질이 일부 떨어지는 문제가 있었을 것으로 예측된다.

2차 예비 조사는 1차 예비 조사와 동일한 환경과 설문 순서를 지켜 조사하였지만 본 조사에서는 문항 수 조절 혹은 말하기 영역을 앞으로 배치하는 등의 조치가 필요해 보인다.

4. 1차 예비 조사 결과

1) 1차 예비 조사 결과의 채점

(1) 선택형 문항(듣기)의 채점

1차 예비 조사는 태블릿 PC를 활용했기 때문에 응답자가 화면상에서 특정 보기를 터치하면, 응답 결과가 곧바로 CAPI 시스템에 입력되는 방식으로 진행했다. 따라서 종이 설문지를 이용한 기존 조사처럼 별도의 전산 입력 과정을 거치거나, OMR 카드 사용 등을 병행할 필요가 없었다. 현장 조사 종료 후 서울대 국어교육연구소로부터 각 문항의 정답을 제공받았고, 이를 통계 처리된 데이터와 비교하여 각 문항별 정답률을 계산했다.

(2) 수행형 문항(듣기·말하기, 말하기)의 채점

1차 예비 조사의 주관식 문항은 (가), (나)형 2종으로 출제되었으며, 전체 문항 유형은 가교 문항을 포함하여 모두 5개로 구성되었다. 1차 예비 조사에는 240명의 피험자가 참여하였으며, 수행형 문항에서 총 1,200건의 응답이 수집되어 이를 채점 대상으로 확정하였다. 채점은 본 연구의 목적 달성과 신뢰도 확보를 위해 다음의 채점 계획을 수립하였다.

첫째, 각 문항 유형별로 전문가 3인의 공통 채점을 위해 전체 5개 문항 유형에 해당하는 15인의 채점단을 확보한다. 채점단은 국어교육학 전공의 석사 및 박사 과정생으로서 국어교육 전문 인력으로 구성한다.

둘째, 신뢰도 확보에 부정적인 영향을 끼칠 수 있는 개별 채점자의 채점자 효과(rater effect)를 제한하고, 일정 수준 이상의 채점자 간 신뢰도 확보를 위해 각 문항 유형별로 3인 공통 채점 훈련 및 본 채점을 진행한다.

셋째, 채점 대상 문제 및 채점 척도 사용 수준의 눈높이를 맞추기 위한 웹 기반 훈련⁹⁾에 모든 채점자가 반드시 참여하며, 훈련 과정에서 실시한 사전 채점 및 소집단 협의, 종합 협의를 통해 세부적인 채점 실행 사항을 점검한다.

9) 2022년 11월 4일(금) 19시부터 21시까지 비대면 화상 회의(zoom) 방식으로 실시하였음.

<표 IV-15> 수행형 문항의 영역 및 문항, 채점자 구성

영역	문항 유형 및 응답 수		채점자
듣기.말하기 (LS)	1번	(가)형(LS1-가) 120개 응답	오지은, 이혜린, 윤현아
		(나)형(LS1-나) 120개 응답	
	2번	(가)형(LS2-가) 120개 응답	최혜진, 임지원, 송미지
		(나)형(LS2-나) 120개 응답	
말하기 (S)	3번	가교(S1) 240개 응답	김다연, 이은석, 이민경
	4번	(가)형(S2-가) 120개 응답	김동섭, 권근영, 이정민
		(나)형(S2-나) 120개 응답	
	5번	(가)형(S3-가) 120개 응답	권아람, 박장범, 안이진
		(나)형(S3-나) 120개 응답	

1차 예비 조사의 문항 유형별 채점자 집단은 화법 및 문법, 문학 전공의 서울대 대학원 석사 및 박사 과정생으로 구성하였으며, 채점자 집단 설계에서 이들의 전공 특성 영향을 상쇄하기 위해 전체 5개 집단 모두 화법, 문법, 문학 전공자가 각각 한 명씩 배치되도록 하였다.

1차 예비 조사 채점의 목적은 각 문항 유형 및 검사지별 세부 문제에 대한 양호도를 확인하는 것이며, 채점의 효율성과 신뢰성을 동시에 확보하기 위하여 각 문항 유형을 전문적으로 채점할 채점자 집단을 선정하고, 해당 문항 유형의 전체 응답을 집단 내에서 동시에 채점하는 집단별 공통 피험자 설계를 적용하였다. 이는 채점자 간 신뢰도 확보의 측면에서 단일 문항 유형에 한정하여 채점을 하기 때문에 여러 문항을 동시에 채점해야 하는 가교 피험자 설계나 가교 문항 설계에 비해 각 문항 유형별로 전문적인 접근이 가능하다는 장점이 있다. 이에 따라 듣기·말하기 1,440건, 말하기 2,160건의 채점 결과값이 분석 대상이 되었다. 1차 예비 조사의 본 채점은 다음과 같은 방법으로 진행되었다.

첫째, 각 문항 유형 및 검사지별 세부 문제에 대한 피험자 응답 파일을 분류한다.

둘째, 채점 수행의 신뢰도 확보를 위해 문항 유형별로 채점 척도를 구성하고, 세부 문제에 따라 실제적인 사항을 반영하기 위하여 채점자 워크숍에서 논의된 사항을 반영하여 본 채점 수행을 위한 채점 척도를 확정한다.

셋째, 채점자들의 일관된 채점 척도 적용을 위해 채점 척도에 각 문항 유형별로 설정한 내용, 조직, 표현, 전달의 채점 준거별 세부 수행 항목을 제시하고, 이를 계량적으로 판단할 수 있는 지침을 제시하였다.

<표 IV-16> 사례: 말하기 3번 문항(S3) 채점 기준 및 세부 채점 기준

조사 항목		채점 요소	채점 척도(점수)					점수 비중	
			매우 미흡	미흡	보통	우수	매우 우수		
내용 구성	내용 선정	- 질문에서 요구하는 대답을 한다. - 청자가 이해하기 쉽게 구체적으로 설명한다.	3	6	9	12	15	30%	60%
	조직	- 내용의 흐름이 설명에 적합하다. - 시작과 마무리가 자연스럽다.	3	6	9	12	15	30%	
표현 및 전달	표현	- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.	2	4	6	8	10	20%	40%
	전달	- 발음, 속도, 어조가 적절하다. - 적절한 시간을 지켜 말한다.	2	4	6	8	10	20%	
조사 항목		세부 채점 기준							
내용 선정		이 항목에서 가장 중요한 기준은 ‘①해당 문제에 대한 이해’, ‘②설명’하는 말하기 - 매우 미흡: ①과 ②의 내용을 하나도 갖추지 않은 경우. 질문과 전혀 상관없는 이야기를 한 경우. - 미흡: ①과 ②중 하나 이상의 요소가 있지만 내용이 모호하거나 설명이 구체적 예시로 나타나지 않는 경우. 예를 들어, 초등학생이 넘어져서 다쳤을 때 취해야 하는 대처법에 대한 설명이 나타나야 함. 그런데 이것이 아닐 경우(즉, 넘어져서 다쳤으면 일단 조심을 해야 하고, 집에 가서 밴드 붙이자.) 이는 질문에 요구하는 대답이 아님. - 보통: ①과 ②에 해당하는 요소가 있는 경우. ‘설명’이 있는지에 관한 판단은 구체적인 예시가 있는지를 기준으로 판단. - 우수: ①과 ②를 만족하며 ③구체적인 설명(상황에 대한 설명, 행동의 이유 등)이 나 몇 가지 예시를 추가적으로 더 드는 경우. - 매우 우수: ①과 ②를 만족하며 ③구체적인 설명(상황에 대한 설명, 행동의 이유 등)과 함께 한두 가지 예시를 추가적으로 더 드는 경우.							
조직		이 항목에서 가장 중요한 기준은 ‘①내용의 흐름(상황 파악, 선후관계 등)’과 ‘②시작과 마무리의 표현’ - 매우 미흡: ①을 명백하게 위배한 경우, 즉 내용이 주제와 관련이 없는 경우, 상황을 잘못 이해한 경우 - 미흡: ①을 위배한 경우로, 내용은 주제와 관련이 있지만 일관된 흐름 없이 뒤죽박죽 말하는 경우. - 보통: ①을 지켜 말한 경우 - 우수: ①과 ②를 지켜 말하며 상황을 정확하게 파악하고 자연스럽게 내용이 이어지는 경우. - 매우 우수: ①과 ②를 지켜 말하며 시작과 마무리 내용이 포함되는 경우. 즉 상황을 정확하게 파악하고 자연스럽게 내용이 이어지며, 청자의 주의를 환기하는 도							

	입부까지 있는 경우.
표현	이 항목에서 가장 중요한 기준은 ‘①적절한 어휘 사용’, ‘②문법에 맞는 표현’ ①적절성: 상황에 맞는 어휘의 사용(초등학생에게 적합한 쉽고 구체적인 어휘의 사용), 어휘의 의미를 잘 알고 사용하는 경우 등 ②명확성: 어법에 어긋난 문장이나 표현을 사용하는 경우, 간명하지 못한 표현 등 - 매우 미흡: ①과 ②가 모두 부족한 경우, 매우 어려운 어휘를 쓰거나 어법에 어긋나는 표현이 지나치게 많은 경우 - 미흡: ①과 ② 중 하나가 부족한 경우. 예를 들어 아이가 이해하기에 적합하지 않은 어려운 어휘나 어법을 사용하는 경우. - 보통: ①과 ②가 보통 수준으로 사용된 경우. 기본적으로 아이에게 이야기하듯이 말하는 경우. 예를 들어 ‘여름에 밖에 둔 음식을 먹으면 식중독에 걸릴 수 있으니 까 조심해야 해요’와 같이 아이에게 말하는 것처럼 표현한 경우. - 우수: ①과 ②를 모두 적절하게 사용하는 경우. 즉, 초등학생이 이해하기 쉬운 어휘를 사용하는 경우. - 매우 우수: ①과 ②를 모두 적절하게 사용하는 경우, 즉 초등학생에게 맞는 어휘나 표현을 사용하고, 어려울 수 있는 표현은 풀어서 설명하며 어법에 맞게 말한 경우.
전달	이 항목의 기준은 ‘①시간 준수’와 ‘②반언어적 표현’ - 매우 미흡: ①시간이 10초 이하인 경우 - 미흡: ①시간이 10초에서 30초의 분량인 경우. ②말을 지나치게 더듬거나 너무 빠르게 하여 내용 전달이 명확하지 않게 되는 경우. 또한 분량이 보통 수준이더라도, 지나치게 느리게 말하거나 더듬어서 말이 길어진 경우에는 미흡에 해당. - 보통: ①시간이 30초에서 40초의 분량인 경우 ②적절한 수준으로 평범하게(천천히 또박또박) 말하는 경우. - 우수: ①시간이 40초를 넘기며 ②반언어적 표현을 적절히 사용하여 이야기를 유창하게 지속하는 경우. - 매우 우수: ①시간이 1분 30초 이상이며, ②반언어적 표현을 적절히 사용한 경우

넷째, 채점의 효율적인 진행과 채점자 피로도 관리를 위해 본 채점은 1차(40%), 2차(60%)로 나누어 진행하며, 채점의 결과를 실시간으로 확인 및 관리하기 위해 개인화된 웹 채점표 링크에 접속하여 점수를 입력한다.

① 듣기·말하기 영역

듣기·말하기는 이번 조사를 통해 새롭게 도입된 영역으로, 1차 예비 조사에는 2개 문항 유형, 세부 영

역별로 (가)형과 (나)형의 총 4개 문제가 출제되었다. 이 영역에서 확보된 응답은 총 480건이며, 이에 대하여 내용, 조직, 표현, 전달 준거별로 비중에 따라 3~9점, 2~6점 척도로 수준을 나누어 채점을 진행하였다. 듣기·말하기 문항의 점수 척도는 3~15점, 2~10점으로 구성된 말하기 문항과 달리 실제성을 고려하여 피험자의 응답이 즉각적으로 이루어진다는 점과 이에 따라 간단한 형태로 응답이 이루어지기 때문에 세밀한 척도를 적용할 경우 변별이 어렵다는 점 등을 고려하여 1~3점으로 구성하였다.

<표 IV-17> 듣기·말하기 1번 (가)형 문항 채점 척도

조사 항목		채점 요소	채점 척도(점수)			점수 비중	
			미흡	보통	우수		
내용 구성	내용 선정	- 문제에서 요구하는 응답을 한다. (배려) - 상대의 부담을 최소화하고, 합리적인 대안을 제시한다.	3	6	9	30%	60%
	조직	- 응답의 흐름이 배려를 나타내는 데 적합하다. - 발화의 시작과 마무리가 자연스럽다.	3	6	9	30%	
표현 및 전달	표현	- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.	2	4	6	20%	40%
	전달	- 발음, 속도, 어조 등이 자연스럽다(유창성). - 발음이 명료하다(정확성).	2	4	6	20%	

<표 IV-18> 듣기·말하기 1번 (나)형 문항 채점 척도

조사 항목		채점 요소	채점 척도(점수)			점수 비중	
			미흡	보통	우수		
내용 구성	내용 선정	- 문제에서 요구하는 응답을 한다. (사과) - 잘못된 내용에 대해 언급하고, 상대의 기분을 헤아려 말한다.	3	6	9	30%	60%
	조직	- 응답의 흐름이 사과하는 데 적합하다. - 발화의 시작과 마무리가 자연스럽다.	3	6	9	30%	
표현 및 전달	표현	- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.	2	4	6	20%	40%
	전달	- 발음, 속도, 어조 등이 자연스럽다(유창성). - 발음이 명료하다(정확성).	2	4	6	20%	

<표 IV-19> 듣기·말하기 2번 (가)형 문항 채점 척도

조사 항목		채점 요소	채점 척도(점수)			점수 비중	
			미흡	보통	우수		
내용 구성	내용 선정	- 문제에 부합하는 적절한 응답을 한다. - 문제에 대한 설명과 해결을 위한 반응이 잘 드러난다(적극성).	3	6	9	30%	60%
	조직	- 응답이 이전 발화의 흐름에 잘 대응된다. - 발화의 시작과 마무리가 자연스럽다.	3	6	9	30%	
표현 및 전달	표현	- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.	2	4	6	20%	40%
	전달	- 발음, 속도, 어조 등이 자연스럽다(유창성). - 발음이 명료하다(정확성).	2	4	6	20%	

<표 IV-20> 듣기·말하기 2번 (나)형 문항 채점 척도

조사 항목		채점 요소	채점 척도(점수)			점수 비중	
			미흡	보통	우수		
내용 구성	내용 선정	- 과제에서 요구하는 응답을 한다(거절). - 거절의 이유에 대해 설명하고, 대안을 제시한다.	3	6	9	30%	60%
	조직	- 응답이 이전 발화의 흐름에 잘 대응된다. - 발화의 시작과 마무리가 자연스럽다.	3	6	9	30%	
표현 및 전달	표현	- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.	2	4	6	20%	40%
	전달	- 발음, 속도, 어조 등이 자연스럽다(유창성). - 발음이 명료하다(정확성).	2	4	6	20%	

② 말하기 영역

말하기 영역은 이번 1차 예비 조사에서 2018년 조사와의 동등화 가능성을 마련하기 위해 선정한 가교 문항(1개)과 동형 문항(1개), 그리고 신규 출제된 문항 유형(1개)으로 구성하였다. 이 중에서 가교 문항은 2018 조사에서 안정적인 곤란도 및 변별도 수치를 나타낸 개념 및 경험 설명하기를 조사하는 문항이다. 기존 조사와 형식적으로 같은 문항 유형에서는 토론 상황에서의 주장하는 말을 조사하며, 신규 출제된 문항은 공적 상황에서의 발표하는 말하기를 조사하는 문항이다. 말하기 영역에 대한 채점은 기존 조사와 마찬가지로 응답을 준비하는 시간을 제공한다는 점, 응답을 위해 고려해야 할 구체적인 조건이 있다는

점 등을 고려하여 기존의 말하기 조사 문항과 같이 5단계의 척도(3-6-9-12-15, 2-4-6-8-10)로 이루어졌다.

<표 IV-21> 말하기 1번 가교 문항 채점 척도

조사 항목		채점 요소	채점 척도(점수)					점수 비중	
			매우 미흡	미흡	보통	우수	매우 우수		
내용 구성	내용 선정	- 질문에서 요구하는 대답을 한다. - 청자가 이해하기 쉽게 구체적으로 설명한다.	3	6	9	12	15	30%	60%
	조직	- 내용의 흐름이 설명에 적합하다. - 시작과 마무리가 자연스럽다.	3	6	9	12	15	30%	
표현 및 전달	표현	- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.	2	4	6	8	10	20%	40%
	전달	- 발음, 속도, 어조가 적절하다. - 적절한 시간을 지켜 말한다.	2	4	6	8	10	20%	

<표 IV-22> 말하기 2번 문항 채점 척도

조사 항목		채점 요소	채점 척도(점수)					점수 비중	
			매우 미흡	미흡	보통	우수	매우 우수		
내용 구성	내용 선정	- 슬라이드의 내용을 충분히 설명한다. - 복잡한 내용을 간명하게 정리하여 말한다.	3	6	9	12	15	30%	60%
	조직	- 발표의 흐름이 자연스럽다. - 발화의 시작과 마무리가 효과적이다.	3	6	9	12	15	30%	
표현 및 전달	표현	- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.	2	4	6	8	10	20%	40%
	전달	- 발음, 속도, 어조가 등이 자연스럽다(유창성). - 발음이 명료하다(정확성).	2	4	6	8	10	20%	

<표 IV-23> 말하기 3번 문항 채점 척도

조사 항목		채점 요소	채점 척도(점수)					점수 비중	
			매우 미흡	미흡	보통	우수	매우 우수		
내용 구성	내용 선정	- 문제에서 요구하는 응답을 한다. - 문제 상황에 대한 설명과 주장의 근거가 잘 드러난다.	3	6	9	12	15	30%	60%
	조직	- 내용의 흐름이 논증에 적합하다. - 발화의 시작과 마무리가 효과적이다.	3	6	9	12	15	30%	
표현 및 전달	표현	- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.	2	4	6	8	10	20%	40%
	전달	- 발음, 속도, 어조가 등이 자연스럽다(유창성). - 적절한 시간을 지켜 말한다(10초 미만: 매우 미흡, 30초: 미흡, 30초 이상은 감점 없음.)	2	4	6	8	10	20%	

(3) 수행형 문항(듣기·말하기, 말하기) 채점의 채점자 신뢰도 및 영향 분석

① 듣기·말하기 영역

<표 IV-24> 1차 예비 조사 듣기·말하기 영역의 채점자 간 신뢰도

영역	문항 유형		채점자	급내상관	df	유의확률
듣기·말하기 (LS)	1번	LS1-가 LS1-나	KR011 KR012 KR013	.735	951	.000
	2번	LS2-가 LS2-나	KR021 KR022 KR023	.767	479	.000

1차 예비 조사의 채점 설계에 따라 듣기·말하기 영역의 채점자 신뢰도를 알아보기 위해 급내상관을 확인하였다. 급내상관은 여러 명의 채점자가 같은 피험자를 채점한 경우, 채점자 간의 입력 점수가 개별 채점자 또는 평균값을 기준으로 서로 떨어진 거리인 분산에 따라 0을 기준으로 1에 가까울수록 일치도가 높음을 나타낸다.

1차 예비 조사에 대한 채점자 간 신뢰도 분석 결과, 듣기·말하기 1, 2번 유형 모두 급내상관이 .7 이상으로 양호한 수준인 것으로 나타났다. 이를 통해 두 문항 유형 모두 신뢰할 만한 수준으로 채점이 이루어졌음을 확인하였다.

<표 IV-25> 1차 예비 조사 듣기·말하기 영역의 채점자 영향

영역	문항 유형		채점자	logit	SE
듣기· 말하기 (LS)	1번	LS1-가	KR011	-1.88	.060
		LS1-나	KR012	-1.84	.060
			KR013	-1.56	.060
	2번	LS2-가	KR021	-2.55	.080
		LS2-나	KR022	-1.81	.070
			KR023	-2.70	.110

이번 조사에서 새롭게 출제된 듣기·말하기 문항의 채점 결과에서 채점자 영향을 알아보기 위해 다국면 라쉬모형(MFRM) 분석을 실시하였다. 채점자 영향을 의미하는 로짓(logit)은 0을 기준으로 값이 양의 방향으로 커질수록 엄격한 채점 경향을 나타낸다. 분석 결과 1, 2번 문항 유형을 담당한 채점자 모두 매우 관대하게 채점을 한 것으로 나타났다. 특히 2번 문항 유형을 담당한 KR021의 경우 듣기·말하기 영역에서 가장 관대한 경향을 갖고 있는 것으로 나타났다. 관대한 채점 경향이 채점 척도 사용의 문제와 관련이 되었는지 알아보기 위해 채점 척도 사용의 적합도를 확인하였다. 그 결과, KR023의 경우 척도 사용의 일관성이 다소 부족한 것으로 나타났으나, 문제가 되는 수준으로 판정되지는 않았다. 이상의 채점 경향 분석 결과를 바탕으로 2차 예비 조사의 듣기·말하기 문항 유형 채점에서는 다음의 방안을 고려할 필요가 있다.

첫째, 채점자들이 채점 척도 사용에서 관대한 경향으로 넘어가는 문지방 수준을 높이기 위해 최고 수준인 ‘우수’ 수준에 대한 엄격한 기준 적용이 필요하다. 내용 측면에서는 듣기·말하기 문항이 짧고 간단한 형태의 응답이 주를 이루는 상황에서 문제가 되지 않는 선의 응답도 우수 수준으로 판정하였던 것을 요건을 갖추었을 때에만 ‘우수’ 수준으로 판정하는 것으로 수정하고, 전달 측면에서도 정해진 시간을 지키지 않은 경우에는 ‘우수’ 수준이 아니라 ‘보통’ 수준 이하에서만 채점이 가능하도록 제한할 필요가 있다.

둘째, 채점자들은 관대한 채점 특성을 보이는 가운데 일정한 수준 이상의 양호한 신뢰도를 나타내었다. 다만 채점 워크숍의 효과를 보다 지속적으로 유지할 수 있도록 채점 기한을 지나치게 여유를 두지 않으면서 계획적으로 채점을 실시하도록 채점 계획을 조정할 필요가 있다.

② 말하기 영역

1차 예비 조사의 채점 설계에 따라 말하기 영역의 채점자 신뢰도를 알아보기 위해 급내상관을 확인하였다. 분석 결과, 말하기 1, 2, 3번 유형 모두 급내상관이 .8 이상으로 채점자 간 신뢰도가 우수한 수준인 것으로 나타났다. 이를 통해 모든 말하기 문항 유형에서 신뢰할 만한 수준으로 채점이 이루어졌음을 확인하였다.

<표 IV-26> 1차 예비 조사 말하기 영역의 채점자 간 신뢰도

영역	문항 유형		채점자	급내상관	df	유의확률
말하기 (S)	1번	S1(가교)	KR031	.923	959	.000
			KR032			
			KR033			
	2번	S2-가 S2-나	KR041	.892	955	.000
			KR042			
			KR043			
	3번	S3-가 S3-나	KR051	.902	935	.000
			KR052			
			KR053			

<표 IV-27> 1차 예비 조사 말하기 영역의 채점자 영향

영역	문항 유형		채점자 집단	logit	SE
말하기 (S)	1번	S1(가교)	KR031	.440	.040
			KR032	.630	.040
			KR033	.010	.040
	2번	S2-가 S2-나	KR041	.630	.040
			KR042	-.060	.040
			KR043	.290	.040
	3번	S3-가 S3-나	KR051	.410	.040
			KR052	.840	.040
			KR053	.220	.040

1차 예비 조사 말하기 문항의 채점 결과에서 채점자 효과를 알아보기 위해 다국면라쉬모형(MFRM) 분석을 실시하였다. 분석 결과 말하기 1, 2, 3번 문항 유형을 담당한 채점자들은 대체로 다소 엄격한 경향으로 채점한 것으로 나타났다. 이 중에서 KR052의 경우 담당한 말하기 3번 문항 유형에서 가장 엄격하게 채점을 하였으며, 반대로 말하기 2번 문항 유형을 맡은 KR042의 경우 가장 관대한 경향으로 채점을 한 것으로 나타났다.

이상의 듣기·말하기 및 말하기 문항의 1차 예비 조사 채점자 간 신뢰도 및 채점자 영향 분석을 바탕으로 2차 예비 조사 채점에 대해 얻은 시사점은 다음과 같다.

첫째, 듣기·말하기 문항 유형의 신뢰도 수준을 말하기 문항 유형의 수준으로 끌어올릴 수 있도록 단답 형태의 응답 유형에 대한 일관된 판정 기준을 수립하여 2차 예비 조사 채점자 훈련 및 본 채점에 적용한다.

둘째, 각 문항 유형별로 3인 공통 채점 설계를 적용한 결과, 듣기·말하기 영역은 관대한 경향이, 말하기 영역은 엄격한 경향이 나타난 것과 관련된 사항을 2차 예비 조사의 채점 워크숍에 반영하여 관대한 경향의 집단에는 문지방 수준을 높이기 위해 채점 척도의 세부 기준 적용을 강조하고, 엄격한 집단에는 문지방 수준을 다소 낮출 수 있도록 엄격한 경향이 두드러졌던 문항을 2차 조사에서 제외할 수 있도록 한다.

셋째, 2차 예비 조사 채점에서는 전체 채점자의 특성을 비교하기 위해 공통 문항 피험자 설계를 추가하여 단일 문항 채점으로 인한 채점 특성 심화 경향을 상쇄하도록 한다.

2) 1차 예비 조사 결과 분석

(1) 검사 설계상의 특징

<표 IV-28> 듣기 영역 검사지별 문항 구성

검사지	문항 구성			문항 수	응시자 수
	1, 2번 (공통/가교)	3~15번(가)	3~15번(나)		
듣기(가)	포함	포함		15	120
듣기(나)	포함		포함	15	120

<표 IV-29> 듣기·말하기, 말하기 영역 검사지별 문항 구성

검사지	문항 구성					문항 수	응시자 수
	듣기 ·말하기(가)	듣기 ·말하기(나)	말하기(가)	말하기(나)	3번 (가교)		
듣·말(가) /말(가)	포함		포함		포함	5	60
듣·말(가) /말(나)	포함			포함	포함	5	60
듣·말(나) /말(가)		포함	포함		포함	5	60
듣·말(나) /말(나)		포함		포함	포함	5	60

검사의 설계상 영역별 원점수 총점은 검사지에 따라서 다른 의미를 가진다. 각 검사지에 포함된 문항이 지니는 변별도나 곤란도에 다소 차이가 있게 마련이기 때문이다. 따라서 각기 다른 검사지로 시험을 치른 피험자들의 원점수 총점을 비교하는 것은 적절하지 않고, 문항반응이론을 적용하여 피험자의 능력을 추정하는 방식으로 비교가 가능하다.

(2) 영역별 점수 분포 분석

① 기술통계

<표 IV-30> 영역별 점수 분포의 기술통계치

영역	사례 수	최솟값	최댓값	평균	표준편차
듣기	240	28.01	63.15	49.93	8.32
듣기.말하기	240	38.66	60.82	51.94	4.03
말하기	240	36.69	69.34	50.00	6.81

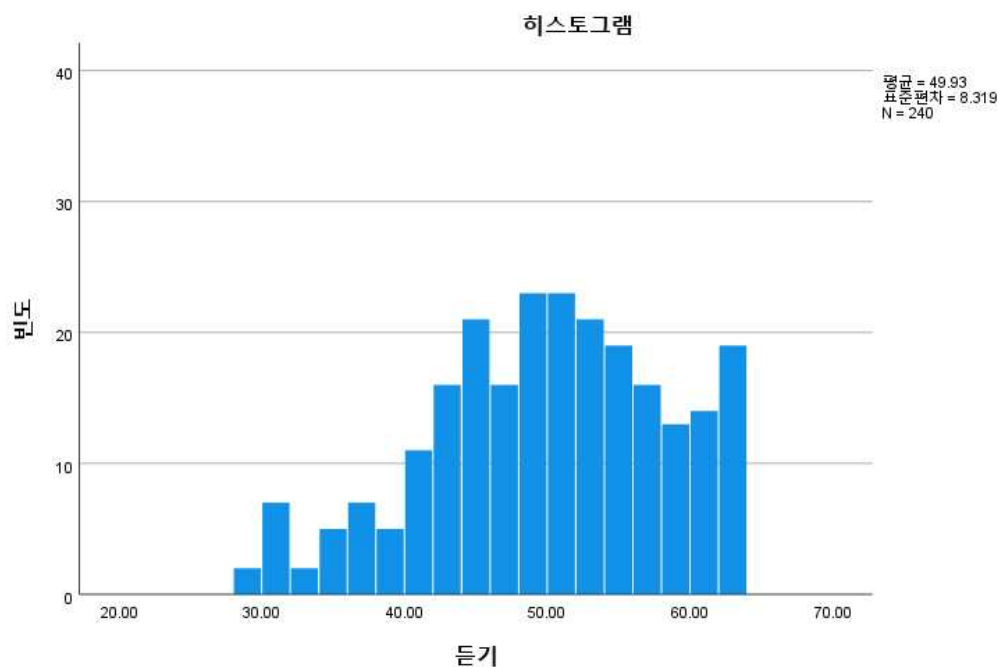
② 영역별 점수 분석 방법

선택형 문항으로 구성된 듣기 영역의 경우, 2모수 문항반응이론 모형을 적용하여 분석하였다. 피험자 능력 모수의 추정치(평균 0, 표준편차 1)를 산출한 후, 이 점수에 10점을 곱하고, 50점을 더하여 평균 50, 표준편차 10인 t점수로 영역별 점수를 최종 산출하였다.

능력 모수 추정치는 표본 집단이 아닌 모집단에 대한 추정치이므로 기술통계 결과표에서 각각의 영역별 점수의 평균과 표준편차는 각각 50점과 10점에서 다소 차이가 있다.

수행형 문항으로 구성된 듣기·말하기 영역과 말하기 영역의 경우, 문항 양호도를 파악하기 위해 문항 곤란도는 각 문항별로 원점수를 t점수로 변환하여 계산한 평균값과 다국면라쉬모형 분석 결과에서 제시된 로짓 척도값을 이용하였다. 문항 변별도의 경우에는 Facets 측정 결과 중 문항반응이론에 따른 2모수 일반화 부분 점수 모형(2PL Generalized partial credit model)을 적용하여 산출하였다.

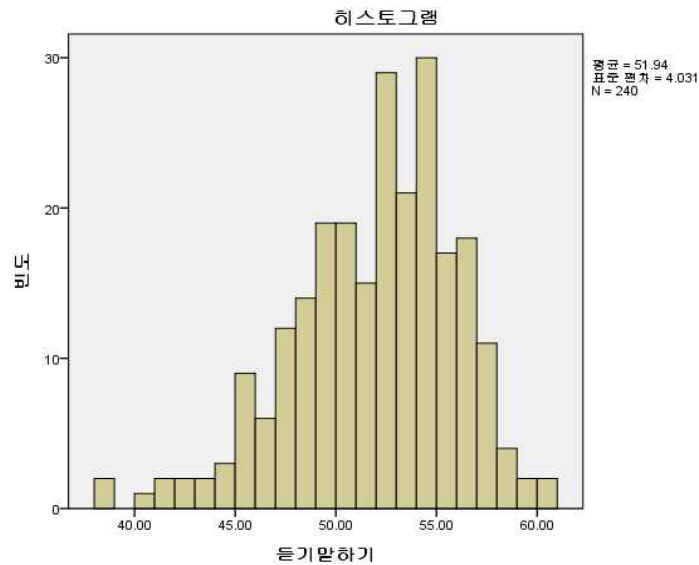
③ 듣기 영역



[그림 IV-4] 1차 예비 조사의 듣기 영역 점수 분포

듣기 영역 점수 분포는 정규분포에 가깝지만, 왜도가 -.408로 오른쪽으로 약간 치우치는 부정 편포를 보이고, 첨도가 -.366으로 중심이 다소 낮고 분포가 비교적 퍼져 있는 모양을 보인다.

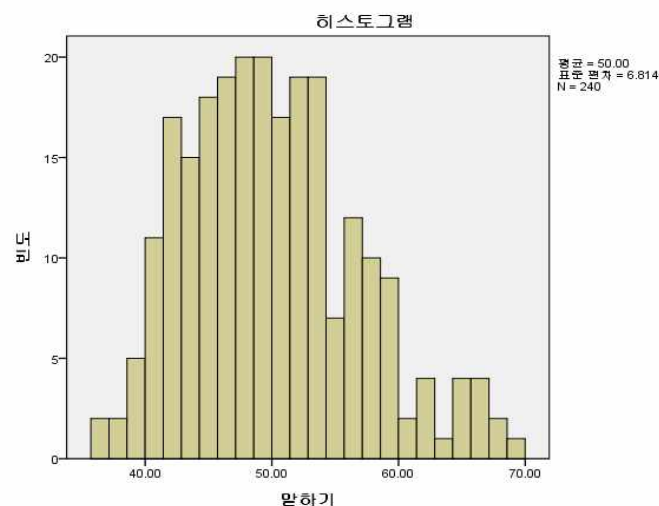
④ 듣기.말하기 영역



[그림 IV-5] 1차 예비 조사의 듣기·말하기 영역 점수 분포

듣기·말하기 영역 점수 분포는 정규분포에 가깝지만, 왜도가 -.612로 오른쪽으로 약간 치우치는 부정 편포를 보이고, 첨도가 .386으로 중심이 다소 높고 분포가 비교적 뾰족한 모양을 보인다.

⑤ 말하기 영역



[그림 IV-6] 1차 예비 조사의 말하기 영역 점수 분포

말하기 영역 점수 분포는 정규분포에 가깝지만, 왜도가 .541로 왼쪽으로 약간 치우치는 정적 편포를 보이고, 첨도가 -.045로 중심의 높이가 정규분포에 매우 가까운 모양을 보인다.

(3) 영역별 문항 곤란도 및 변별도 분석

① 듣기 영역

<표 IV-31> 1차 예비 조사 듣기 영역 문항 곤란도 및 변별도 분석

문항	사례 수	정답률	문항 모수 추정치				적합도 ¹⁰⁾		
			변별도	SE	곤란도	SE	χ^2	df	p-value
듣기_01(공통)	240	56.7	3.733	8925.395	-.184	108.745	36.672	8	.000
듣기_02(공통)	240	72.9	.288	125.140	-3.503	1466.144	234.264	8	.000
듣기_03(가)	120	91.7	1.603	.629	-2.099	.504	2.578	8	.958
듣기_04(가)	120	62.5	.355	.264	-1.516	1.175	13.395	8	.099
듣기_05(가)	120	88.3	1.547	.570	-1.835	.432	4.140	8	.844
듣기_06(가)	120	64.2	1.230	.431	-0.642	.239	18.053	8	.021
듣기_07(가)	120	66.7	.181	.253	-3.894	5.421	10.747	8	.216
듣기_08(가)	120	60.8	-.178	.249	2.465	3.610	1.886	8	.984
듣기_09(가)	120	87.5	.955	.407	-2.401	.820	3.910	8	.865
듣기_10(가)	120	85.8	.703	.368	-2.836	1.292	2.985	8	.935
듣기_11(가)	120	88.3	1.166	.462	-2.169	0.635	3.678	8	.885
듣기_12(가)	120	64.2	.961	.333	-.754	.295	14.995	8	.059
듣기_13(가)	120	75.8	1.413	.448	-1.129	.281	4.928	8	.765
듣기_14(가)	120	92.5	1.166	.519	-2.644	.865	1.538	8	.992
듣기_15(가)	120	52.5	1.027	.363	-.148	.217	5.047	8	.753
듣기_03(나)	120	92.5	.391	.425	-6.565	6.854	11.705	8	.165
듣기_04(나)	120	80.8	1.813	.578	-1.186	.265	3.189	8	.922
듣기_05(나)	120	67.5	1.593	.460	-.658	.199	5.345	8	.720
듣기_06(나)	120	61.7	1.426	.389	-.456	.191	13.110	8	.108
듣기_07(나)	120	93.3	3.660	1.407	-1.620	.247	1.752	8	.988
듣기_08(나)	120	38.3	.957	.306	.602	.273	10.746	8	.233
듣기_09(나)	120	90.8	2.156	.746	-1.684	.332	2.018	8	.981
듣기_10(나)	120	67.5	1.405	.414	-.702	.221	4.716	8	.788
듣기_11(나)	120	78.3	1.742	.545	-1.086	.252	6.442	8	.598
듣기_12(나)	120	24.2	.188	.256	6.134	8.252	1.715	8	.989
듣기_13(나)	120	60.8	.983	.313	-0.529	.257	5.145	8	.742
듣기_14(나)	120	78.3	1.972	.625	-1.027	.227	4.309	8	.828

듣기_15(나)	120	95.0	1.822	.753	-2.264	.561	2.716	8	.951
----------	-----	------	-------	------	--------	------	-------	---	------

듣기 영역 28문항에 대하여 2모수 문항반응이론 모형을 적용하여 분석한 결과, 일부 문항에서 양호하다고 보기 어려운 변별도와 곤란도 분석 결과를 보였다.

듣기 4번(가), 7번(가), 8번(가), 3번(나), 12번(나) 문항의 경우, 변별도가 .35 미만으로 거의 없는 것으로 나타났다. 특히 8번(가) 문항은 변별도가 음수인 역변별 문항 즉, 듣기 능력이 부족할수록 정답을 더 잘 맞히는 좋지 못한 특성을 지닌 문항에 해당한다.

듣기 3번(가), 7번(가), 9번(가), 10번(가), 11번(가), 3번(나), 14번(가), 15번(나) 문항의 경우 곤란도가 -2.0 미만으로 매우 쉬운 것으로 나타났고, 8번(가), 12번(나) 문항의 경우 곤란도가 2.0 이상으로 매우 어려운 것으로 나타났다. 검사지에 다양한 곤란도를 지닌 문항이 고루 포함되어야 다양한 능력을 지닌 피험자를 고루 변별할 수 있으므로 검사지에 이들 문항이 일부 포함되는 것은 적절할 수 있으나 지나치게 많이 포함되어서는 안 될 것이다.

② 듣기·말하기 영역

<표 IV-32> 1차 예비 조사 듣기·말하기 문항 양호도

문항	사례 수	t점수 평균	문항 모수 추정치				적합도			
			변별도	SE	곤란도	SE	내적합	내적합 표준화	외적합	외적합 표준화
듣기·말하기 1-가	120	50.97	1.00	.060	-.310	.050	.990	-.200	1.020	.300
듣기·말하기 1-나	120	49.38	1.09	.060	.640	.050	.900	-2.800	.950	-1.100
듣기·말하기 2-가	120	55.07	1.01	.060	.370	.060	1.030	.800	.950	-.900
듣기·말하기 2-나	120	53.61	.920	.060	-.710	.080	1.110	1.6	1.150	1.400

듣기·말하기 영역의 2개 문항 유형 4 문제에 대하여 다국면라쉬모형 분석을 실시한 결과, 전체적으로 채점 척도 사용의 경향은 안정적으로 나타났으나, 듣기·말하기 2번 유형의 (나)형 문제의 경우 문제의 곤란도가 낮고 변별도도 떨어지는 것으로 나타났다. 해당 문제는 척도 사용에서도 다른 문제에 비해 일관성이 부족한 경향이 있어 2차 예비 조사에서는 사용하지 않기로 하였다.

10) Yen(1981)의 제안에 따라 검정함.

③ 말하기 영역

<표 IV-33> 1차 예비 조사 말하기 문항 양호도

문항	사례 수	t점수 평균	문항 모수 추정치				적합도			
			변별도	SE	곤란도	SE	내적합	내적합 표준화	외적합	외적합 표준화
말하기 1-가교	240	50.01	.980	.030	-.030	.020	.980	-.800	1.010	.300
말하기 2-가	120	50.17	1.160	.030	-.250	.030	1.160	4.300	1.170	4.400
말하기 2-나	120	49.83	1.110	.030	.050	.030	1.110	2.900	1.120	3.200
말하기 3-가	120	50.22	.770	.030	.160	.030	.990	-.200	.970	-.900
말하기 3-나	120	49.79	.740	.030	.060	.030	.790	-6.200	.780	-6.700

말하기 영역의 경우 3개 문항 유형 전체 5개 문제에 대해 Facets 분석을 통해 문항 양호도를 파악하였다. 그 결과, 문항 모수 추정치에서는 말하기 문항 3의 (가)형에서 변별도가 낮음에도 곤란도가 높은 역전 현상이 나타났다. 적합도 측면에서는 전체적으로 채점 척도의 사용은 적절하게 이루어졌으나, 말하기 문항 2 유형에서는 표준화 값이 0.000을 기준으로 값의 분포가 정규적인 특성이 나타나지 않았으며, 문항 3의 (나)형에서는 표준화 값이 -2 이하로 나타나 기계적인 채점 패턴이 나타났다.

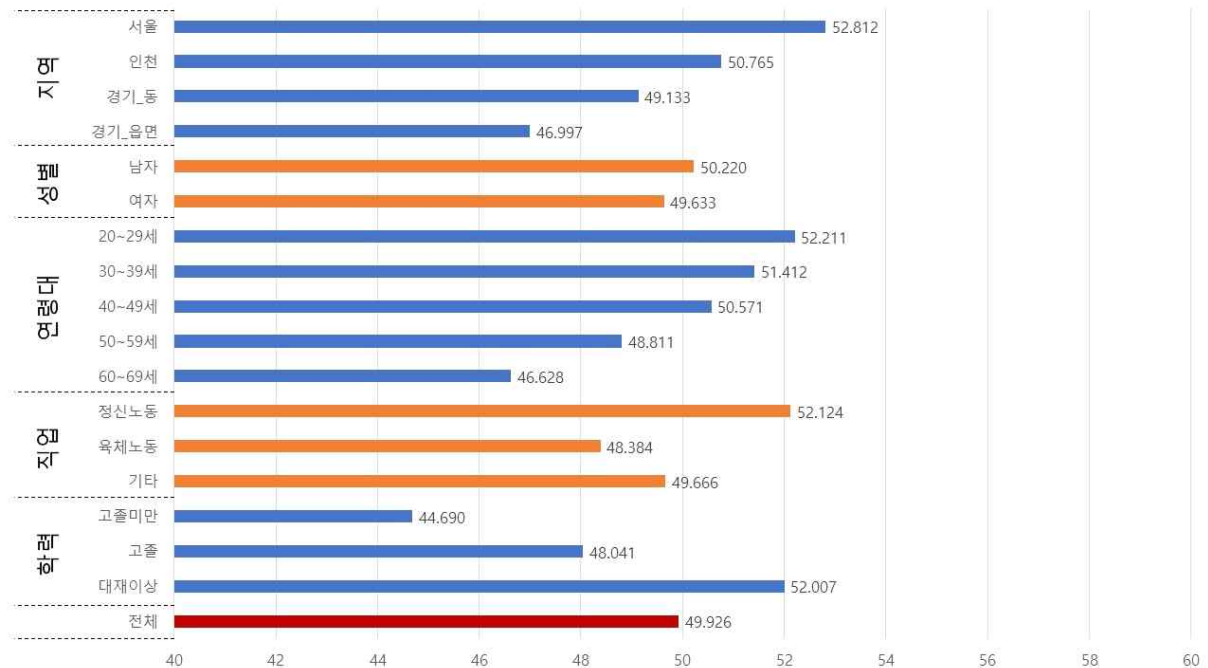
이상의 문항 양호도 분석 결과를 바탕으로 말하기 영역에서는 문항 모수 추정치에서 변별도와 곤란도가 역전된 경향이 나타난 말하기 3번 유형의 (가)형을 2차 예비 조사에서 제외하기로 하였다. 아울러 채점 척도 사용의 특이성이 드러난 말하기 2번, 3번 유형에 대해서는 2차 예비 조사 채점 워크숍을 통해 명세화된 채점 방법을 제시하기로 하였다.

(4) 배경 변인별 분석

① 듣기 영역

<표 IV-34> 1차 예비 조사 듣기 영역 배경 변인별 분석

구분		사례 수	평균	표준편차
지역	서울	60	52.812	8.830
	인천	60	50.765	6.381
	경기_동	60	49.133	7.308
	경기_읍면	60	46.997	9.472
성별	남자	120	50.220	8.779
	여자	120	49.633	7.858
연령대	20-29세	48	52.211	6.902
	30-39세	48	51.412	9.262
	40-49세	48	50.571	7.800
	50-59세	48	48.811	8.726
	60~69세	48	46.628	7.817
직업	정신노동	75	52.124	7.961
	육체노동	95	48.384	8.507
	기타	70	49.666	8.034
학력	고졸 미만	20	44.690	7.683
	고졸	89	48.041	7.903
	대재 이상	131	52.007	8.082
전체		240	49.926	8.319



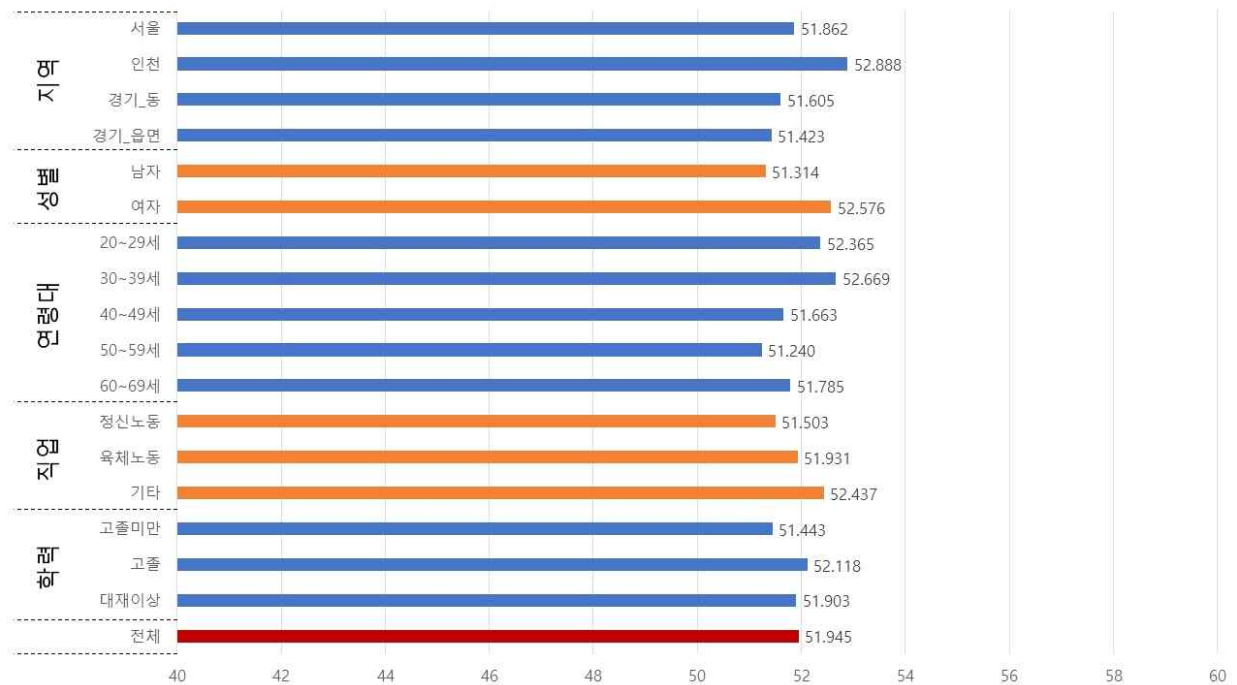
[그림 IV-7] 1차 예비 조사 듣기 영역 배경 변인별 분석

듣기 영역 척도 점수의 배경 변인별 차이를 살펴보면, 지역에 따라서 ‘서울’이 52.812점으로 전체 평균에 비해 다소 높은 수준이었고, ‘인천’과 ‘경기 동’은 각각 50.765점과 49.133점으로 평균과 비슷한 수준이었으며, ‘경기 읍면’은 46.997점으로 평균에 비해 낮은 수준이었다. 연령대에 따라서는 ‘20~29세’와 ‘30~39세’가 각각 52.211점과 51.412점으로 전체 평균에 비해 다소 높은 수준이었고, ‘40~49세’가 50.571점으로 평균과 비슷한 수준이었으며, ‘50~59세’와 ‘60~69세’가 각각 48.811점과 46.628점으로 평균에 비해 낮은 수준이었다. 직업에 따라서는 ‘정신노동’이 52.124점으로 ‘육체노동’의 48.384점에 비해 높았다. 학력에 따라서는 ‘대재 이상’이 52.007점으로 가장 높고, ‘고졸’이 48.041점으로 뒤를 이었으며, ‘고졸 미만’이 44.690점으로 가장 낮은 수준이었다.

② 듣기·말하기 영역

<표 IV-35> 1차 예비 조사 듣기·말하기 영역 배경 변인별 분석

구분		사례 수	평균	표준편차
지역	서울	60	51.862	4.261
	인천	60	52.888	3.620
	경기_동	60	51.605	3.977
	경기_읍면	60	51.423	4.176
성별	남자	120	51.314	4.207
	여자	120	52.576	3.759
연령대	20-29세	48	52.365	3.600
	30-39세	48	52.669	3.136
	40-49세	48	51.663	4.331
	50-59세	48	51.240	4.590
	60~69세	48	51.785	4.320
직업	정신노동	75	51.503	4.100
	육체노동	95	51.931	4.002
	기타	70	52.437	3.996
학력	고졸 미만	20	51.443	4.551
	고졸	89	52.118	3.963
	대재 이상	131	51.903	4.018
전체		240	51.945	4.031



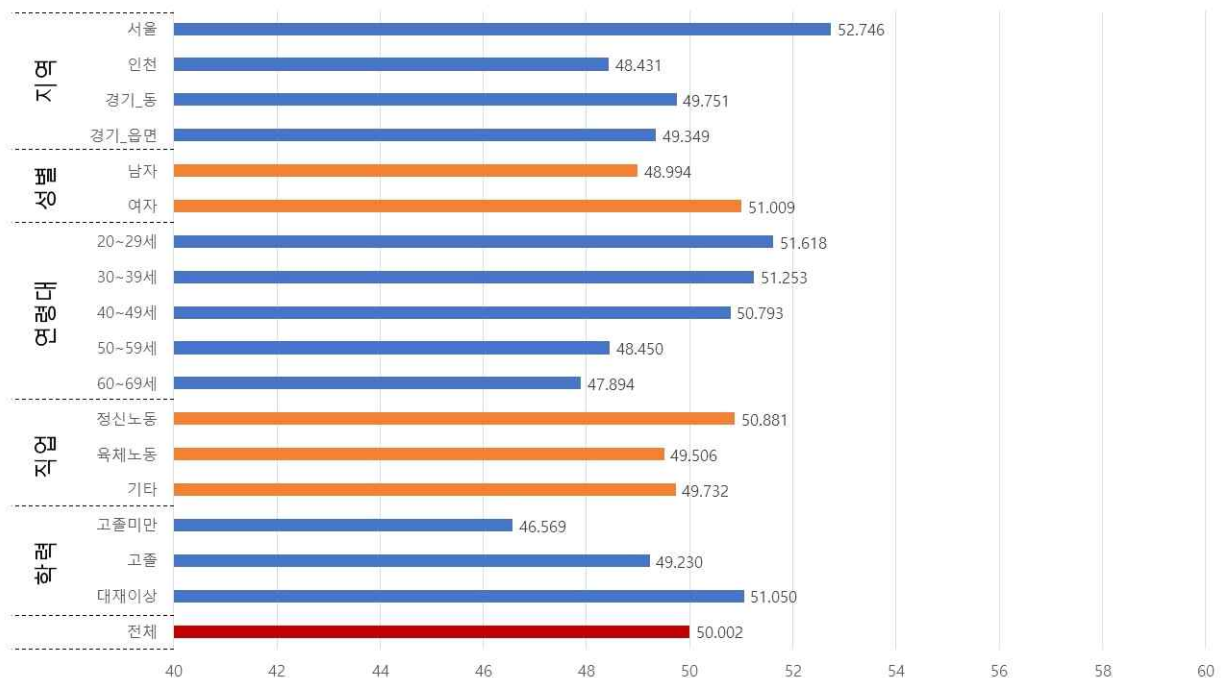
[그림 IV-8] 1차 예비 조사 듣기·말하기 영역 배경 변인별 분석

듣기·말하기 영역 척도 점수의 배경 변인별 차이를 살펴보면, 지역에 따라서 ‘인천’이 52.888점으로 전체 평균에 비해 다소 높은 수준이었고, ‘서울’, ‘경기 동’, ‘경기 읍면’은 각각 51.862점, 51.605점, 51.423점으로 평균과 비슷하거나 약간 낮은 수준이었다. 성별에 따라서는 ‘여자’가 52.576점으로 ‘남자’의 51.314점에 비해 높았다. 연령대에 따라서는 ‘20~29세’와 ‘30~39세’가 각각 52.365점과 52.669점으로 전체 평균에 비해 다소 높은 수준이었고, ‘40~49세’, ‘50~59세’, ‘60~69세’는 각각 51.663점, 51.240점, 51.785점으로 평균과 비슷하거나 약간 낮은 수준이었다. 직업에 따라서는 ‘육체노동’이 51.931점으로 ‘정신노동’의 51.503점에 비해 높았다. 학력에 따라서는 ‘고졸’이 52.118점으로 가장 높고, ‘대재 이상’이 51.903점으로 뒤를 이었으며, ‘고졸 미만’이 51.443점으로 가장 낮은 수준이었다.

③ 말하기 영역

<표 IV-36> 1차 예비 조사 말하기 영역 배경 변인별 분석

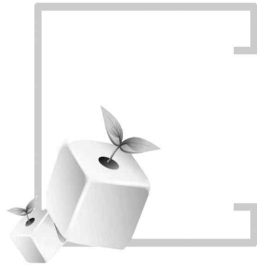
구분		사례 수	평균	표준편차
지역	서울	60	52.476	6.091
	인천	60	48.431	7.424
	경기_동	60	49.751	6.951
	경기_읍면	60	49.349	6.190
성별	남자	120	48.994	6.324
	여자	120	51.009	7.157
연령대	20~29세	48	51.618	7.534
	30~39세	48	51.253	7.143
	40~49세	48	50.793	6.210
	50~59세	48	48.450	6.142
	60~69세	48	47.894	6.337
직업	정신노동	75	50.881	6.883
	육체노동	95	49.506	6.777
	기타	70	49.732	6.798
학력	고졸 미만	20	46.569	6.401
	고졸	89	49.230	6.478
	대제 이상	131	51.050	6.903
전체		240	50.002	6.814



[그림 IV-9] 1차 예비 조사 말하기 영역 배경 변인별 분석

말하기 영역 척도 점수의 배경 변인별 차이를 살펴보면, 지역에 따라서 ‘서울’이 52.746점으로 전체

평균에 비해 다소 높은 수준이었고, '인천', '경기 동', '경기 읍면'은 각각 48.431점, 49.751점, 49.349점으로 평균에 비해 낮은 수준이었다. 성별에 따라서는 '여자'가 51.009점으로 '남자'의 48.994점에 비해 높았다. 연령대에 따라서는 '20~29세'와 '30~39세'가 각각 51.618점과 51.253점으로 전체 평균에 비해 다소 높은 수준이었고, '40~49세'가 50.793점으로 평균과 비슷한 수준이었으며, '50~59세'와 '60~69세'가 각각 48.450점과 47.894점으로 평균에 비해 낮은 수준이었다. 직업에 따라서는 '정신노동'이 50.881점으로 '육체노동'의 49.506점에 비해 높았다. 학력에 따라서는 '대재 이상'이 51.050점으로 가장 높고, '고졸'이 49.230점으로 뒤를 이었으며, '고졸 미만'이 46.569점으로 가장 낮은 수준이었다.



제 5 장

2차 예비 조사 시행 및 결과 분석



1. 2차 예비 조사 설계

1) 2차 예비 조사를 위한 검사지 구성

(1) 영역별 검사지 구성

1차 예비 조사와 마찬가지로 검사지 영역은 듣기, 말하기, 듣기·말하기 총 3가지 영역이며, 각 영역은 (가)형과 (나)형으로 구분된다. 영역별 2가지 유형(가형, (나)형)을 조합하여 검사지를 총 8세트로 구성하였다.

(2) 영역별 검사 문항의 실제

영역별 검사 문항도 1차 조사와 마찬가지로 ‘문두’ 및 ‘선지’는 화면에 읽을 수 있는 형태로 제공하였고 ‘지문’과 ‘문두’까지 음성 녹음 자료로 제공하였다. 아래 그림은 듣기 영역의 예시이며 지문 플레이를 조사원이 누르면 문두와 함께 지문이 재생되고 피조사자는 문두 및 선지를 읽은 후 답안을 제출한다.

듣기·말하기와 말하기 영역도 듣기 영역과 마찬가지로 ‘문두’ 및 ‘선지’는 화면에 읽을 수 있는 형태로 제공하였고 ‘지문’과 ‘문두’ 모두 음성 녹음 자료로 제공하였다. 지문 플레이를 조사원이 누르면 문두와 함께 지문이 재생되고 피조사자는 말하기를 위한 준비 시간을 가진 뒤 녹음 버튼을 눌러 말하기 답안을 제출한다.

2) 2차 예비 조사의 설계

(1) 모집단 및 표본 수

2차 예비 조사의 모집단은 1차와 동일한 서울, 경기, 인천에 거주하는 만 20세~69세의 남녀로 정의하였다. 본 조사의 목적은 개인 차원의 행동과 태도를 측정하는 것이므로, 조사 기본 단위(elementary unit : 조사의 대상이 되는 가장 최소의 요소)는 개인으로 정의된다.

목표 표본 수는 조사 목적과 조사비용, 기간 등을 고려해 총 240명으로 결정하였다. 240명 표본의 오차 한계는 단순 확률 표본을 기준으로 95% 신뢰 수준에서 $\pm 6.3\%$ 로 계산된다.

(2) 표본 수 배분

예비 조사는 본 조사에서 사용할 문항의 2배수 분량을 조사하기 때문에, 모든 응답자에게 전체 문항을 조사한다면 응답자의 피로감이 높아질 가능성이 매우 크다. 이로 인해 각 영역별로 (가)형과 (나)형을 구분하였고, 검사지 유형별 표본 수는 동일하게 배분하였다.

(3) 모집단 층화와 표본 수 배분

층화추출은 모집단을 서로 겹치지 않은 그룹들, 즉 층(strata)들로 나누고 각 층에서 표본을 뽑는 방법이다. 표본의 크기가 같은 경우, 층화추출은 단순 확률 추출을 통해 얻은 표본보다 추정 오차의 한계를 줄일 수 있고, 특히 층 내의 관측값들이 동질적인 경우에 더욱 효과적인 추출법으로 알려져 있다. 이 밖에도 모집단 조사 단위들을 서로 편리한 그룹으로 묶음으로써 조사 시 비용을 절감할 수 있고, 모집단 내의 하위 그룹들에 대한 모수 추정값을 따로 구할 수 있다는 장점이 있다.

각 층을 나누는 기준은 연구 목적에 따라 여러 가지를 고려할 수 있지만, 광범위한 지역을 대상으로 이루어지는 조사에서는 행정 구역이나 지리적 위치를 층화의 1차적 기준으로 삼는 경우가 많다. 이 연구에서는 전체 모집단을 서울, 인천, 경기도 읍면지역, 경기도 동지역이라는 4개의 층으로 구분하고, 각 층에서 표본을 추출하는 층화추출을 진행했다.

다음으로 4개 권역으로 구분된 각 층별로 몇 명의 표본을 조사해야 하는지를 결정하였다. 이번 예비 조사는 모집단 전체의 국어능력을 추정하려는 목적보다는 다양한 인구학적 특성을 지닌 표본을 추출해 국어능력 평가 조사의 평가지의 타당성과 신뢰도를 검토하기 위한 목적에서 진행됐다. 이러한 조사 목적을 감안해 각 층별로 최대한 같은 크기의 표본 수를 배분하고자 했다.

(4) 집락(읍·면·동) 추출과 응답자 추출

각 층별 표본 수를 결정한 후에는 각 층을 통계적으로 대표할 수 있도록 표본을 추출하였다. 이때에는 2단계 집락추출법(cluster sampling)을 사용하였다. 통계적 개념에서의 집락(cluster)은 서로 인접해 있는 개별 조사 단위들이 모여 있는 집단을 의미한다. 그리고 집락추출은 먼저 조사를 진행할 집락을 선정한 후에 그 집락에서 표본을 추출해서 조사를 진행하는 방식이다. 이때 집락에 속한 기본 단위를 전수 조사한다면 1단계 집락추출이 되지만, 일부 표본을 다시 추출해서 조사한다면 2단계 집락추출이 되는 것이다.

이 조사에서는 행정 구역상의 읍·면·동을 집락으로 간주하고, 2단계 집락추출을 진행하였다. 간단히 정리하자면 각 층에서 통계적인 방법을 통해 읍·면·동을 먼저 추출한 후에, 그 지역에서 설문 조사에 참여할 개인을 추출한 것이다.

각 층 내에서는 크기비례확률추출(Sampling with probabilities proportional to size)을 통해 읍·면·동을 선정하였다. 크기비례확률추출은 규모가 큰 집락일수록 표본으로 선정될 확률이 비례해서 높아지게 된다. 특히 집락들의 규모가 심하게 차이가 나거나 대규모 집락이 전체 모집단에 미치는 영향이 클 경우에는 각 집락의 크기에 비례하여 뽑는 방법인 크기비례확률추출법이 널리 사용되고 있다. 이렇게 크기비례확률추출을 통해 집락을 추출하였을 경우에는 각 집락에서 동일한 수의 표본을 추출하는 것을 권장한다. 이렇게 하면 각 표본별로 추출 확률이 같아지는 자체가중표본(self-weighting sample)을 얻을 수 있다. 자체가중표본을 활용하면 결과 분석 과정에서 보다 간편하고 경제적인 작업이 가능하다. 각

층별로는 추가적인 가중치 보정 없는 단순 평균값으로도 결과를 산출할 수 있기 때문이다. 이 조사에서는 여러 가지 여건을 감안해 하나의 집락에서 조사되는 표본 수를 10명으로 결정하였다. 따라서 목표 표본 수 240명을 조사하기 위한 집락 수는 총 24개로 설계하였다. 이러한 과정을 거쳐 최종적으로 추출한 집락과 집락별 표본 수는 다음과 같다.

<표 V-1> 2차 예비 조사 집락(읍·면·동) 추출 결과

총 구분	집락(읍·면·동)	목표 표본 수	조사 완료 표본 수
서울	서울특별시 강북구 수유1동	10	10
	서울특별시 성북구 정릉제3동	10	10
	서울특별시 서대문구 남가좌제2동	10	10
	서울특별시 양천구 신정3동	10	10
	서울특별시 동작구 사당제1동	10	10
	서울특별시 강동구 길동	10	10
인천	인천광역시 동구 송림1동	10	10
	인천광역시 미추홀구 용현5동	10	10
	인천광역시 남동구 만수5동	10	10
	인천광역시 부평구 청천2동	10	10
	인천광역시 서구 가정1동	10	10
	인천광역시 계양구 계산4동	10	10
경기 동지역	경기도 권선구 금곡동	10	10
	경기도 의정부시 신곡2동	10	10
	경기도 부천시 신중동	10	10
	경기도 일산서구 대화동	10	10
	경기도 수정구 태평3동	10	10
	경기도 구리시 수택3동	10	10
경기 읍면지역	경기도 남양주시 진접읍	10	10
	경기도 김포시 고촌읍	10	10
	경기도 안성시 공도읍	10	10
	경기도 화성시 매송면	10	10
	경기도 광주시 퇴촌면	10	10
	경기도 양주시 광적면	10	10

(5) 조사 대상자(개인) 추출 방법

추출된 집락(읍·면·동)에서 최종적으로 조사 대상자(개인)를 선정하는 과정에서는 할당추출법을 사용했다. 할당추출은 모집단을 구성하는 하부 집단(예를 들어 성별이나 연령대)의 구성비에 맞게 표본을 추출하는 방식이며, 사전 할당한 조건에 맞는 응답자를 찾아서 표집하는 방법이다. 비확률추출법이기는 하지만 조사를 신속하고 효율적으로 수행할 수 있다는 장점이 있으며, 확률추출을 적용하기 어려운 여건에서 모집단의 주요한 특성과 유사한 표본을 추출하기 위한 대안으로 여러 조사에서 널리 활용되고 있다.

2차 예비 조사에서는 개인의 국어능력과 밀접한 연관성을 지녔다고 볼 수 있는 성별과 연령대, 직업군, 학력을 기준으로 한 할당추출을 진행하였다. 성별과 연령대 변인은 집단별로 균등하게 할당했으며,

직업과 학력 변인은 2021년 인구주택총조사자료를 기준으로 비례 할당했다. 이미 모집단을 총화하는 단계에서 지역 변인을 고려했기 때문에 표본 추출 시 고려한 변인은 ‘지역, 성별, 연령대, 학력, 직업’의 5개로 구성된다. 앞에서 설명한 지역 배분 기준을 제외하고, 나머지 각 변인 내 세부 구분은 다음과 같다.

① 성별 변인

성별 변인은 남성과 여성으로 구분하였다.

② 연령대 변인

연령대 변인은 만 20세에서 만 69세로 설정하였다. 만 20세~69세의 각 1세 연령별로는 표본을 할당하기는 어렵기 때문에, 20대(만 20-29세), 30대(만 30-39세), 40대(만 40-49세), 50대(만 50-59세), 60대(만 60-69세)의 5개 구간으로 연령을 묶고 각 구간별로 표본을 할당했다.

③ 학력 변인

학력 변인은 ‘①무학, ②초등(국민)학교 졸업, ③중학교 중퇴, ④중학교 졸업, ⑤고등학교 중퇴, ⑥고등학교 재학, ⑦고등학교 졸업, ⑧대학교 재학, ⑨대학교 졸업, ⑩대학원 재학, ⑪대학원 졸업’의 11가지 항목으로 조사했다. 그러나 이러한 11가지 학력 수준별로 표본을 추출하는 것은 불가능하기 때문에, ‘중졸 이하(①~⑥), 고등학교 졸업(⑦), 대재 이상(⑧~⑪)’의 세 요소로 구분하여 표본을 추출했다.

④ 직업 변인

직업 변인은 한국 표준 직업 분류(KSCO)를 참고하여 13개 항목으로 나누어 조사했다. 표본 할당을 위한 직업군은 ‘정신노동(white collar)’, ‘육체노동(blue collar)’, 학생, 주부, 무직자를 포함한 ‘기타’의 세 집단으로 나누었다. 이때, 각 직업군에 속하는 세부 직업들은 다음의 표와 같다.

<표 V-2> 직업 구분

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

(6) 표본 구성

집락추출 및 조사 영역 배분, 응답자 추출 과정을 통해 표집된 표본은 다음과 같이 구성된다. 평가 영

역 내에서 설문지 유형은 무작위로 순환해서 조사를 진행했기 때문에 유형별 응답자 특성도 비교적 유사한 경향을 보이고 있다.

① 성×연령대×지역별 표본 구성

2차 예비 조사에서도 4개의 지역(서울, 인천, 경기 동부, 경기 읍면부), 성별(남자, 여자), 연령대(20대, 30대, 40대, 50대, 60대)로 총 40개의 구간을 설정하여 구간별로 동일 크기의 표본을 할당하였다. 2차 예비 조사에서는 40개 구간별로 동일한 표본을 조사하도록 조사원을 독려하여 목표 표본에 100% 맞춰 조사를 완료할 수 있었다.

<표 V-3> 2차 예비 조사 성×연령대×지역별 목표 표본

구 분	남	남	남	남	남	여	여	여	여	여	합계
	20대	30대	40대	50대	60대	20대	30대	40대	50대	60대	
합계	24	24	24	24	24	24	24	24	24	24	240
서울	6	6	6	6	6	6	6	6	6	6	60
인천	6	6	6	6	6	6	6	6	6	6	60
경기 동부	6	6	6	6	6	6	6	6	6	6	60
경기 읍면부	6	6	6	6	6	6	6	6	6	6	60

<표 V-4> 2차 예비 조사 성×연령대×지역별 조사 완료 표본

구 분	남	남	남	남	남	여	여	여	여	여	합계
	20대	30대	40대	50대	60대	20대	30대	40대	50대	60대	
합계	24	24	24	24	24	24	24	24	24	24	240
서울	6	6	6	6	6	6	6	6	6	6	60
인천	6	6	6	6	6	6	6	6	6	6	60
경기 동부	6	6	6	6	6	6	6	6	6	6	60
경기 읍면부	6	6	6	6	6	6	6	6	6	6	60

② 지역×직업별 표본 구성

직업 변인은 2021년 인구주택총조사자료를 기준으로 4개 지역별로 비례 할당했다. 지역×직업별 목표 표본과 조사 완료 표본은 아래 <표 V-5>, <표 V-6>과 같으며, 구간별로 일부 차이를 허용하여 조사하였다.

<표 V-5> 2차 예비 조사 지역×직업별 목표 표본

구분	정신노동	육체노동	기타	합계
합계	73	94	73	240
서울특별시	22	19	19	60
인천광역시	16	25	19	60
경기 동부	20	22	18	60
경기 읍면부	15	28	17	60

<표 V-6> 2차 예비 조사 지역×직업별 조사 완료 표본

구분	정신노동	육체노동	기타	합계
합계	72	95	73	240
서울특별시	21	20	19	60
인천광역시	15	26	19	60
경기 동부	20	23	17	60
경기 읍면부	16	26	18	60

③ 지역×학력별 표본 구성

학력 변인도 마찬가지로 2021년 인구주택총조사자료를 기준으로 4개 지역별로 비례 할당했다. 지역×학력별 목표 표본과 조사 완료 표본은 아래 <표 V-7>, <표 V-8>과 같으며, 이 역시 구간별로 일부 차이를 허용하여 조사하였다.

<표 V-7> 2차 예비 조사 지역×학력별 목표 표본

구분	중졸 이하	고졸	대재 이상	합계
합계	21	86	133	240
서울	4	16	40	60
인천	6	24	30	60
경기 동부	4	20	36	60
경기 읍면부	7	26	27	60

<표 V-8> 2차 예비 조사 지역×학력별 조사 완료 표본

구분	중졸 이하	고졸	대재 이상	합계
합계	23	88	129	240
서울	5	15	40	60
인천	7	24	29	60
경기 동부	5	21	34	60
경기 읍면부	6	28	26	60

④ 검사지 유형별 표본 구성

8종의 검사지 유형별로 지역, 성별, 연령대, 직업, 학력이 고르게 조사될 수 있도록 표본을 할당하였으며, 검사지 유형별 표본 역시 구간별로 일부 차이를 허용하여 조사하였다. 목표 표본과 조사 완료 표본은 아래 <표 V-9>, <표 V-10>과 같다.

<표 V-9> 2차 예비 조사 검사지 유형별 목표 표본

구분		검사지1	검사지2	검사지3	검사지4	검사지5	검사지6	검사지7	검사지8	합계
합계		30	30	30	30	30	30	30	30	240
지역	서울	7	8	7	8	7	8	7	8	60
	인천	8	7	8	7	8	7	8	7	60
	경기 동부	7	8	8	7	7	8	8	7	60
	경기 읍면부	8	7	7	8	8	7	7	8	60
성별	남자	15	15	15	15	15	15	15	15	120
	여자	15	15	15	15	15	15	15	15	120
연령대	20대	6	6	6	6	6	6	6	6	48
	30대	6	6	6	6	6	6	6	6	48
	40대	6	6	6	6	6	6	6	6	48
	50대	6	6	6	6	6	6	6	6	48
	60대	6	6	6	6	6	6	6	6	48
직업	정신노동	10	9	9	10	9	8	9	9	73
	육체노동	11	12	12	11	12	12	12	12	94
	기타	9	9	9	9	9	10	9	9	73
학력	중졸 이하	2	3	3	3	2	3	2	3	21
	고졸	11	10	11	11	11	10	11	11	86
	대재 이상	17	17	16	16	17	17	17	16	133

<표 V-10> 2차 예비 조사 검사지 유형별 조사 완료 표본

구분		검사지1	검사지2	검사지3	검사지4	검사지5	검사지6	검사지7	검사지8	합계
합계		30	30	30	30	30	30	30	30	240
지역	서울	7	8	7	8	7	8	7	8	60
	인천	8	7	8	7	8	7	8	7	60
	경기 동부	7	8	8	7	7	8	8	7	60
	경기 읍면부	8	7	7	8	8	7	7	8	60
성별	남자	15	15	15	15	15	15	15	15	120
	여자	15	15	15	15	15	15	15	15	120
연령대	20대	5	7	5	7	8	4	5	7	48
	30대	8	4	5	7	5	7	5	7	48
	40대	8	5	4	7	4	4	8	8	48
	50대	5	6	8	5	5	7	8	4	48
	60대	4	8	8	4	8	8	4	4	48
직업	정신노동	11	9	8	9	7	5	12	11	75
	육체노동	13	11	15	9	12	14	11	10	95
	기타	6	10	7	12	11	11	7	9	70
학력	중졸 이하	1	4	4	3	3	4	2	2	20
	고졸	13	10	8	11	13	12	11	10	89
	대재 이상	16	16	18	16	14	14	17	18	131

2. 2차 예비 조사 진행

1) 조사 방법

본 조사는 태블릿 PC를 활용한 1:1 대인면접조사(TAPI: Tablet Assisted Personal Interview)로 진행됐으며, 태블릿 PC의 디지털 검사지를 통해 자료를 수집했다. 디지털 검사지 구동을 위한 소프트웨어는 안드로이드 OS 기반의 Nfield CAPI 프로그램을 이용했다. Nfield CAPI는 NIPO software에서 만든 조사 솔루션으로 오프라인 상태에서 조사가 진행되기 때문에 인터넷망의 접속이나 속도 문제로 인해 조사가 방해받는 현상을 막을 수 있다. 대신 현장에서 수집된 자료는 1차적으로 태블릿 PC에 저장되고, 나중에 인터넷 연결과 프로그램 동기화 과정을 통해 중앙 서버로 데이터가 전송되는 방식으로 작동한다.

2) 조사원 선발 및 교육

이번 조사는 응답자들이 일반 국민들이라는 점, 설문 조사라기보다는 시험과 같은 성격을 띠고 있다는 점, 응답자에게 말하기 등을 요구하는 점 때문에 조사의 곤란도가 매우 높다고 할 수 있다. 이를 고려하여 2017년과 2018년에 국어능력 실태 조사에 참여한 경험이 있는 조사원과 TAPI 조사 경험이 있는 경력 10년 이상의 조사원을 우선 선발하였으며, 총 18명을 선발하여 2차 예비 조사에 투입하였다. 이 중 1차 예비 조사에 참여한 인원은 12명이며, 6명을 추가로 선발하였다. 조사원 18명이 240건의 설문을 진행하여 조사원 1인당 진행 표본 수는 약 13명이었다.

조사원 교육은 여의도에 위치한 KANTAR PUBLIC 교육장에서 2022년 11월 22일에 2시간가량 진행하였으며, 담당 연구원이 직접 교육을 진행했다. 조사원 교육은 인쇄된 설문지를 이용해 문항 구조와 개념을 이해하고, TAPI 시스템을 시연하는 방식으로 진행했다.

2022년 국민의 국어능력 평가 실태조사 예비조사 (2차) 조사원 지침서

02 설문 진행 방법 : 듣기/말하기

03 듣기/말하기 영역 - 제한 시간: 5분(지문 듣기 시간 포함)

- 지문을 듣고 응답자가 말하는 영역으로 듣기와 말하기를 동시에 평가하는 영역입니다.
- 피말림 상형사건 소리가 잘 들리는지 확인 후, 이어폰을 착용하여 응답자가 적당하다고 생각하는 수준으로 소리를 조절합니다. (이 때 소리 조절은 듣기 문항 인쇄 후드 부분을 이용하여 조절)
- 기기 이상과 문제가 발생한 경우를 제외하고, 모든 문항은 반드시 한 번만 들리고 말아야 합니다.
- 녹음이 잘 될 수 있도록 조용한 주변 환경이 유지될 수 있도록 합니다.
- 각 문항 모두 준비시간 5초와 말하는 시간 20초~30초로 정해져 있으니 반드시 시간을 지키도록 합니다.
- 녹음이 완료되면 녹음 상태를 반드시 확인해야 하고, 저장에 잘 되어야 합니다.
- 녹음 버튼을 누르면 녹음 시작, 종결한 버튼 한번 더 누르면 녹음 저장입니다. 만약 녹음 종료 후 녹음 버튼을 또 누르면 기종의 녹음이 삭제됩니다.
- 즉시 모뎀 상황에 대비하여 개인 휴대전화로 동시에 녹음을 진행합니다(듣기/말하기 영역 시작부터 말하기 영역 끝날 때까지 한번에 녹음, 메모장으로 간략하여 동시에 녹음)

STEP 01 문항별 5초의 준비시간 확보 (이 때 조사원은 간단히 녹음 테스트)

STEP 02 20초~30초 간 응답자가 말하는 내용 녹음 (응답자 이름과 문항번호를 말하고 시작)

STEP 03 저장 후 녹음상태 확인 필수

[그림 V-1] 2차 예비 조사 조사원 지침서

<표 V-11> 영역별 조사 진행 방법 및 지침

영역	조사 진행 방법 및 숙지 사항
듣기	● 파일을 실행시켜 소리가 잘 들리는지 확인 후, 이어폰을 전달하여 응답자에게 적당한 수준으로 소리 크기를 조절함 (소리 크기는 듣기 문항 안내 멘트 부분을 이용하여 조절) ● 기기 이상의 문제가 발생한 경우를 제외하고, 모든 문항을 반드시 한 번만 들려주어야 한다는 원칙 엄수
말하기	● 녹음이 잘 될 수 있도록 조용한 주변 환경이 유지될 수 있도록 함 ● 각 문항별로 준비 시간 1~2분과 제공 및 녹음 테스트 ● 2분간 응답자의 응답 내용 녹음 (응답자 ID와 문항 번호를 말하고 응답 시작) ● 녹음 버튼을 누르면 녹음 시작, 동일한 버튼을 한 번 더 누르면 녹음 저장(만약 녹음 종료 후 녹음 버튼을 또 누르면 기존의 녹음이 삭제되므로 주의) ● 녹음 완료 후 녹음 상태 확인 및 저장 후 다음 말하기 문항 진행 ● 혹시 모를 상황에 대비하여 개인 휴대전화로 동시에 녹음 진행
듣기 말하기	● 파일을 실행시켜 소리가 잘 들리는지 확인 후, 이어폰을 전달하여 응답자에게 적당한 수준으로 소리 크기를 조절함 (소리 크기는 듣기 문항 안내 멘트 부분을 이용하여 조절) ● 기기 이상의 문제가 발생한 경우를 제외하고, 모든 문항을 반드시 한 번만 들려주어야 한다는 원칙 엄수 ● 녹음이 잘 될 수 있도록 조용한 주변 환경이 유지될 수 있도록 함 ● 각 문항별로 준비 시간 5초 제공 및 녹음 테스트 ● 20~30초간 응답자의 응답 내용 녹음 (응답자 ID와 문항 번호를 말하고 응답 시작) ● 녹음 버튼을 누르면 녹음 시작, 동일한 버튼을 한 번 더 누르면 녹음 저장(만약 녹음 종료 후 녹음 버튼을 또 누르면 기존의 녹음이 삭제되므로 주의) ● 녹음 완료 후 녹음 상태 확인 및 저장 후 다음 말하기 문항 진행 ● 혹시 모를 상황에 대비하여 개인 휴대전화로 동시에 녹음 진행

3) 2차 예비 조사 진행 일정

<표 V-12> 2차 예비 조사 진행 일정

단계	일정
표본 설계 및 설문지 설계	~ 11월 15일
TAPI 시스템 구축	11월 16일 ~ 11월 21일
조사원 교육	11월 22일
TAPI 시스템 최종 점검 및 보완	11월 22일
현장 조사 진행	11월 23일 ~ 12월 9일
사후 검증	12월 2일 ~ 12월 9일
통계 처리	12월 10일 ~ 12월 12일
결과 보고	12월 12일

4) 조사 검증

각 조사원이 조사한 표본 중 20% 이상 검증하는 것을 목표로 표본을 무작위로 추출해 전화 검증을 진행했다. 전화 검증은 별도의 검증 요원이 응답자에게 직접 전화를 걸어 조사 참여 여부, 주요 인구특성 등을 재확인하는 과정으로 진행했다.

2차 예비 조사 완료 표본 240건 중 155건에 대해 검증을 시도하여 통화 시도율은 65%였으며, 이 중 검증을 완료한 표본은 53건으로 검증률은 22%다. 세부 검증 결과는 다음과 같다.

<표 V-13> 2차 예비 조사 검증 결과

구분	표본 수
전체	240
A. 검증 설문지 총수	155
B. 검증 완성 설문지	53
1. 인터뷰했음	53
C. 검증 미완성 설문지 수	102
2. 인터뷰한 적 없음	-
3. 인터뷰했으나 검증 질문의 응답 틀림	-
4. 응답 거절(끊음, 통화 곤란 등)	-
5. 응답자 부재중	1
6. 통화 중	-
7. 전화 안 받음 / 10. 자동응답전화	97
8. 번호 틀림	-
9. 결번	-
11. 기타 (나중에, 회의 중)	4
통화 시도율(B/A)	65%
검증률(B/전체 표본)	22%

일반적인 검증 절차와는 별도로 말하기 녹음 파일 전수를 다시 청취해 녹음 상태가 불량하거나 누락된 경우가 있는지 추가로 확인하는 과정도 거쳤다. 녹음 파일 1,200개 전체에 대한 검수 과정을 거쳤으며, 이 중 3건의 녹음 파일 오류가 발견되어 재조사를 진행하였다.

5) 결과물 제출

‘듣기’ 영역의 조사 결과는 SPSS 데이터와 엑셀(EXCEL) 데이터, 주요 인구학적 변인과의 교차분석 통계표 형태로 제출하였다. ‘말하기’와 ‘듣기·말하기’ 영역은 응답자의 말하기 내용을 녹음한 디지털 음성파일(MP3)로 제출하였다.

3. 2차 예비 조사 관련 특이 사항

1) TAPI 활용 관련

1차 조사에서와 마찬가지로 일부 조사원의 조작 실수로 녹음 버튼을 한 번 더 누르게 되면, 기존 파일이 삭제되고 새로운 파일이 생성되며 응답 결과가 소실되는 문제가 발견되었다. 이에 조사원의 휴대폰 녹음 기능을 이용해 말하기 조사 과정을 동시 녹음하는 방식으로 응답 파일 유실에 대비했다. 그리고 실사 관리자가 사후 검증 차원에서 말하기 음성 파일 1,200개를 전수 재청취했는데, 이로 인해 조사 종료 후 데이터 점검 과정에 많은 시간이 소요되는 문제가 있었다. 대규모의 본 조사를 진행하기 위해서는 예산을 추가로 투입하여 사후 데이터 점검 과정까지 고려한 시스템 도입이 필요해 보인다.

2) 실사 과정상의 문제점

2차 예비 조사 역시 1차 조사와 표본 수, 표본 할당은 동일하였다. 5개 변인과 검사지 유형별로 표본을 할당하여 표본을 선정하는 과정에서 여전히 어려움이 있었으며, 녹음 자체에 대한 거부감으로 인해 조사를 거절한 경우도 많았다. 후반부에 말하기 영역을 제시하였기 때문에 응답자의 부담감과 피로감도 더 컸으며, 이로 인해 말하기 영역의 응답 질이 일부 떨어지는 문제가 있었을 것으로 예측된다. 추후 본 조사에서는 문항 수 조절 혹은 말하기 영역을 앞으로 배치하는 등의 조치가 필요해 보인다.

4. 2차 예비 조사 결과

1) 2차 예비 조사 결과의 채점

(1) 선택형 문항(듣기)의 채점

2차 예비 조사는 태블릿 PC를 활용했기 때문에 응답자가 화면상에서 특정 보기를 터치하면, 응답 결과가 곧바로 CAPI 시스템에 입력되는 방식으로 진행했다. 따라서 종이 설문지를 이용한 기존 조사처럼 별도의 전산 입력 과정을 거치거나, OMR 카드 사용 등을 병행할 필요가 없었다. 현장 조사 종료 후 서울대 국어교육연구소로부터 각 문항의 정답을 제공받았고, 이를 통계 처리된 데이터와 비교하여 각 문항별 정답률을 계산했다.

(2) 수행형 문항(듣기·말하기, 말하기)의 채점

2차 예비 조사의 주관식 문항은 (가), (나)형 2종으로 출제되었으며, 전체 문항 유형은 1차 조사와 마찬가지로 가교 문항을 포함하여 모두 5개로 구성되었다. 2차 예비 조사에도 240명의 피험자가 참여하였

으며, 총 1,200건의 응답이 수집되어 이를 채점 대상으로 확정하였다. 2차 조사의 채점은 본 연구의 목적 달성을 위해 다음과 같이 진행하였다.

첫째, 채점자 집단 간 가교 설계를 적용하여 문항 유형별로 채점자 영향을 정교하게 확보할 수 있도록 한다.

둘째, 각 문항 유형별로 전문가 3인의 공동 채점을 위해 전체 5개 문항 유형에 해당하는 15인의 채점단을 확보한다. 채점단은 국어교육학 전공의 석사 및 박사 과정생으로서 국어교육 전문 인력으로 구성한다.

셋째, 채점 대상 문제 및 채점 척도 사용 수준의 높이를 맞추기 위한 웹 기반 훈련¹¹⁾에 모든 채점자가 반드시 참여하며, 훈련 과정에서 실시한 사전 채점 및 문항 유형별 협의, 종합 협의를 통해 세부적인 채점 실행 사항을 점검한다.

<표 V-14> 수행형 문항의 영역 및 문항, 채점자 구성

영역	문항 유형 및 응답 수		채점자
듣기.말하기 (LS)	1번	(가)형(LS1-가) 120개 응답	오지은, 이혜린, 윤현아 권아람, 박장범, 안이진
		(나)형(LS1-나) 120개 응답	
	2번	(가)형(LS2-가) 120개 응답	오지은, 이혜린, 윤현아 최혜진, 임지원, 송미지
말하기 (S)	3번	가교(S1) 240개 응답	최혜진, 임지원, 송미지 김다연, 이은석, 이민경
	4번	(가)형(S2-가) 120개 응답	김동섭, 권근영, 이정민 김다연, 이은석, 이민경
		(나)형(S2-나) 120개 응답	
	5번	(나)형(S3-나) 120개 응답	김동섭, 권근영, 이정민 권아람, 박장범, 안이진

2차 예비 조사의 문항 유형별 채점자 집단은 1차 조사와 마찬가지로 화법 및 문법, 문학 전공의 서울대 대학원 석사 및 박사 과정생으로 구성하였으며, 채점자 집단 설계에서 전공 특성 영향을 상쇄하기 위해 전체 5개 집단 모두 화법, 문법, 문학 전공자를 각각 한 명씩 배치하였다.

2차 예비 조사 채점에서는 1차 예비 조사 채점을 통해 확보한 문항별 채점 전문성을 바탕으로 보다 엄밀한 문항 양호도 추정을 위해 5개의 채점자 집단을 가교 문항으로 연계하는 설계 방법을 적용하였다. 이는 1차 예비 조사의 단일 문항 유형 채점과 달리 2개 문항 유형을 채점하는 방식으로, 채점자 영향을 고려하여 듣기·말하기와 말하기 영역 문항 간에 비교 가능성을 확보할 수 있다는 장점이 있다. 채점

11) 2022년 12월 23일(금) 19시 30분부터 21시 30분까지 비대면 화상 회의(zoom) 방식으로 실시하였음.

대상 자료는 듣기·말하기 1,440건, 말하기 2,160건이었으며, 2차 예비 조사의 본 채점은 다음과 같이 진행되었다.

첫째, 각 문항 유형 및 검사지별 세부 문제에 대한 피험자 응답 파일을 분류한다.

둘째, 채점 수행의 신뢰도 확보를 위해 문항 유형별로 채점 척도를 바탕으로 워크숍 기반의 채점자 훈련을 실시한다.

셋째, 채점자들의 일관된 채점 척도 적용을 위해 각 문항 유형별로 설정한 내용, 조직, 표현, 전달의 채점 준거별 세부 수행 항목을 제시한다.

넷째, 채점의 효율적인 진행과 채점자 피로도 관리를 위해 본 채점은 1차(50%), 2차(50%)로 나누어 진행하며, 채점의 결과를 실시간으로 확인 및 관리하기 위해 채점자는 개인화된 웹 채점표 링크에 접속하여 점수를 입력한다.

① 듣기·말하기 영역

2차 예비 조사에는 1차 예비 조사 결과에 따라 2개 문항 유형, 세부 영역별로 (가)형과 (나)형의 총 3개 문제가 포함되었다. 이 영역에서 확보된 응답은 총 480건이며, 문제별로 내용, 조직, 표현, 전달 준거별로 비중에 따라 3~9점, 2~6점 척도로 수준을 나누어 채점을 진행하였다. 듣기·말하기 문항의 점수 척도는 말하기 문항과 달리 실제성을 고려하여 피험자의 응답이 즉각적으로 이루어진다는 점과 이에 따라 간단한 형태로 응답이 이루어진다는 점, 세밀한 척도를 적용할 경우 변별이 어렵다는 점 등을 고려하여 구성하였다.

<표 V-15> 듣기·말하기 1번 (가)형 문항 채점 척도

조사 항목		채점 요소	채점 척도(점수)			점수 비중	
			미흡	보통	우수		
내용 구성	내용 선정	- 문제에서 요구하는 응답을 한다. (배려) - 상대의 부담을 최소화하고, 합리적인 대안을 제시한다.	3	6	9	30%	60%
	조직	- 응답의 흐름이 배려를 나타내는 데 적합하다. - 발화의 시작과 마무리가 자연스럽다.	3	6	9	30%	
표현 및 전달	표현	- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.	2	4	6	20%	40%
	전달	- 발음, 속도, 어조 등이 자연스럽다(유창성). - 발음이 명료하다(정확성).	2	4	6	20%	

<표 V-16> 듣기·말하기 1번 (나)형 문항 채점 척도

조사 항목		채점 요소	채점 척도(점수)			점수 비중	
			미흡	보통	우수		
내용 구성	내용 선정	- 문제에서 요구하는 응답을 한다. (사과) - 잘못된 내용에 대해 언급하고, 상대의 기분을 헤아려 말한다.	3	6	9	30%	60%
	조직	- 응답의 흐름이 사과하는 데 적합하다. - 발화의 시작과 마무리가 자연스럽다.	3	6	9	30%	
표현 및 전달	표현	- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.	2	4	6	20%	40%
	전달	- 발음, 속도, 어조 등이 자연스럽다(유창성). - 발음이 명료하다(정확성).	2	4	6	20%	

<표 V-17> 듣기·말하기 2번 (가)형 문항 채점 척도

조사 항목		채점 요소	채점 척도(점수)			점수 비중	
			미흡	보통	우수		
내용 구성	내용 선정	- 문제에 부합하는 적절한 응답을 한다. - 문제에 대한 설명과 해결을 위한 반응이 잘 드러난다(적극성).	3	6	9	30%	60%
	조직	- 응답이 이전 발화의 흐름에 잘 대응된다. - 발화의 시작과 마무리가 자연스럽다.	3	6	9	30%	
표현 및 전달	표현	- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.	2	4	6	20%	40%
	전달	- 발음, 속도, 어조 등이 자연스럽다(유창성). - 발음이 명료하다(정확성).	2	4	6	20%	

② 말하기 영역

말하기 영역에서도 1차 예비 조사의 결과를 바탕으로 2018년 조사와의 동등화 가능성을 마련하기 위해 선정한 가교 문항(1개)과 신규 출제된 문항 유형(1개), 동형 문항(1개)으로 구성하였다. 이 중에서 가교 문항은 1차 예비 조사와 같은 문항을 사용하였으며, 신규 출제된 문항은 공적 상황에서의 발표하는 말하기를 조사하는 문항이다. 동형 문항은 공적 상황에서 논제에 대해 자신의 주장을 밝히는 문항이다. 2차 조사에서의 말하기 영역 채점은 기존 조사와 마찬가지로 5 수준의 척도(3-6-9-12-15, 2-4-6-8-10)

로 이루어졌다.

<표 V-18> 말하기 1번 가교 문항 채점 척도

조사 항목		채점 요소	채점 척도(점수)					점수 비중	
			매우 미흡	미흡	보통	우수	매우 우수		
내용 구성	내용 선정	- 질문에서 요구하는 대답을 한다. - 청자가 이해하기 쉽게 구체적으로 설명한다.	3	6	9	12	15	30%	60%
	조직	- 내용의 흐름이 설명에 적합하다. - 시작과 마무리가 자연스럽다.	3	6	9	12	15	30%	
표현 및 전달	표현	- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.	2	4	6	8	10	20%	40%
	전달	- 발음, 속도, 어조가 적절하다. - 적절한 시간을 지켜 말한다.	2	4	6	8	10	20%	

<표 V-19> 말하기 2번 문항 채점 척도

조사 항목		채점 요소	채점 척도(점수)					점수 비중	
			매우 미흡	미흡	보통	우수	매우 우수		
내용 구성	내용 선정	- 슬라이드의 내용을 충분히 설명한다. - 복잡한 내용을 간명하게 정리하여 말한다.	3	6	9	12	15	30%	60%
	조직	- 발표의 흐름이 자연스럽다. - 발화의 시작과 마무리가 효과적이다.	3	6	9	12	15	30%	
표현 및 전달	표현	- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.	2	4	6	8	10	20%	40%
	전달	- 발음, 속도, 어조가 등이 자연스럽다(유창성). - 발음이 명료하다(정확성).	2	4	6	8	10	20%	

<표 V-20> 말하기 3번 문항 채점 척도

조사 항목		채점 요소	채점 척도(점수)					점수 비중	
			매우 미흡	미흡	보통	우수	매우 우수		
내용 구성	내용 선정	- 문제에서 요구하는 응답을 한다. - 문제 상황에 대한 설명과 주장의 근거가 잘 드러난다.	3	6	9	12	15	30%	60%
	조직	- 내용의 흐름이 논증에 적합하다. - 발화의 시작과 마무리가 효과적이다.	3	6	9	12	15	30%	
표현 및 전달	표현	- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.	2	4	6	8	10	20%	40%
	전달	- 발음, 속도, 어조가 등이 자연스럽다(유창성). - 적절한 시간을 지켜 말한다(10초 미만: 매우 미흡, 30초: 미흡, 30초 이상은 감점 없음.)	2	4	6	8	10	20%	

(3) 수행형 문항(듣기·말하기, 말하기) 채점의 채점자 신뢰도 및 영향 분석

① 듣기·말하기 영역

<표 V-21> 2차 예비 조사 듣기·말하기 영역의 채점자 간 신뢰도

영역	문항 유형		채점자	급내상관	df	유의확률
듣기·말하기 (LS)	1번	LS1-가 LS1-나	KR011	.816	479	.000
			KR012			
			KR013	.877	479	.000
			KR051			
	2번	LS2-가	KR052	.791	479	.000
			KR053			
			KR011	.798	479	.000
			KR012			
			KR013	.798	479	.000
			KR021			
			KR022			
			KR023			

2차 예비 조사 듣기·말하기 영역의 채점자 신뢰도를 알아보기 위해 급내상관을 확인하였다. 채점자 간 신뢰도 분석 결과, 듣기·말하기 1, 2번 유형 모두 급내상관이 .7 이상으로 양호한 수준으로 나타났으며, 두 문항 유형 모두 전체 채점자 집단에서 1차 예비 조사에 비해 향상된 신뢰도 수준을 나타내었다.

<표 V-22> 2차 예비 조사 듣기·말하기 영역의 채점자 영향

영역	문항 유형		채점자	logit	SE
듣기. 말하기 (LS)	1번	LS1-가 LS1-나	KR011	.446	.037
			KR012	.155	.039
			KR013	.108	.039
			KR051	.654	.042
			KR052	.820	.041
			KR053	.258	.041
	2번	LS2-가	KR011	.446	.037
			KR012	.155	.039
			KR013	.108	.039
			KR021	-.308	.162
			KR022	.113	.162
			KR023	-2.247	.192

2차 예비 조사 듣기·말하기 문항 채점의 채점자 영향을 알아보기 위해 다분 반응에 대한 IRT 분석 프로그램인 PASCAL을 이용하여 채점자 모수 분석을 실시하였다. 분석 결과 1번 문항 유형을 담당한 채점자들은 보통 수준의 채점을 수행하였으며, 2번 문항에서는 대체로 매우 관대하게 채점을 한 것으로 나타났다. 이는 1차 예비 조사에 비해서는 관대한 정도가 감소한 수준이었다. 2번 문항 유형을 담당한 KR023의 경우 듣기·말하기 영역에서 가장 관대한 경향을 갖고 있는 것으로 나타났다. 관대한 채점 경향이 채점 척도 사용의 문제와 관련이 되었는지 알아보기 위해 다국면라쉬모형(MFRM) 분석을 통해 채점 척도 사용의 적합도를 확인하였다. 그 결과, KR023의 경우 척도 사용의 일관성과 수준 판별의 일관성이 부족한 것으로 나타났다. 듣기·말하기 영역에서 1차 예비 조사와 2차 예비 조사에서 나타난 공통적인 채점 특성은 관대한 채점 경향이었으며, 일부 채점자의 경우 채점 척도 사용의 어려움을 나타내기도 하였다. 이상의 채점 경향 분석 결과를 바탕으로 향후 본 조사의 듣기·말하기 문항 유형 채점에서는 다음의 방안을 고려할 필요가 있다.

첫째, 듣기·말하기 문항의 경우 채점자들의 관대성 경향이 높게 나타날 수 있으며, 이는 현재 사용하고 있는 채점 척도를 구성하는 채점 준거와 수준 척도에서 비롯된 것으로 볼 수 있다. 특히 채점자들에 의해 보통 수준으로 판정되는 응답들에 대해서도 보다 엄격한 채점이 이루어질 수 있도록 할 필요가 있다. 둘째, 현재 듣기·말하기 문항에 적용하고 있는 점수 척도는 말하기 문항과 달리 3점 척도로 구성되어 있으며, 이로 인해 채점자의 판정 경향이 5점 척도에 비해 편향되어 나타날 가능성이 높다. 따라서 이를 보완하기 위해 평가 준거의 수준별 기준 기술을 보완하는 동시에 점수 척도를 4점 혹은 5점 척도로 변경하는 방안을 고려할 필요가 있다. 셋째, 채점자들은 채점자 간 신뢰도 측면에서 1차 예비 조사 채점에 비해 2차 예비 조사 채점에서 향상된 결과를 나타내었으며, 이는 반복된 채점 경험과 워크숍 방식의 채점자 교육의 효과가 있었음을 의미한다. 따라서 향후 진행될 본 조사에서도 채점자 간 신뢰도 확보를 위해 이와 같은 채점자 선정 및 훈련 방법을 고려할 필요가 있다.

② 말하기 영역

2차 예비 조사의 채점 설계에 따라 말하기 영역의 채점자 신뢰도를 알아보기 위해 급내상관을 확인하였다. 분석 결과, 말하기 1, 2, 3번 유형 모두 급내상관이 .8 이상으로 채점자 간 신뢰도가 우수한 수준인 것으로 나타났다. 이를 통해 모든 말하기 문항 유형에서 신뢰할 만한 수준으로 채점이 이루어졌음을 확인하였다.

<표 V-23> 2차 예비 조사 말하기 영역의 채점자 간 신뢰도

영역	문항 유형		채점자	급내상관	df	유의확률
말하기 (S)	1번	S1(가교)	KR021	.904	479	.000
			KR022			
			KR023			
			KR031	.922	479	.000
			KR032			
			KR033			
	2번	S2-가 S2-나	KR031	.905	479	.000
			KR032			
			KR033			
			KR041	.908	479	.000
			KR042			
			KR043			
	3번	S3-나	KR041	.916	479	.000
			KR042			
			KR043			
			KR051	.897	479	.000
			KR052			
			KR053			

<표 V-24> 2차 예비 조사 말하기 영역의 채점자 영향

영역	문항 유형		채점자 집단	logit	SE
말하기 (S)	1번	S1(가교)	KR021	-.108	.032
			KR022	.280	.033
			KR023	-.004	.032
			KR031	-.067	0.18
			KR032	.276	.021
			KR033	-.097	.020
	2번	S2-가 S2-나	KR031	-.067	0.18
			KR032	.276	.021
			KR033	-.097	.020
			KR041	.099	.017
			KR042	-.236	.016
			KR043	.021	.017
	3번	S3-나	KR041	.099	.017
			KR042	-.236	.016
			KR043	.021	.017
			KR051	-.267	.023
			KR052	.293	.024
			KR053	-.191	.022

2차 예비 조사 말하기 문항 채점 결과에 대한 채점자 효과를 알아보기 위해 다분 반응에 대한 IRT 분석 프로그램인 PASCAL을 이용하여 채점자 모수 분석을 실시하였다. 분석 결과 말하기 1, 2, 3번 문항 유형을 담당한 채점자들은 대체로 다소 관대한 경향으로 채점한 것으로 나타났다. 이는 1차 예비 조사에서의 엄격한 경향이 보완된 결과인데, 이러한 결과는 각 문항 유형별로 새로운 채점자들이 추가된 점, 채점자 훈련을 통해 엄격성 조정에 대한 협의가 이루어진 점 등에 따른 것으로 판단된다. 말하기 문항 채점자 중 KR052의 경우 담당한 말하기 3번 문항 유형에서 가장 엄격하게 채점을 하였으며, 반대로 말하기 2번 문항 유형을 맡은 KR051의 경우 가장 관대한 경향으로 채점을 한 것으로 나타났다.

이상의 듣기·말하기 및 말하기 문항의 2차 예비 조사 채점자 간 신뢰도 및 채점자 영향 분석을 바탕으로 본 조사 채점에 대해 얻은 시사점은 다음과 같다.

첫째, 워크숍 기반 채점자 교육은 듣기·말하기 및 말하기 문항 유형 모두 채점자 간 신뢰도 확보의 측면에서 기여할 수 있다. 이는 1·2차 채점자 워크숍에서의 채점 연습 결과 및 예비 조사 채점자 간 신뢰도 수치를 통해 확인되었다. 따라서 본 조사의 수행형 문항 채점자 교육에서도 그룹 내 협의에 기반하는 채점자 훈련 방식을 반영할 필요가 있다.

둘째, 각 문항 유형별로 가교 문항 채점 설계를 적용한 결과, 듣기·말하기 영역과 말하기 영역에서 1차 예비 조사에 비해 채점 경향이 상쇄되는 효과가 일부 나타났다. 이는 각 문항 유형별로 채점자의 다양성이 증가하면서 나타난 현상으로 볼 수 있으나, 종합해 볼 때 개인의 채점 특성이 지나치게 관대한 경우에는 그 효과가 미미하였다. 특히 듣기·말하기 영역의 경우에는 이러한 측면이 제한된 채점 척도의 사용과 관련이 있는 것으로 판단되며, 따라서 본 조사에서는 해당 영역에 대한 채점 척도 사용의 적절성을 재고할 필요가 있다.

2) 2차 예비 조사 결과 분석

(1) 검사 설계상의 특징

<표 V-25> 듣기 영역 검사지별 문항 구성

검사지	문항 구성			문항 수	응시자 수
	1, 2번 (공통/가교)	3~15번(가)	3~15번(나)		
듣기(가)	포함	포함		15	120
듣기(나)	포함		포함	15	120

<표 V-26> 듣기·말하기, 말하기 영역 검사지별 문항 구성

검사지	문항 구성					문항 수	응시자 수
	듣기 ·말하기(가)	듣기 ·말하기(나)	말하기(가)	말하기(나)	3번 (가교)		
듣·말(가) /말(가)	포함		포함		포함	5	60
듣·말(가) /말(나)	포함			포함	포함	5	60
듣·말(나) /말(가)		포함	포함		포함	5	60
듣·말(나) /말(나)		포함		포함	포함	5	60

검사의 설계상 영역별 원점수 총점은 검사지에 따라서 다른 의미를 가진다. 각 검사지에 포함된 문항이 지니는 변별도나 곤란도에 다소 차이가 있게 마련이기 때문이다. 따라서 각기 다른 검사지로 시험을 치른 피험자들의 원점수 총점을 비교하는 것은 적절하지 않고, 문항반응이론을 적용하여 피험자의 능력을 추정하는 방식으로 비교가 가능하다. 특히 이번 2차 예비 조사에서는 듣기·말하기 및 말하기 영역의 문항 양호도 추정에도 곤란도 모수와 변별도 모수를 함께 고려하는 2모수 IRT 분석 모형을 적용하여 문항 간 비교를 위한 객관화된 자료를 확보하였다.

(2) 영역별 점수 분포 분석

① 기술통계

<표 V-27> 영역별 점수 분포의 기술통계치

영역	사례 수	최솟값	최댓값	평균	표준편차
듣기	240	27.84	63.59	50.12	8.48
듣기.말하기	240	24.78	72.28	49.82	8.50
말하기	240	27.32	80.91	50.04	8.99

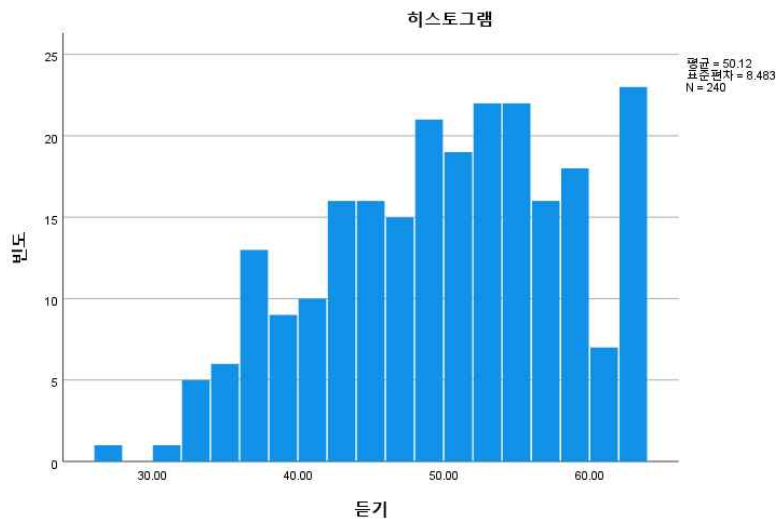
② 영역별 점수 분석 방법

선택형 문항으로 구성된 듣기 영역의 경우, 2모수 문항반응이론 모형을 적용하여 분석하였다. 피험자 능력 모수의 추정치(평균 0, 표준편차 1)를 산출한 후, 이 점수에 10점을 곱하고, 50점을 더하여 평균 50, 표준편차 10인 t점수로 영역별 점수를 최종 산출하였다.

능력 모수 추정치는 표본 집단이 아닌 모집단에 대한 추정치이므로 기술통계 결과표에서 각각의 영역별 점수의 평균과 표준편차는 각각 50점과 10점에서 다소 차이가 있다.

수행형 문항으로 구성된 듣기·말하기 영역과 말하기 영역의 경우, 곤란도와 변별도를 동시에 고려하는 2모수 문항반응이론 일반화 부분 점수 채점자 모형을 사용하여 개별 문항의 채점 준거별로 곤란도와 변별도를 추정하였다. 분석은 PASCAL을 통해 이루어졌으며, 이를 통해 산출한 피험자의 능력 모수치(평균 0, 표준편차 1)는 듣기 영역과 마찬가지로 평균 50, 표준편차 10인 t점수로 최종 산출하였다.

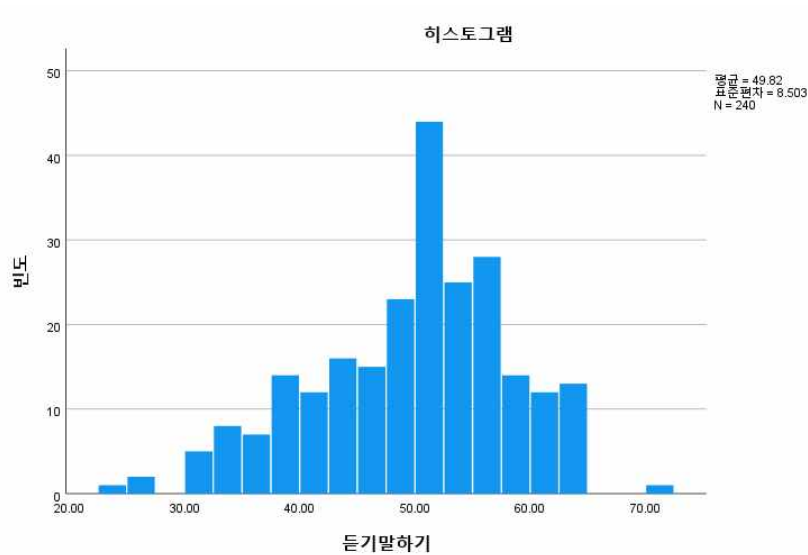
③ 듣기 영역



[그림 V-2] 2차 예비 조사의 듣기 영역 점수 분포

듣기 영역 점수 분포는 정규분포에 가깝지만, 왜도가 -.255로 오른쪽으로 약간 치우치는 부정적 편포를 보이고, 첨도가 -.731로 중심이 다소 낮고 분포가 비교적 퍼져 있는 모양을 보인다. 그럼에도 불구하고 콜모고로프스미르노프(Kolmogorov-Smirnov) 검증 방법으로 정규성 검정을 수행한 결과 유의확률 .064로 분포가 정규분포에 해당한다는 영가설을 채택할 수 있음을 확인하였다.

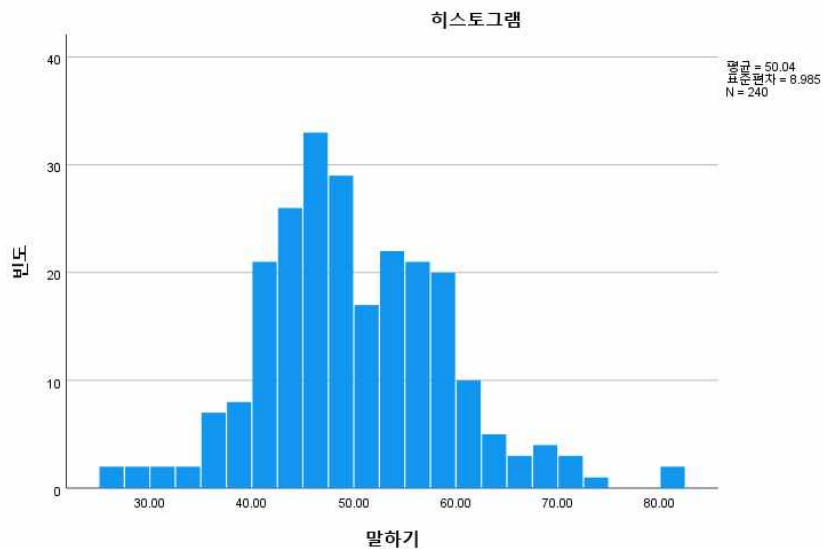
④ 듣기·말하기 영역



[그림 V-3] 2차 예비 조사의 듣기·말하기 영역 점수 분포

듣기·말하기 영역 점수 분포는 정규분포에 가깝지만, 왜도가 -.452로 오른쪽으로 약간 치우치는 부정적 편포를 보이고, 첨도는 -.058로 정규분포에 거의 근접한 모양을 보인다.

⑤ 말하기 영역



[그림 V-4] 2차 예비 조사의 말하기 영역 점수 분포

말하기 영역 점수 분포는 정규분포에 가깝지만, 왜도가 .425로 왼쪽으로 약간 치우치는 정적 편포를 보이고, 첨도가 .770으로 중심이 다소 높고 뾰족한 모양을 보인다.

(3) 영역별 문항 곤란도 및 변별도 분석

① 듣기 영역

<표 V-28> 2차 예비 조사 듣기 영역 문항 곤란도 및 변별도 분석

문항	사례 수	정답률	문항 모수 추정치				적합도 ¹²⁾		
			변별도	SE	곤란도	SE	χ^2	df	p-value
듣기_01(공통)	240	64.6	.342	.173	-1.805	.952	20.858	8	.008
듣기_02(공통)	240	75.8	.763	.221	-1.681	.447	12.052	8	.149
듣기_03(공통)	240	98.3	3.821	3.822	-2.315	.679	4.495	8	.810
듣기_04(공통)	240	79.2	1.182	1.182	-1.481	.371	3.546	8	.896
듣기_05(가)	120	87.5	.676	.390	-23.035	1.600	10.327	8	.243
듣기_06(가)	120	65.8	.884	.315	-.795	.340	10.892	8	.208
듣기_07(공통)	240	100.0	7.238	7.239	-4.185	93.265	0.000	8	1.000
듣기_08(공통)	240	30.8	.126	.126	6.411	12.206	6.663	8	.573
듣기_09(가)	120	66.7	1.914	.618	-.519	.173	6.225	8	.622
듣기_10(가)	120	37.5	1.464	.439	.543	.198	10.708	8	.219
듣기_11(가)	120	76.7	1.182	.411	-1.194	.365	8.856	8	.355
듣기_12(공통)	240	55.8	.817	.817	-.383	.274	5.216	8	.734
듣기_13(가)	120	80.8	1.839	.627	-1.116	.264	4.197	8	.839
듣기_14(가)	120	90.0	2.621	1.168	-1.438	.302	5.533	8	.699
듣기_15(공통)	240	56.7	1.038	1.038	-.373	.228	11.419	8	.179
듣기_05(나)	120	64.2	1.330	1.330	-.647	.214	11.072	8	.198
듣기_06(나)	120	71.7	1.824	1.824	-.857	.196	8.121	8	.422
듣기_09(나)	120	74.2	2.876	2.876	-.844	.162	3.841	8	.871
듣기_10(나)	120	72.5	2.789	2.789	-.788	.160	2.836	8	.944
듣기_11(나)	120	66.7	.887	.887	-.968	.339	3.916	8	.865
듣기_13(나)	120	51.7	.795	.795	-.149	.265	14.402	8	.072
듣기_14(나)	120	76.7	1.397	1.397	-1.212	.279	3.421	8	.905

듣기 영역 28문항에 대하여 2모수 문항반응이론 모형을 적용하여 분석한 결과, 일부 문항에서 양호하다고 보기 어려운 변별도와 곤란도 분석 결과를 보였다.

듣기 8번(공통) 문항의 경우 변별도가 .35 미만으로 매우 낮게 나타났다. 따라서 이 문항은 듣기 능력을 거의 변별하지 못하는 문항으로 볼 수 있다. 다만, 8번(공통) 문항은 모든 문항 중 유일하게 곤란도가 2.0 이상으로 매우 어려운 문항으로 나타났는데, 이러한 결과를 고려하면 이 문항에서 변별도가 낮게 나타난 것은 난도가 지나치게 높았기 때문이라고 판단된다.

듣기 3번(공통), 5번(가), 7번(공통) 문항의 경우 곤란도가 -2.0 미만으로 매우 쉬운 것으로 나타났고,

12) Yen(1981)의 제안에 따라 검정함.

앞서 언급한 것처럼 8번(공통) 문항의 경우 곤란도가 2.0 이상으로 매우 어려운 것으로 나타났다. 검사지에 다양한 곤란도를 지닌 문항이 고루 포함되어야 다양한 능력을 지닌 피험자를 고루 변별할 수 있으므로 검사지에 이들 문항이 일부 포함되는 것은 적절할 수 있으나 지나치게 많이 포함되어서는 안 될 것이다.

이상의 결과를 고려하되, 그 변별도와 곤란도에 특별한 문제가 없는 경우 다양한 곤란도의 문항을 선별하는 데 초점을 두어 최종 문항을 선택하였다. 그리하여 본 조사에 사용할 최종 문항으로는 듣기 1번(공통), 2번(공통), 3번(공통), 4번(공통), 5번(나), 6번(나), 7번(공통), 8번(공통), 9번(나), 10번(가), 11번(가), 12번(공통), 13번(나), 14번(가), 15번(공통)을 선정하였고, 예비 문항으로 6번(가), 9번(가), 10번(나)를 선정하여 1.2배수에 해당하는 18문항을 최종 선정하였다.

② 듣기·말하기 영역

<표 V-29> 2차 예비 조사 듣기·말하기 문항 양호도

문항	준거	사례 수	문항 모수 추정치				단계 모수			
			변별도	SE	곤란도	SE	1	SE	2	SE
듣기·말하기 1-가	내용	120	1.407	.185	-.466	.070	.881	.074	-.881	.068
	조직	120	4.857	2.281	-.809	.060	.595	.038	-.595	.030
	표현	120	.695	.099	-1.833	.129	.997	.214	-.997	.106
	전달	120	.471	.083	-2.523	.226	1.354	.361	-1.354	.150
듣기·말하기 1-나	내용	120	1.075	.128	-.758	.079	1.101	.102	-1.101	.081
	조직	120	2.638	.734	-.905	.066	.694	.057	-.694	.042
	표현	120	.603	.100	-.883	.144	(어분 반응 분석으로 단계 모수 없음.) ¹³⁾			
	전달	120	.580	.104	-.968	.161	(이분 반응 분석으로 단계 모수 없음.)			
듣기·말하기 2-가	내용	240	.471	.060	.949	.134	(이분 반응 분석으로 단계 모수 없음.)			
	조직	240	.446	.036	-1.366	.107	1.826	.212	-1.826	.111
	표현	240	.150	.034	-7.913	1.303	2.935	1.284	-2.935	.365
	전달	240	.054	.043	-11.296	9.053	(이분 반응 분석으로 단계 모수 없음.)			

듣기·말하기 영역의 2개 문항 유형 채점 준거에 대하여 2모수 문항반응이론 분석을 실시한 결과, 곤란도 측면에서 대체로 매우 쉬운 수준으로 나타났으며, 변별도 측면에서는 대부분의 준거가 양호한 것으로 나타났다. 문항의 곤란도가 가장 낮았던 준거는 듣기·말하기 2번 유형의 (가)형에서 ‘전달’ 준거였으며, 2번 유형의 ‘내용’ 준거는 가장 높은 것으로 나타났다. 변별도 측면에서는 1번 문제의 ‘조직’ 준거가 가장 높은 것으로 나타났으며, 2번 유형에서의 ‘전달’ 준거의 변별도가 매우 낮게 추정되었다. 듣기·말하기 영역의 문항 양호도 분석 결과에서 전반적으로 낮은 곤란도와 양호한 변별도가 나타난 것과 관련하여 평가 영역 특성인 채점자들이 갖고 있는 채점 경향과 사용된 채점 척도 사용의 적절성의 영향이 있을 수 있

13) 듣기·말하기 1-나의 ‘표현’, ‘전달’, 듣기·말하기 2-가의 ‘전달’의 경우 2모수 문항반응이론 분석 과정에서 다분 반응 모형 분석 프로그램이 요구하는 분석 자료 설계에 부합하지 않아 3점 척도로 이루어진 채점 결과에서 채점 빈도가 적은 하위 두 구간을 묶어 양분된 척도로 수정하여 분석이 이루어졌으며, 이에 따라 분석 결과에서도 단계 모수 추정치가 산출되지 않았다.

다. 문항 유형 1의 가, 나 중에서는 (나)형 준거의 곤란도 모수가 (가)형에 비해 평균(0)에 가까운 분포를 보이고 있다는 점에서 (나)형을 향후 이루어질 본 검사 문항으로 최종 선정하였다.¹⁴⁾

③ 말하기 영역

<표 V-30> 2차 예비 조사 말하기 문항 양호도

문항	준거	사례 수	문항 모수 추정치				단계 모수							
			변별도	SE	곤란도	SE	1	SE	2	SE	3	SE	4	SE
말하기 1-가교	내용	240	3.538	.447	.050	.040	1.874	.054	.318	.024	-.692	.029	-1.500	.048
	조직	240	1.692	.105	-.136	.047	1.719	.090	1.122	.052	-.888	.044	-1.954	.083
	표현	240	.399	.024	-.366	.079	3.443	.348	1.130	.148	-1.562	.144	-3.011	.254
	전달	240	.880	.058	1.245	.058	1.735	.062	-.783	.120	-.061	.171	-.891	.216
말하기 2-가	내용	120	1.096	.095	.257	.062	1.181	.095	.133	.085	-.485	.103	-.829	.123
	조직	120	2.856	.402	.279	.053	.993	.048	.280	.040	-.275	.045	-.998	.061
	표현	120	1.401	.152	.085	.058	1.291	.089	.240	.069	-.432	.075	-1.099	.095
	전달	120	1.106	.120	.162	.062	1.298	.104	.365	.083	-.499	.092	-1.164	.122
말하기 2-나	내용	120	2.025	.226	.465	.054	1.173	.056	.324	.052	-.292	.061	-1.205	.091
	조직	120	3.831	.621	.350	.050	1.009	.038	.386	.033	-.308	.038	-1.086	.056
	표현	120	1.501	.155	.054	.058	1.219	.089	.500	.068	-.338	.066	-1.381	.093
	전달	120	1.629	.167	.101	.056	1.116	.077	.321	.065	-.163	.064	-1.274	.085
말하기 3-나	내용	240	2.773	.242	.211	.042	1.486	.042	.536	.030	-.614	.033	-1.408	.051
	조직	240	4.051	.471	.340	.041	1.272	.029	.356	.023	-.430	.026	-1.198	.039
	표현	240	1.560	.093	.162	.046	1.630	.066	.508	.044	-.650	.048	-1.488	.074
	전달	240	1.072	.068	.582	.048	1.841	.069	-.043	.063	-.314	.078	-1.485	.121

2차 예비 조사의 말하기 영역도 듣기·말하기 영역과 마찬가지로 3개 문항 유형 전체 4개 문제의 하위 준거에 대해 2모수 문항반응이론 분석을 통해 문항 양호도를 확인하였다. 그 결과, 변별도 모수에서 말하기 문항 3의 ‘조직’ 준거가 가장 높았으며, 가교 문항이었던 말하기 문항 1의 ‘표현’ 준거가 가장 낮은 것으로 나타났다. 곤란도 모수에서는 말하기 문항 1의 ‘표현’ 준거가 가장 곤란도가 낮고 ‘전달’ 준거의 곤란도가 가장 높아 점수를 받기 어려운 것으로 나타났다.

이상의 말하기 영역 분석 결과를 바탕으로 최종 문항 선정을 위해 말하기 문항 2의 (가)형과 (나)형을 살펴본 결과, (가)·(나)형 모두 양호한 수준의 변별도와 곤란도 수준을 나타내어 최종 1.2배수 문항에 포함시키기로 하였다.

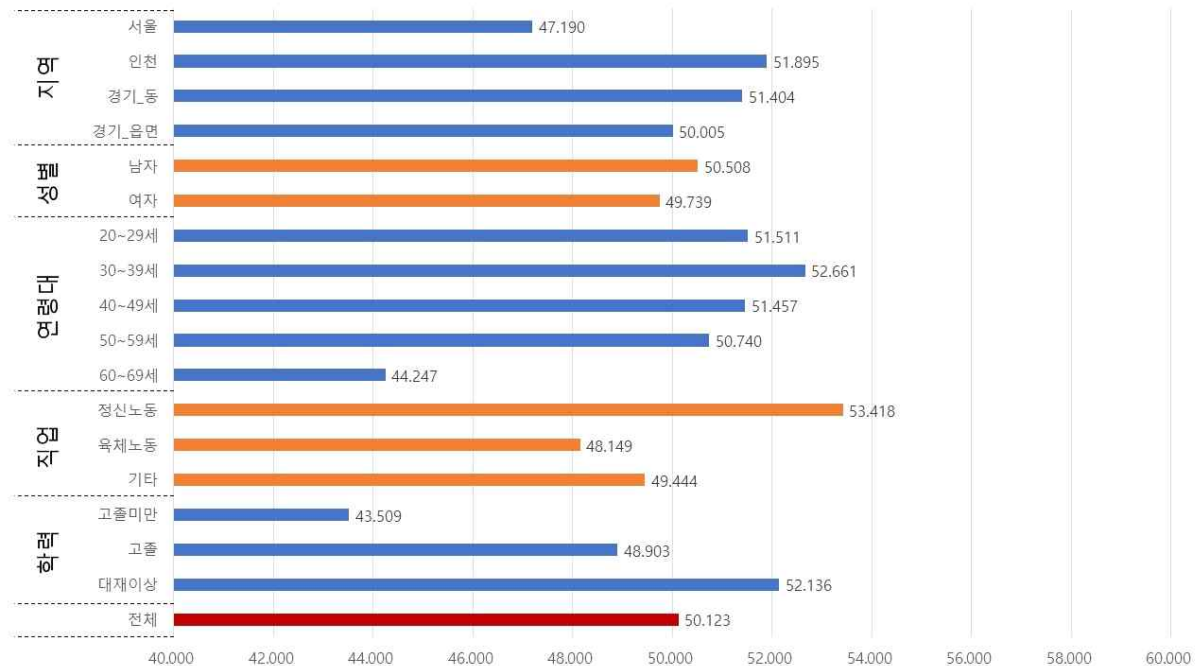
14) 2차 예비 조사의 듣기·말하기 문항 유형 간에 뚜렷한 우위가 나타난 것은 아니었기 때문에 차후 실시하는 본 조사에서 1·2차 예비 조사 결과를 종합적으로 검토하여 듣기·말하기 영역 본 조사 도구를 확정할 필요가 있다.

(4) 배경 변인별 분석

① 듣기 영역

<표 V-31> 2차 예비 조사 듣기 영역 배경 변인별 분석

구분		사례 수	평균	표준편차
지역	서울	60	47.190	9.062
	인천	60	51.895	7.900
	경기_동	60	51.404	8.706
	경기_읍면	60	50.005	7.584
성별	남자	120	50.508	8.433
	여자	120	49.739	8.549
연령대	20-29세	48	51.511	8.731
	30-39세	48	52.661	8.037
	40-49세	48	51.457	8.498
	50-59세	48	50.740	6.710
	60~69세	48	44.247	7.863
직업	정신노동	72	53.418	7.642
	육체노동	95	48.149	8.138
	기타	73	49.444	8.850
학력	고졸 미만	23	43.509	7.980
	고졸	88	48.903	7.667
	대재 이상	129	52.136	8.395
전체		240	50.123	8.483



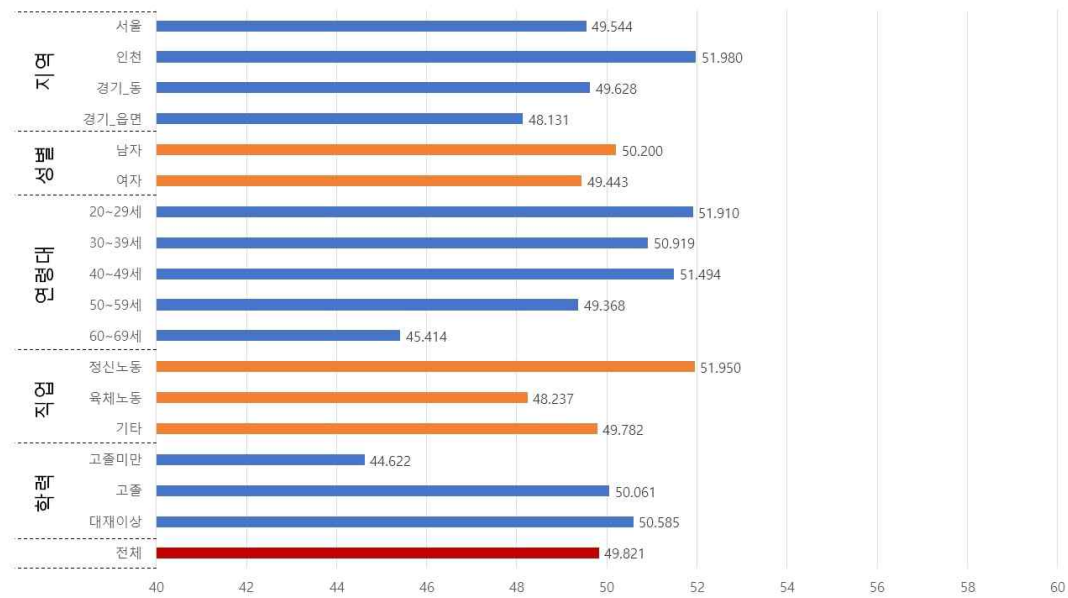
[그림 V-5] 2차 예비 조사 듣기 영역 배경 변인별 분석

듣기 영역 척도 점수의 배경 변인별 차이를 살펴보면, 지역에 따라서 ‘서울’이 47.190점으로 전체 평균에 비해 다소 낮은 수준이었고, ‘인천’과 ‘경기 동’은 각각 51.895점과 51.404점으로 평균보다 근소하게 높은 수준이었으며, ‘경기 읍면’은 50.005점으로 평균에 해당하는 수준이었다. 연령대에 따라서는 ‘20~29세’, ‘30~39세’, ‘40~49세’가 각각 51.511점, 52.661점, 51.457점으로 전체 평균에 비해 다소 높은 수준이었고, ‘50~59세’가 50.740점으로 평균과 비슷한 수준이었으며, ‘60~69세’가 44.247점으로 평균에 비해 낮은 수준이었다. 직업에 따라서는 ‘정신노동’이 53.418점으로 ‘육체노동’의 48.149점에 비해 높았다. 학력에 따라서는 ‘대재 이상’이 52.136점으로 가장 높고, ‘고졸’이 48.903점으로 뒤를 이었으며, ‘고졸 미만’이 43.509점으로 가장 낮은 수준이었다.

② 듣기·말하기 영역

<표 V-32> 2차 예비 조사 듣기·말하기 영역 배경 변인별 분석

구분	사례 수	평균	표준편차
지역	서울	60	49.544
	인천	60	51.980
	경기_동	60	49.628
	경기_읍면	60	48.131
성별	남자	120	50.200
	여자	120	49.443
연령대	20-29세	48	51.910
	30-39세	48	50.919
	40-49세	48	51.494
	50-59세	48	49.368
	60-69세	48	45.414
직업	정신노동	72	51.950
	육체노동	95	48.237
	기타	73	49.782
학력	고졸 미만	23	44.622
	고졸	88	50.061
	대재 이상	129	50.585
전체	240	49.821	8/503



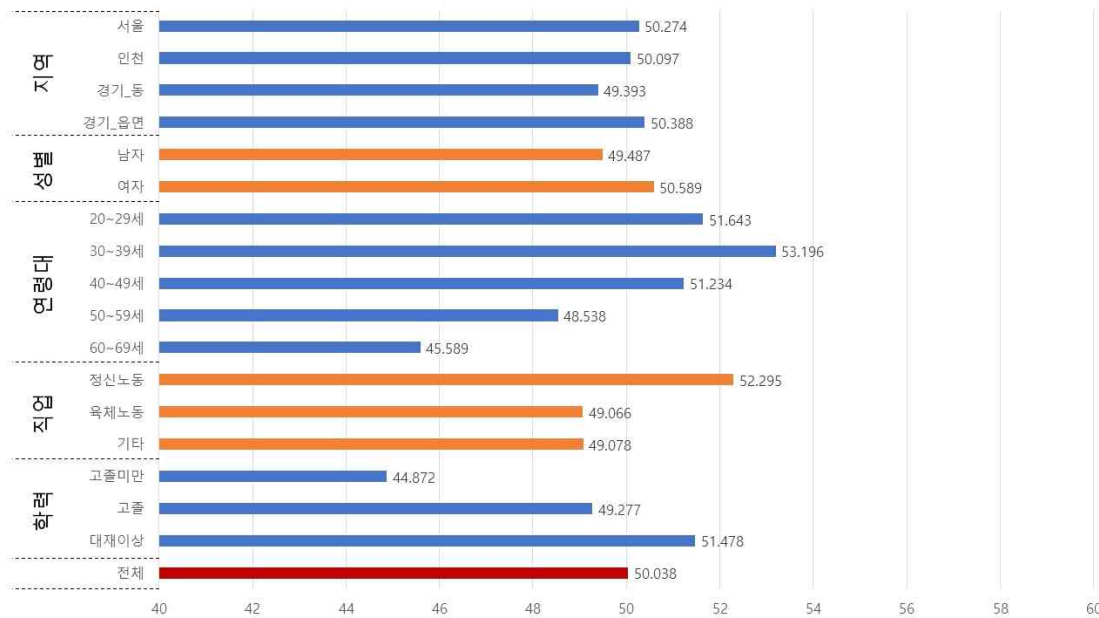
[그림 V-6] 2차 예비 조사 듣기·말하기 영역 배경 변인별 분석

듣기·말하기 영역 척도 점수의 배경 변인별 차이를 살펴보면, 지역에 따라서 ‘인천’이 51.980점으로 전체 평균에 비해 다소 높은 수준이었고, ‘서울’, ‘경기 동’, ‘경기 읍면’은 각각 49.544점, 49.628점, 48.131점으로 평균보다 약간 낮은 수준이었다. 성별에 따라서는 ‘남자’가 50.200점으로 ‘여자’의 49.443점에 비해 높았다. 연령대에 따라서는 ‘20~29세’, ‘30~39세’, ‘40~49세’가 각각 51.910점, 50.919점, 51.494점으로 전체 평균에 비해 약간 높은 수준이었고, ‘50~59세’와 ‘60~69세’는 각각 49.368점과 45.414점으로 평균보다 낮은 수준이었다. 직업에 따라서는 ‘정신노동’이 51.950점으로 ‘육체노동’의 48.237점에 비해 높았다. 학력에 따라서는 ‘대재 이상’이 50.585점으로 가장 높고, ‘고졸’이 50.061점으로 뒤를 이었으며, ‘고졸 미만’이 44.622점으로 가장 낮은 수준이었다.

③ 말하기 영역

<표 V-33> 2차 예비 조사 말하기 영역 배경 변인별 분석

구분		사례 수	평균	표준편차
지역	서울	60	50.274	10.033
	인천	60	50.097	9.548
	경기_동	60	49.393	7.357
	경기_읍면	60	50.388	8.972
성별	남자	120	49.487	8.431
	여자	120	50.589	9.512
연령대	20~29세	48	51.643	10.075
	30~39세	48	53.196	6.744
	40~49세	48	51.234	8.450
	50~59세	48	48.538	9.864
	60~69세	48	45.589	7.623
직업	정신노동	72	52.295	9.604
	육체노동	95	49.066	8.140
	기타	73	49.078	9.125
학력	고졸 미만	23	44.872	7.082
	고졸	88	49.277	8.886
	대재 이상	129	51.478	9.011
전체		240	50.038	8.986



[그림 V-7] 2차 예비 조사 말하기 영역 배경 변인별 분석

말하기 영역 척도 점수의 배경 변인별 차이를 살펴보면, 지역에 따라서 ‘서울’, ‘인천’, ‘경기 읍면’이 50.274점, 50.097점, 50.388점으로 전체 평균에 비해 약간 높은 수준이었고, ‘경기 동’은 49.393점으로 평균에 비해 낮은 수준이었다. 성별에 따라서는 ‘여자’가 50.589점으로 ‘남자’의 49.487점에 비해 높았

다. 연령대에 따라서는 ‘20~29세’, ‘30~39세’, ‘40~49세’가 각각 51.643점, 53.196점, 51.234점으로 전체 평균에 비해 다소 높은 수준이었고, ‘50~59세’와 ‘60~69세’가 각각 48.538점과 45.589점으로 평균에 비해 낮은 수준이었다. 직업에 따라서는 ‘정신노동’이 52.295점으로 ‘육체노동’의 49.066점에 비해 높았다. 학력에 따라서는 ‘대재 이상’이 51.487점으로 가장 높고, ‘고졸’이 49.277점으로 뒤를 이었으며, ‘고졸 미만’이 44.872점으로 가장 낮은 수준이었다.

3) 최종 도구의 확정

최종 도구의 확정 과정은 다음과 같다. 우선 공동 연구원들이 모여 2차 예비 조사의 분석 결과를 보고 받고 2023년 시행 예정인 본 조사를 위한 조사 도구 선정 및 수정의 방향을 논의하였다. 이후 각 영역을 담당하는 공동 연구원들이 수정한 영역별 문항을, 모든 공동 연구원들이 협의를 통해 검토하여 최종적으로 1.2배수 문항을 선정하였다. 최종적으로 선정된 문항의 전체적인 문항 구성 및 개별 문항에 관해 각 영역 전문가의 자문 및 검토를 의뢰하고 그 결과를 반영하여 최종 도구를 확정하였다. 자문 및 검토를 의뢰한 전문가는 <표 V-34>와 같다.

<표 V-34> 최종 조사 도구 자문 및 검토 전문가

전문가	소속
이선영 교수	경인교육대학교 국어교육과
최영인 교수	서울교육대학교 국어교육과



제 6 장

제 언



1. 듣기·말하기 영역 본 조사 방안 수립

1) 조사 도구 구성

2023년부터 영역별, 연차별 실태 조사로 전환됨에 따라 실태 조사 결과에 대한 심층적인 분석이 가능해지므로 본 연구는 환류 정책을 구체적으로 수립하기 위한 본 조사 시행 및 결과 활용 방안을 제안한다. 이를 위해 본 연구에서 진행될 1, 2차 예비 조사 결과 및 앞서 수립한 국민의 국어능력 영역별, 연차별 실태 조사 고도화 방안 계획을 고려하여 본 조사의 일정, 조사 방법 등의 시행 방안을 구체적으로 설계한다. 이에 더하여 듣기·말하기 영역의 예비 조사 결과를 분석하여 관련 국어 정책을 수립하기 위한 방향에 대해 입론한다. 특히 영역별 국어능력 본 조사 결과의 정책적·교육적 활용 방안을 탐색하여 상세화하고자 한다. 예컨대 2023년에 시행될 듣기·말하기 영역의 본 조사를 중심으로 추후 단계별로 진행될 읽기, 쓰기, 디지털 리터러시 영역에 따라 국어능력 제고가 가장 크게 요구되는 영역은 무엇이며, 성별, 세대, 지역, 직업, 학력 등의 주요 변인에 따라 국어교육이 가장 크게 요구되는 집단은 어떤 집단인지 등을 도출하여 이를 토대로 관련 국어 정책 수립의 방향성을 제안할 것이다.

이를 위해 먼저 2년 차인 2023년도 듣기·말하기 영역 본 조사의 조사지 구성 방식을 보이면 다음과 같다.

<표 VI-1> 2년 차(2023년) 듣기·말하기 본 조사의 조사지 구성 방식

조사 영역	조사 방식	문항 수	시간
듣기	선택형(객관식)	15	25
듣기·말하기	수행형(주관식)	2	5
말하기	수행형(주관식)	3	15
계		20	45분

위의 표와 같이 기본적으로 단일 조사 도구의 크기는 본 연구의 예비 조사와 유사하게 구성할 수 있다. 다만, 단일 조사 도구를 투입할 것인지, 공통 문항 비동등 집단 설계 방식의 이원화된 조사 도구를 투입할 것인지에 대해서는 논의가 필요하다. 후자의 경우 2013 국민의 국어능력 실태 조사에서 취한 방식이자 일반적으로 국가 수준의 대단위 성취도 평가에서 활용된다. 무엇보다 이 방식은 하위 평가 요소를 보다 다양하게 상정할 수 있다는 장점이 있으므로 이 같은 지점들을 종합적으로 고려하여 조사 도구를 설계할 필요가 있다. 아울러 3년 차(읽기, 쓰기), 4년 차(디지털 리터러시)의 조사 시행도 동일한 방식으로 설계하는 것까지를 고려하여 조사 방식을 결정해야 할 것이다.

한편 단일 조사 도구, 이원화된 조사 도구 등과 같은 조사 방식에 앞서 적정한 조사 시간을 따져 볼 필요가 있다. 본 연구에서는 위의 표와 같이 문제 수와 유형을 고려하여 45분의 시간을 상정하였지만, 피험자에게 45분이라는 조사 시간은 부담으로 작용할 가능성이 있다. 특히 60대 이상의 피험자를 고려하면 최대 30분을 넘기지 않는 선에서 듣기·말하기 영역의 본 조사에 적절한 조사 시간을 확정할 필요가 있을 것이다. 이 과정에서 국민의 국어능력 평가의 목적과 대상이 개인별 국어능력의 측정 및 피드백 제공이 아니라 전체 국민의 국어능력 실태라는 점을 주지하면 굳이 한 개인이 모든 영역의 문제를

풀어야만 하는 당위성은 낮다. 따라서 시간적인 부담을 줄이기 위해서는 집단별로 영역을 보다 세부적으로 나눠서 평가한 후 그 결과를 종합하여 국민 전체의 추이를 분석하는 방안을 고려해 볼 필요가 있다. 물론 추후 인공지능 활용 국민의 국어능력 진단 체계 개발 사업의 데이터 활용 여부에 따라 평가 방식이 달라질 수 있을 것이다. 그러나 해당 데이터를 활용한다고 하더라도 집단별로 영역을 나눠서 조사하는 경우 인공지능 데이터 역시 집단으로 가져와 처리가 가능하다는 점에서 문제가 될 지점은 없어 보인다.

2) 조사 인원 및 방식

(1) 모집단 및 표본 수

2023년 듣기·말하기 영역 본 조사의 모집단은 이전의 국민의 국어능력 본 조사와 마찬가지로 전국 17개 시도에 거주하는 만 20세~69세의 남녀로 정의한다. 본 조사의 목표 표본 수는 하나의 영역만을 조사한다는 점을 고려하여 2018 국민의 국어능력 본 조사 인원인 5,500명보다 적은 3,000명으로 제안하고자 한다.

(2) 층화 후 집락추출

본 조사의 표본 추출은 예비 조사 및 기존의 1, 2차 조사와 동일하게 층화 후 집락추출법을 적용한다. 전체 모집단을 17개 시도로 나눈 뒤, 하나의 시도를 다시 동부, 읍면부를 나누어 층을 구분하며, 서울, 대전, 광주는 읍면 지역이 없기 때문에 전국을 총 31개의 층으로 나눈 후, 각 층의 주민 등록 인구 통계상 만20~69세 인구수에 비례하도록 층별 표본 수를 배분한다. 각 층별 표본 수를 결정한 후에는 층 내의 읍·면·동을 집락으로 간주하고, 2단계 집락추출을 진행한다. 집락추출 과정에서는 크기비례확률추출(Sampling with probabilities proportional to size)을 사용한다.

(3) 조사 대상자 추출

위와 같이 추출된 집락(읍·면·동)에서 조사 대상자(개인)를 선정하는 과정에서는 그간의 실태 조사에서와 동일하게 할당추출법을 사용한다. 할당추출은 모집단을 구성하는 하부 집단의 구성비에 맞게 표본을 추출하는 방식이며, 사전에 할당한 조건에 맞는 응답자를 찾아서 표집하는 방법이다. 비확률추출법이기는 하지만 조사를 신속하고 효율적으로 수행할 수 있다는 장점이 있으며, 확률추출을 적용하기 어려운 여건에서 모집단의 주요한 특성과 유사한 표본을 추출하기 위한 대안이 된다. 이에 예비 조사에서와 마찬가지로 개인의 국어능력과 관련성이 높다고 볼 수 있는 성별, 연령대, 직업군, 학력을 기준으로 한 할당추출을 진행한다. 이미 모집단을 층화하는 단계에서 지역 변인을 고려하기 때문에 2023 본 조사 표본 추출을 위한 변인은 ‘지역, 성별, 연령대, 직업, 학력’의 5개이다.

(4) 조사 방법

2023년 본 조사에서는 2차 실태 조사(2018년)와 마찬가지로 기본적으로 TAPI를 이용한 1:1 방문 면접 조사로 진행한다. 이때 태블릿 PC에 어려움을 겪는 표본 집단을 대상으로는 지면 검사지로 보완하는 병행 방식으로 조사를 수행한다. 그런데 이러한 조사 방식은 앞서 제기한바, 기본적으로 가구 방문을 통한 1:1 조사라는 점에서 조사 환경 및 피조사자들의 적극적인 참여도 측면에서 한계가 있다. 따라서 일부 피험자 집단을 대상으로 '내방형 집단 조사'를 실시하여 조사 결과를 비교해 볼 필요가 있다. 그러나 이에 대한 적절한 유인책을 마련하기 위해서는 관련 예산의 증대 편성, 지역 국어문화원의 협조 등이 필수적으로 요구된다. 더하여, 내방형 집단 조사와 기존의 1:1 방문 면접 조사의 결과가 상이한 경우에 대한 문제, 완전한 내방형 집단 조사로 전환하는 데에 따르는 지역, 교통, 연령 문제 등을 총체적으로 고려할 필요가 있다.

이상의 고려 사항 중 '예산의 증대 편성'과 관련하여 다음은 3,000명을 기준으로 '1:1 방문 조사', '1:1 방문 조사(2,700명) + 내방형 집단 조사(300명)' 시 각각 투입되어야 할 조사원과 소요 예산을 조사 업체(KANTAR PUBLIC)를 통해 예측한 것이며, 동일한 방식으로 5,000명의 표본을 대상으로 했을 때의 관련 정보를 정리한 것이다. 이때 내방형 집단 조사의 인원으로는 표본 수의 10%를 상정한 것이며, 해당 인원이 적정한지에 대해서는 예산 및 변인 등을 고려하여 결정해야 할 것이다.

<표 VI-2> 2년 차(2023년) 듣기·말하기 본 조사 방식에 따른 예산안

구분	본 조사 피험자 수	조사원 수	예산 (vat 포함)	비고
2013년 국민의 국어능력 평가 조사	3,000명	118명	100,000,000원	-예비 조사 400명, 본 조사 3,000명 -가/나 set별 1,500명 조사
2018년 국민의 국어능력 평가 조사	5,500명	111명	363,225,600원	-예비 조사 300명, 본 조사 5,500명 -가/나 set별 2,500명, 통합 500명
2023년 듣기·말하기 본 조사 (방문 면접 조사)	3,000명	120명	230,000,000원	
2023년 듣기·말하기 본 조사 (방문 면접 + 내방형 집단 조사)	3,000명	110명	250,000,000원	-방문 면접 2,700명, 내방형 300명 -내방형 진행자/안내자 별도 구성
2023년 듣기·말하기 본 조사 (방문 면접 조사)	5,000명	200명	370,000,000원	
2023년 듣기·말하기 본 조사 (방문 면접 + 내방형 집단 조사)	5,000명	180명	400,000,000원	-방문 면접 4,500명, 내방형 500명 -내방형 진행자/안내자 별도 구성

위의 표에서 제시한 예산에는 피험자 사례비가 최소한으로 포함된 것이다. 방문 면접 조사의 경우

2018년도 2주기 조사에서와 동일한 금액으로 책정되었으나, 내방형 집단 조사의 경우 피험자가 특정 장소에 방문하여 조사에 응해야 한다는 점에서 최소 1시간 ~ 1시간 30분 정도가 소요될 것으로 예상되는 바, 피험자에게 제공해야 하는 사례비 역시 방문 면접 조사에 비해 5배 이상의 금액으로 책정되었다. 그러나 본 연구에서 제안한 바와 같이 앞으로 매 본 조사마다 조사 영역 특정의 배경 설문에 대한 응답이 더해질 시 조사 시간은 상술한 시간보다 더 늘어날 수 있으며, 일반적으로 2시간 정도 소요되는 FGI의 피험자 사례비가 8~10만 원 정도라는 점을 고려하면 위에서 제시한 예산은 최소한의 금액으로 보아야 한다. 주지하듯이 본 조사 방식과 인원 등에 대한 결정 역시 이 같은 예산을 고려하여 이루어져야 할 것이다.

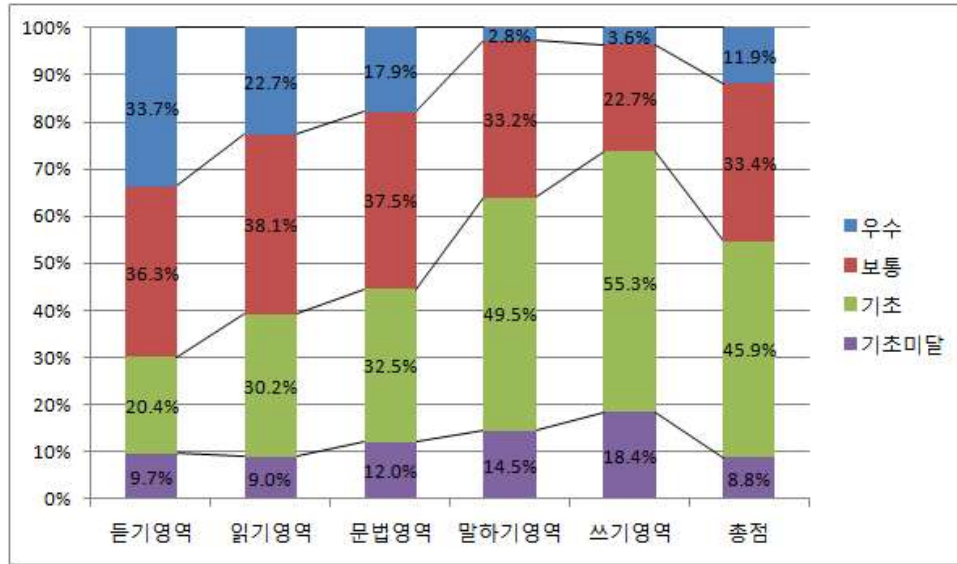
3) 성취 수준의 등급화

2023년 본 조사의 국어능력 등급화를 위해서는 2013년 1차 조사에서 적용한 앵고프(Angoff) 방식을 따른다. 앵고프 방식은 전문가 패널 집단이 영역별로 각 문항에 대한 기대 정답률을 판별하여 분할 점수를 산출하는 것인데, 이는 절대적인 능력 수준을 구분하는 기준을 산출하는 방식이다. 이는 원년 연구인 2013년 조사, 이전 주기 조사인 2018년 조사와의 연계성을 유지하며, 다음 주기 조사 결과와의 비교에 있어 안정성을 추구하는 방식이기도 하다.

국민의 국어능력 실태 조사를 바탕으로 국어 정책을 수립·시행하기 위해서는 상대적인 수행 정도가 아니라 준거 참조 평가(criterion-referenced test)를 통해 그 도달 수준을 직접적으로 해석할 수 있어야 한다. 이를 위해 국민의 국어능력을 둘 이상의 범주와 집단으로 분류하는 준거에 해당하는 평가 척도상의 점수인 분할 점수(cut-off score)를 설정하여 절대적인 능력 정도를 파악할 필요가 있다. 분할 점수는 추상적인 준거를 가시적인 수치로 전환하는 과정으로, 각 수준의 최소 능력 보유자가 지닌 능력치를 기준으로 하기 때문에 국어 정책을 수립하는 데 있어 질적인 정보를 제공한다는 점에서도 유의미하다.

검사 동등화를 실시하는 검사의 경우는 기본 검사에 기준 설정을 하고 이후에 개발되는 검사의 분할 점수는 기본 검사의 분할 점수와 동등한 점수를 찾는 방식으로 이루어진다(김선, 김성훈, 반재천, 2009: 1037). 즉, 새로운 검사의 분할 점수는 검사들 간의 곤란도를 조정하는 검사 동등화 과정을 통하여 찾을 수 있다(Kolen & Brennan, 2004; 김선, 김성훈, 반재천, 2009: 1037에서 재인용). 검사 동등화의 방법에는 단일 집단 설계, 동등 집단 설계, 가교 검사 설계가 있는데(성태제, 1994: 212), 2018년 조사에서는 2013년 원년 연구에서 설정한 가교 문항을 사용하여 두 검사를 동등화함으로써 분할 점수를 산출하였다. 마찬가지로 2023년의 본 조사에서도 가교 문항을 통하여 2013년 조사, 2018년 조사와의 동등화를 시행하고, 이에 따라 분할 점수를 설정할 수 있게 된다.

구체적으로, 원년 조사인 2013년 연구에서는 앞으로 시행될 조사들과의 연계를 위하여 앵고프(Angoff) 방식으로 전문가 패널의 합의를 통하여 각 영역별 분할 점수를 산출하고, ‘우수’, ‘보통’, ‘기초’, ‘기초 미달’로 국민의 국어능력 수준을 등급화하였다. 이처럼 각 영역별로 해당 등급을 구획하기 위한 분할 점수를 산출한 것은 다음 주기 조사 결과 사이의 종단적 비교가 가능하도록 하기 위함이었다. 그리고 전체 국어능력의 분할 점수는 각 영역별 분할 점수를 합산한 총점으로 설정하고, 이를 바탕으로 전체 국어능력을 등급화하였다. 각 영역별 등급화와 전체 국어능력의 등급화 결과를 보이면 다음 [그림 VI-1]과 같다.



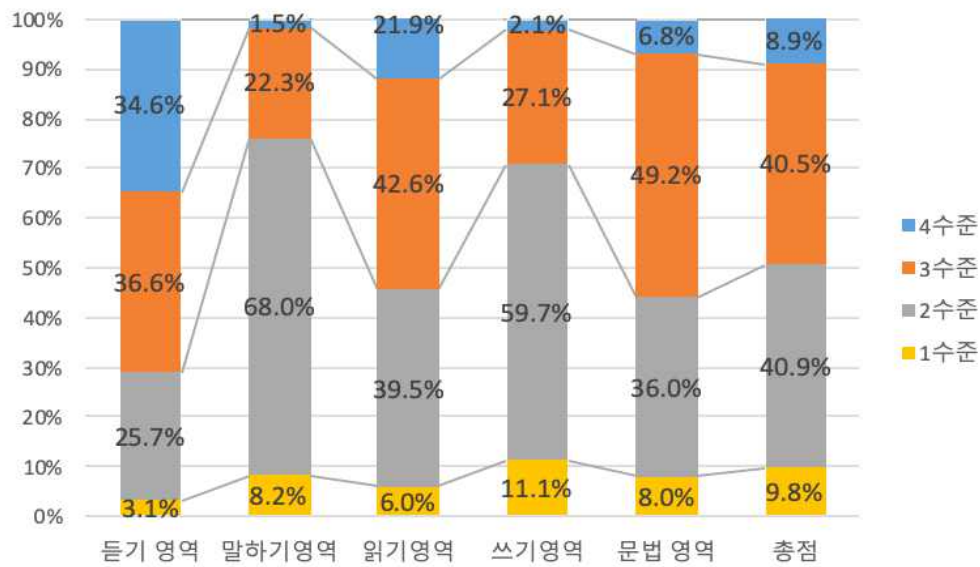
[그림 VI-1] 2013년 조사에서의 영역별 및 전체 국어능력의 등급별 분포

이어서 2018년 조사에서는 원년 조사에서 도출한 각 영역별 분할 점수를 다음 주기 조사에서도 비교의 기준으로 유지하기 위하여, 2013년에 도출된 분할 점수를 그대로 준용하되, 2018년과 2013년의 조사 도구를 동등화하고 이 과정에서 2018년 조사의 분할 점수를 새로이 도출하였다. 그리고 원년 조사의 4등급 체제를 유지하되, ‘우수’, ‘보통’, ‘기초’, ‘기초 미달’의 명칭을 ‘4수준’, ‘3수준’, ‘2수준’, ‘1수준’으로 바꾸어 제시하였다.

한편, 원년 조사와 달리 2018년 조사에서는 전체 국어능력의 성취 수준을 등급화할 때 총점이 아닌 영역별 수준의 합을 기준으로 하였다. 이는 2013년 조사의 방식이 점수 체계가 동일할 경우에는 의미를 지니지만, 영역별 점수 체계가 다를 때에는 총점 합산 시 왜곡이 나타날 수 있다는 점에서였다. 이를 위하여 2013년 조사의 전체 국어능력의 분할 점수를 2018년 조사의 영역별 수준의 합으로 환산하고 해당 값을 전체 국어능력의 성취 수준 분할 기준으로 삼았다. 2018년 조사에서의 전체 국어능력의 성취 수준 분할 기준은 다음 <표 VI-3>과 같으며, 영역별 및 전체 국어능력의 등급화 결과는 [그림 VI-2]와 같다.

<표 VI-3> 2018년 조사에서의 전체 국어능력의 성취 수준 등급화

영역	수준	영역별 등급의 합 (평균 수준)	세부 내용
전체 국어능력	4수준	17(3.4)	영역별 4수준이 최소 2개 이상
	3수준	13(2.6)	영역별 3수준이 최소 3개 이상
	2수준	9(1.8)	영역별 2수준이 최소 4개 이상
	1수준		



[그림 VI-2] 2018년 조사에서의 영역별 및 전체 국어능력의 등급별 분포

앞으로 진행될 2023년 본 조사에서도 마찬가지로 가교 문항을 바탕으로 이전 검사와의 동등화 및 분할 점수 산출이 이루어질 수 있으며, 이렇게 도출된 분할 점수를 바탕으로 조사 결과를 등급화할 수 있을 것이다. 그리고 이러한 일련의 절차를 수행하기 위해서는 원년 조사에서 설정한 분할 점수 및 가교 문항을 준용하여야 한다. 또한, 연차별로 조사가 수행됨에 따라, 연차별 조사가 실시될 때마다 영역별로 등급화를 실시할 수 있을 것이며, 모든 영역의 조사가 실시된 후에는 전체 국어능력의 등급화가 가능할 것이다. 이때, 전체 국어능력 점수를 등급화할 경우에는 만점 및 문항당 배점이 영역별로 다르기 때문에 영역별로 합산한 등급이 같음에도 전체 국어능력 총점이 달라 전체 국어능력에서는 서로 다른 성취 수준으로 구분되는 문제가 발생할 수 있다. 따라서 2013년 조사의 총점 기준보다는 2018년 조사에서 도출하였던 전체 국어능력 점수 분할 기준을 적용하는 것이 타당할 것으로 보인다.

2. 5년 주기 정례화 방안 수립

1) 5년 주기 국어능력 조사의 정례화 방향 및 개선안

그간 이루어진 1, 2차 국민의 국어능력 실태 조사는 5년을 주기로 하여 국민의 국어능력을 진단하기 위해 전 영역을 대상으로 이루어져 왔다. 2차 조사인 2018년 이후 이번 연구는 그간의 조사와는 달리 앞으로 국민의 국어능력을 영역별, 연차별로 조사하기 위한 3차 조사의 첫 단계에 놓여 있다. 주지하듯 2018년도 조사까지 국민의 국어능력 연구는 듣기·말하기, 읽기, 쓰기, 문법의 전 영역에 해당하는 조사 도구를 개발하여 대규모 표본을 대상으로 조사를 수행하였으며, ‘조사 도구 개발 연구’와 ‘본 조사 연구’의 양자가 2년에 걸쳐 연속적으로 이루어져 왔다.

그러나 2018 국민의 국어능력 조사 연구에서도 지적된바, 이론적이면서도 실용적인 가치를 지니는 국민의 국어능력 조사 연구의 잠재력을 더욱더 발현하기 위해서는 1~2년 단위의 개별적인 연구로 그치는 문제점을 해결하고, 보다 장기적이고 체계적인 단계별 연구가 수행되어야 한다. 다시 말해 이전까지의 연구가 ‘도구 개발 연구’와 ‘조사 연구’의 두 단계로 구성되었다면, 향후 연구는 ‘국민의 국어능력 연구’라는 큰 틀 아래에서 ‘도구 개발 연구’ - ‘조사 연구’ - ‘심층 분석 및 후속 연구’의 단계로 진행될 필요가 있는 것이다.

무엇보다 기존의 2013, 2018 실태 조사와 같은 방식에서는 조사 결과에 따른 환류 정책을 시행하는데 한계가 존재하였으며, 전 영역을 통합한 대면 조사로 진행하여 조사 시간뿐만 아니라 조사의 정확성 측면의 문제가 자리하였다. 더하여, 전술하였듯이 변화하는 국어 환경을 그때그때 적절히 반영하기 어렵다는 문제 또한 제기되었다. 따라서 연차별·영역별 분할 실태 조사를 시행하게 되면 5년 동안 특정 영역의 실태 조사 결과에 대한 심층적인 분석은 물론 조사 결과에 대한 구체적인 환류 정책을 수립하는데 효과적이라는 점에서 본 연구에서는 앞서 영역별 연차 조사안을 제안한 바 있다. 이를 정리하여 제시하면 다음과 같다.

<표 VI-4> 국민의 국어능력 실태 조사 고도화

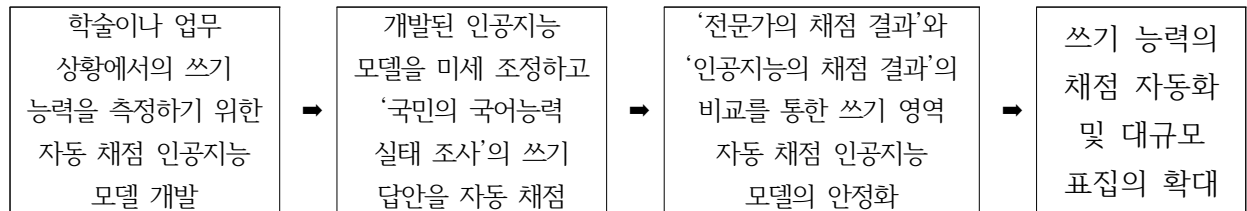
구분	방식	주기	대상
1~2차 조사	대면 조사	5년 주기 4개 영역 통합 조사 (1년 차) 도구 개발('17년) (2년 차) 통합 조사('18년) (3~5년 차) 3년 휴지기('19~'21년)	성인 (20~70세)
1~2차 조사의 문제점 및 한계		1~2차 조사는 4개 영역 통합 대면 조사로 진행하여 조사 시간이 길고 조사 대상의 피로도가 증가하여 정확도가 떨어짐. - 변화하는 국어 환경을 적시에 파악하고 반영하기 어려움. - 거대 규모의 통합 조사로 예비 조사·본 조사 이후 연내 심층 분석이 불가함.	
3차 조사 개선안	온라인 조사 (대면 조사)	- 매년 - 영역별 연차 조사(안)('22년~) - 영역별 분할 조사로 심층 분석 실시 및 조사의 정책 환류도 향상 가능	성인 (20~70세)

보완)	1안				
		듣기말하기영역	읽기 쓰기영역	규범 다룰 랴랴시영역	정책 제언
	1년 차 (’22년)	예비조사→ 도구 개발			연별영역 분할 조사계획 수립
	2년 차 (’23년)	본 조사→ 결과분석 *특수배경 설문	예비조사→ 도구 개발		듣기말하기영역 정책 제언
	3년 차 (’24년)		본 조사→ 결과분석 *특수배경 설문	예비조사→ 도구 개발	읽기 쓰기영역 정책 제언
	4년 차 (’25년)			본 조사→ 결과분석 *특수배경 설문	규범 다룰 랴랴시영역 정책 제언
	5년 차 (’26년)				조사결과종합 및심층분석
	2안				
		듣기말하기영역	읽기 쓰기영역	규범 다룰 랴랴시영역	정책 제언
	1년 차 (’22년)	예비조사→ 도구 개발			연별영역 분할 조사계획 수립
	2년 차 (’23년)	본 조사→ 결과분석 *특수배경 설문			듣기말하기영역 정책 제언
	3년 차 (’24년)		예비조사→ 도구 개발 본 조사→ 결과분석 *특수배경 설문		읽기 쓰기영역 정책 제언
	4년 차 (’25년)			예비 조사→ 도구 개발 본 조사→ 결과분석 *특수배경 설문	규범 다룰 랴랴시영역 정책 제언
	5년 차 (’26년)				조사결과종합 및심층분석

2) 인공지능 활용 국민의 국어능력 진단체계 개발 사업과의 연계안

전술한바, 본 연구에서는 국민의 국어능력 사업과 ‘인공지능 활용 국어능력 진단 체계 개발’ 과업과의 중장기적인 연계 방향을 다음과 같이 구상한 결과, 특히 ‘쓰기’ 영역에서의 활용 방향을 다음과 같이 제안하였다.

<표 VI-5> 인공지능 사업과 국어능력 실태 조사의 연계안



본 연구는 20~60대의 전 국민을 대상으로 한 국어능력 조사라는 점에서 ‘인공지능 활용 국어능력 진단 체계 개발’ 과업에서 목적으로 하는 학술, 업무 상황에서의 쓰기 능력 진단에 비하여 쓰기 문항의 수준이 상대적으로 낮고 요구하는 답안의 길이 자체 역시 짧다. 따라서 ‘인공지능 활용 국어능력 진단 체계 개발’ 과업에서 궁극적으로 개발하고자 하는 자동 채점 인공지능 모델은 추후 본 연구의 ‘쓰기’ 영역 채점 시 원활한 적용이 가능할 것이다. 물론 쓰기 과제의 수준과 유형을 고려하여 이를 채점에 적용할 때 해당 모델을 그대로 사용하기보다는 본 연구의 목적에 맞게 조정 과정을 거쳐야 함은 물론이다.

구체적으로는 본 연구의 쓰기 영역 답안에 대한 전문가 채점 데이터가 확보되어야 한다. 그리고 쓰기 영역 답안 데이터의 일부와 해당 답안의 전문가 채점 결과를 함께 모델에 학습시킨 후, 학습시키지 않고 남아 있는 일부 데이터를 대상으로 학습된 모델이 자동 채점을 수행하도록 해야 할 것이다. 이후, 모델이 자동 채점한 결과를 전문가의 채점 결과와 비교하여 자동 채점을 안정화하는 과정이 후행되어야 한다. 만약 모델이 자동 채점한 결과를 전문가의 채점 결과와 비교하여 정확도가 낮게 나올 경우, 이는 해당 모델을 개발할 때 설정한 여러 가지 요소들이 학술, 직무 작문의 채점에는 용이하게 작용하였으나 본 연구의 쓰기 영역 채점과는 관련성이 부족하였다는 의미로 해석되므로 이 같은 요소들의 기준을 조정하여 본 연구의 쓰기 영역에 보다 적합하도록 자동 채점 인공지능 모델을 안정화할 수 있을 것으로 보인다.

이상으로 ‘인공지능 활용 국어능력 진단 체계 개발’ 과업과의 연계안은 앞으로 국민의 국어능력 조사라는 대규모 조사를 채점하는 것에 따른 비용과 인력 문제를 해결할 수 있게 됨은 물론 더 나아가 이를 바탕으로 더 많은 수의 국민을 대상으로 한 조사도 가능하게 만들 것으로 보인다.

3) 후속 연구를 통한 실효성 있는 정책 수립

국민의 국어능력 조사 결과를 기반으로 한 후속 연구는 무엇보다 실효성 있는 국어 정책을 수립하고 국어능력 함양을 위한 프로그램과 같은 활용 방안을 구체적으로 개발하기 위하여 필요한 과정이다. 다시 말해 실태를 단순히 조사하는 데에서 그치는 것이 아니라 조사를 통해 도출된 결과를 정확하고 깊이 있게 해석하기 위한 연구가 요구되는 것이다. 예컨대 조사 영역별로 국민의 국어능력에 영향을 미친 변

인은 무엇이며, 영역별 특화된 배경 설문과의 영향 관계는 어떠한지, 국어능력의 격차를 해소하기 위하여 필요한 교육적 처치는 무엇인지 등의 연구가 수행될 수 있다.

이 같은 후속 연구는 통상적으로 대규모 조사 연구에서 조사 결과를 해석하기 위해 수행되어 왔다. 예컨대, 국제학업성취도평가(PISA), 국제성인역량조사(PIAAC)에서는 조사 연구와 더불어 조사 결과를 분석하고 해석하는 연구 보고서를 발간하고 있다. 이뿐만 아니라 결과에 영향을 미치는 여러 변인을 수집하여 교육 프로그램 마련에도 힘쓰고 있다. 2018 국민의 국어능력 조사 연구에서도 인구통계학적 변인에 더하여 일부 설문 배경 변인을 추가적으로 설정하여 조사한 바 있다. 그러나 이 역시 영역별 특화된 배경 설문은 아니라는 점에서 본 연구에서는 앞으로 영역별 본 조사 시 특화된 배경 설문을 마련하여 실시할 필요가 있음을 제안하고자 한다.

이처럼 세분화된 영역별 심층 조사 및 연구는 영역별 하위 요소를 세분화하고, 깊이 있는 분석을 가능하게 하여 세밀화된 환류 정책의 수립을 가능하게 한다. 무엇보다 이는 궁극적으로 국어능력의 전반적인 신장을 효과적으로 도모할 것이다. 이에 본 연구에서는 후속 연구 방향을 현재 학업성취도 평가나 대학수학능력시험의 후속 연구 방식을 참조하여 ‘양적 데이터 및 실제 응답 자료(쓰기, 말하기) 공유 → 연구 주제 공모 → 연구 결과에 대한 연구 보고 대회 개최 → 별도의 홈페이지 결과 공유 + 국어원 혹은 관련 기관 발간 소식지 등과 연계’ 및 ‘국어 정책 통계 포털, 홍보 방안과 연계’하는 방식을 제안하였다.

이처럼 영역별 조사 결과를 바탕으로 한 후속 연구 및 정책 수립, 관련 프로그램 개발 등은 국민의 국어능력 연구의 깊이를 더할 뿐 아니라 국가적인 정책의 마련 및 유지에도 도움을 줄 수 있을 것으로 기대된다. 이에 본 연구를 시작으로 진행되는 5년 주기의 연차별·영역별 국어능력 조사의 5년 차에서는 4년간의 조사 결과를 종합하고 이를 심층적으로 분석하는 가운데, 최적화되고 효과적인 정책과 프로그램 등을 제언하여 조사 결과가 유의미하게 활용될 수 있도록 해야 할 것이다. 또한 최종적으로는 국민의 국어능력 향상을 위한 프로그램과 정책의 효과를 면밀히 검토하는 것까지 나아갈 수 있도록 체계적이고 엄밀한 실행 계획을 수립할 필요가 있다. 아울러 정책과 프로그램에 대한 효과를 상시로 검증하여 이후 5년 주기의 영역별·연차별 국민의 국어능력의 실태 조사 시 보완점을 반영할 수 있어야 할 것이다.

4) 국어능력의 전체 등급화 및 특화된 배경 설문과 연계한 분석

전술하였듯 2023년 본 조사의 국어능력 등급화를 위해서는 2013년 1차 조사에서 적용한 앵고프(Angoff) 방식을 따른다. 앵고프 방식은 전문가 패널 집단이 영역별로 각 문항에 대한 기대 정답률을 판별하여 분할 점수를 산출하는 것이며, 이는 절대적인 능력 수준을 구분하는 기준을 산출하는 방식이므로 2023년 본 조사 결과를 통해 도출한 국어능력 수치를 등급화하기 위해서는 2013년 1차 조사의 기준을 동일하게 적용해야 한다. 참고로 1차 조사에서는 ‘우수’, ‘보통’, ‘기초’, ‘기초 미달’로 분할하였으며, 2018년 2차 조사에서는 2013년도의 분할 점수를 준용하되 조사 도구 간 동등화를 거쳐 분할 점수를 새로이 도출하고 각각의 등급을 ‘4수준’, ‘3수준’, ‘2수준’, ‘1수준’으로 명명한 바 있다.

이에 2023년 본 조사 역시 가교 문항을 바탕으로 이전 검사와의 동등화 및 분할 점수 산출을 진행하고, 이를 통해 도출된 분할 점수를 바탕으로 결과를 등급화할 수 있을 것이다. 이 과정에서 원년 조사에서 설정한 분할 점수 및 가교 문항을 준용하여야 함은 물론이다. 더하여 앞으로 연차별, 영역별로 국어능력 조사가 수행됨에 따라, 각 조사가 실시될 때마다 영역별로 등급화를 실시할 수 있을 것이며, 전 영역의 조사가 실시된 후에는 전체 국어능력의 등급화가 가능하게 된다. 이때, 전체 국어능력 점수를 등급

화할 경우에는 만점 및 문항당 배점이 영역별로 다르기 때문에 영역별로 합산한 등급이 같음에도 전체 국어능력 총점이 달라 전체 국어능력에서는 서로 다른 성취 수준으로 구분되는 문제가 발생할 수 있다. 따라서 2013년 조사의 총점 기준보다는 2018년 조사에서 도출하였던 전체 국어능력 점수 분할 기준을 적용하는 것이 타당할 것으로 생각된다.

아울러 1, 2차 조사에서는 4개 영역 통합 조사를 실시하다 보니 해당 영역에 적합한 배경 설문 조사를 실시하지 못하였다. 구체적으로 1차 조사인 2013년 조사에서는 인구통계학적 변인만을 배경 설문의 변인으로 삼은 데 비하여, 2차 조사인 2018년 조사에서는 ‘독서량’, ‘ICT 친숙도’, ‘최종 전공’, ‘국어 관련 학습 활동’ 등 인구통계학적 변인 외에 여러 심층 변인을 추가하여 배경 설문을 실시하였다는 점에서 의의가 있다. 하지만 이 역시 읽기 영역과 관련된 변인들의 비중이 높고, 각 변인들을 다소 포괄적인 문항으로 묻고 있어 제한적이라고 볼 수 있다. 따라서 영역별로 조사에 적합한 심층 변인이 설정될 필요가 있다. 또한, 4개 영역의 통합적인 변인을 조사하는 방식이 아니기 때문에, 각 변인들에 대한 설문을 보다 구체적으로 설계할 필요가 있다. 무엇보다 영역별 심층 변인을 개발하고, 이를 바탕으로 각 영역별 조사 결과를 해석한다면 영역별로 전문적이고 심층적인 해석이 가능할 것이다.

(1) 인구통계학적 변인에 따른 결과 분석

이와 관련하여 영역별 평가 결과는 기존의 1, 2차 조사와 마찬가지로 성별, 연령대, 지역 규모, 직업군, 학력에 따른 변인별 분석이 진행되어야 한다. 2023년 본 조사는 듣기·말하기 영역의 평가가 실시된다는 점에서 다음과 같은 2018년 국민의 국어능력 듣기, 말하기 영역 평가의 변인별 차이를 2023년 본 조사 후의 결과와 함께 고려하여 추이를 살필 필요가 있다.

<듣기 영역>

- 성별: 남성이 여성에 비해 평균 점수가 약간 높지만 유의도가 0.05보다 커 유의미한 차이는 없음.
- 연령대: 통계적으로 유의한 차이는 없음. 다만, 20대부터 50대까지 연령대가 높아질수록 듣기 영역의 점수가 다소 낮아지는 현상을 확인.
- 지역 규모: 평균 점수는 읍면 지역, 대도시, 중소도시 순으로 나타났지만, 지역 간 차이에서 통계적으로 유의한 결과는 나타나지 않음.
- 직업군: 평균 점수는 기타, 육체노동, 정신노동자 순으로 나타났지만, 통계적으로 유의한 차이는 없음.
- 학력: 평균 점수는 대재 이상 집단이 가장 높게 나타났으며 고졸 집단이 가장 낮은 것으로 나타남. 하지만 평균 점수 차이 검정 결과 통계적으로 유의한 차이가 나타나지 않음.

<말하기 영역>

- 성별: 남성이 여성에 비해 평균 점수가 약간 높지만 유의도가 0.05보다 커 유의한 차이는 없음.
- 연령대: 통계적으로 유의한 차이가 있음. 사후 비교에서 30대와 60대의 차이와 40대와 60대의 차이에서 통계적으로 유의한 차이를 보임.
- 지역 규모: 평균 점수는 중소도시, 대도시, 읍면 지역 순으로 나타났지만, 지역 간 차이에서 통계적으로 유의한 결과는 나타나지 않음.

- 직업군: 평균 점수는 정신노동, 육체노동, 기타 순으로 나타났지만, 통계적으로 유의한 차이는 없음.
- 학력: 평균 점수는 대재 이상 집단이 가장 높게 나타났으며 고졸 미만 집단이 가장 낮은 것으로 나타남. 평균 점수 차이 검정 결과 또한 통계적으로 유의한 차이가 나타나서 사후 비교를 한 결과 고졸 미만이 고졸과 대재 이상과 통계적으로 유의한 차이가 있는 것으로 분석됨.

(2) 특화된 배경 설문과 연계한 결과 분석

전술한바, 1, 2차 조사에서는 영역 통합 조사를 실시하여 특정 영역에 특화된 배경 설문 조사를 실시하지 못하였다. 이에 본 연구에서는 영역 특성의 배경 설문을 구성해야 함을 제안하였으며, 따라서 본 조사 결과를 배경 설문과 연계하여 분석하는 일이 요구된다. 또한 배경 설문 구성 시 다음의 국내외 평가 사례의 설문 배경 변인을 고려하되, 평가하고자 하는 영역을 보다 심층적으로 파악할 수 있는 설문으로 구성이 이루어져야 할 것이다.

<표 VI-6> 국내외 평가 사례의 설문 배경 변인

	2013년 국민의 국어능력 평가	2018년 국민의 국어능력 평가	PIAAC (OECD)	PISA (OECD)	2008년 기초문해력 조사 (국립국어원)	2015년 국민의 언어의식 조사 보고서 (국립국어원)
인구통계학적 변인	○	○	○	○	○	○
출생지			○			○
교육 수준			○		○	○
최종 전공		○				
모국어 및 외국어 능력			○	○	○	○
가족 배경			○	○		
고용 상태			○			
소득 수준			○		○	○
업무 관련 능력			○			
문식력과 수리력, 독서량		○	○	○	○	○
국어 관련 학습 활동		○				
컴퓨터 이용 ICT 친숙도		○	○	○	○	○
가족 구성원			○			○

참고 문헌

- 강진숙, 배현순, 김지연, 박유신, 염지홍(2019), 미디어 리터러시 교육과정 운영을 통한 시민역량 제고 방안 연구, 교육부.
- 구자옥, 김성숙, 이해원, 조성민, 박혜영(2016), OECD 국제 학업성취도 평가 연구: PISA 2018 예비검사 시행 기반 구축, 연구보고서(한국교육과정평가원 RRE 2016-2-1).
- 국립국어원(2005), 국민의 국어능력 평가를 위한 기초 연구, 국립국어원.
- 국립국어원(2006), 국어능력 조사 방법 연구, 국립국어원.
- 국립국어원(2008), 국어능력검정 모의시험 방법 연구, 국립국어원.
- 국립국어원(2009), 국어능력검정시험에서 말하기 능력 평가 방법 개발을 위한 연구, 국립국어원.
- 국립국어원(2012), 국어능력 평가를 위한 기초 연구, 국립국어원.
- 길호현, 이지선(2022), 빅데이터 기사문에 나타난 리터러시 개념 사용 양상, 독서연구, 65, 151-199.
- 김남희(2012), PISA 읽기 소양과 21세기 국어 능력, 국어교육, 138, 41-71.
- 김대희(2007), 국어 교육에서 미디어 교육 수용 방안 연구: 초등 교육을 중심으로, 고려대학교 박사학위논문.
- 김선, 김성훈, 반재천(2009), 검사동등화를 이용한 앵고프 방법과 북마크 방법의 비교, 교육평가 연구, 22(4), 1035-1055.
- 김아미(2015), 미디어 리터러시 교육의 이해, 서울: 커뮤니케이션북스.
- 김양은(2007), 미디어교육 교육과정 모델 구성에 관한 연구, 한국언론정보학보, 37, 73-99.
- 김윤주(2021), 문식성 연구 동향 분석과 향후 과제 -지난 25년간(1996~2021)의 학위논문을 중심으로-, 우리어문연구, 70, 253-282.
- 김종윤, 서수현, 김인숙, 조병영, 김지연, 유상희, 김희동, 오은하, 옥현진(2017), 디지털 리터러시의 인지적 영역 평가도구 개발을 위한 기초 연구, 청람어문교육, 62, 7-39.
- 김종철, 민병곤, 박현정, 정희창, 김봉순, 박재현, 주재우, 이관희, 이지수, 박은진, 정재미(2013), 2013년 국민의 국어능력 평가, 연구보고서(국립국어원 2013-01-29).
- 김창영, 이은빈, 이승준, 김지은, 신예림, 조태린, 강현철, 양수경, 김주성, 심주희(2020), 2020년 국민의 언어 의식 조사, 연구보고서(국립국어원 2020-01-35).
- 김창원, 서혁, 윤준채, 이관규, 정건지, 김순임(2008), 국민의 기초 문해력 조사, 연구보고서(국립국어원 2008-01-57).
- 김평원(2007), 의사소통과정으로서의 말하기 평가: 대규모 성인 말하기 평가 시스템을 중심으로, 화법연구, 11, 9-34.
- 김현진, 김현영, 김은영, 류향승, 김보운, 박희락, 최미애(2020), 민주시민육성을 위한 미디어리터러시 역량의 의미와 정규교육과정 통합 방안, 학습자중심교과교육연구, 20(11), 473-500.
- 노들, 옥현진(2020), 텍스트 마이닝 기법을 통한 미디어 리터러시와 디지털 리터러시 개념의 비교 분석 - 신문 기사를 중심으로, 리터러시연구, 11(5), 103-129.
- 노은희, 신호재, 이재진, 정현선(2018), 교과 교육에서의 디지털 리터러시 교육 실태 분석 및 개

- 선 방안 연구, 연구보고서(한국교육과정평가원 RRC 2018-7).
- 민병곤, 김종철, 구본관, 손원숙, 김혜정, 박재현, 김승현, 서보영, 이문수, 장지혜, 이효정, 차경미, 강지영, 김동섭, 허모아(2018), 2018년 국민의 국어능력 실태 조사, 연구보고서(국립국어원 2018-01-53).
- 민병곤, 김종철, 정희창, 박현정, 김혜정, 박재현, 김승현, 서보영, 이기연, 장지혜, 김민재, 차경미, 박민호, 신유진, 김동섭, 허모아(2017), 국민의 국어능력 실태 조사 도구 개발, 연구보고서(국립국어원 2017-01-51).
- 민병곤, 박재현(2015), 고등학생과 성인의 말하기 능력 비교 연구: 정보 전달과 설득적 말하기를 중심으로, 국어교육연구, 35, 283-320.
- 민병곤, 정재미, 박재현(2014), 성인의 국어 말하기 능력 평가 결과 분석, 화법연구, 24, 101-136.
- 박재현(2015), 화법 평가의 쟁점과 발전 방향, 국어교육학연구, 50(2), 6-25.
- 박재현(2016), 국어교육을 위한 의사소통 이론, 사회평론.
- 박주현(2018), 디지털 독서 및 정보 리터러시 평가 문항 분석을 통한 독서 및 정보 서비스의 방향 탐색 -PIAAC와 PISA를 중심으로, 한국문헌정보학회지, 52(3), 61-89.
- 박주현(2021), PISA 2018 독서 리터러시 평가를 통한 디지털 정보 서비스 방안 탐색, 한국문헌정보학회지, 55, 135-159.
- 박종임(2021), 디지털 미디어 리터러시 교육 개선을 위한 국어과 교육과정 현황 분석, 청람어문교육, 81, 7-36.
- 백원근, 이순영, 이기재, 강현철, 이용훈, 안찬수, 김종명, 전상현, 김경희, 정병현(2020), 2019년 국민 독서실태 조사, 문화체육관광부.
- 변숙자(2021), 초·중·고 학생의 미디어 이용 양상 및 미디어 리터러시 교육에 관한 인식 분석, 리터러시연구, 12(3), 89-121.
- 서영진(2012), 상호교섭적 논증 교육의 내용 구성 연구, 부산대학교 박사학위논문.
- 서현석(2007), '친교적 화법' 수업의 내용과 실행 양상 연구, 새국어교육, 75, 193-214.
- 성태제(1994), 대학별고사를 위한 문항분석, 표준점수, 검사동등화, CSAM(Communications for Statistical Applications and Methods), 1(1), 206-214.
- 옥현진, 오은하, 김종윤(2018), 중학생 학습자를 위한 디지털 리터러시 인지적 영역 성취기준 개발, 국어교육연구, 41, 81-112.
- 이동후(2016), 말의 기술화와 디지털 구술성, 제33회 한국화법학회 전국학술대회 자료집.
- 이애화(2015), 디지털 리터러시 교육을 위한 디지털 역량의 개념적 특성과 한계, 교육문화연구, 21(3), 179-200.
- 이창덕(2008), 화법 능력과 화법(듣기, 말하기) 교과서 개발 원리, 화법연구, 12, 113-148.
- 장소영, 신동일(2009), 언어교육평가 연구를 위한 FACETS 프로그램 : 기초 과정편, 글로벌콘텐츠.
- 장은주(2020), 2015 개정 교육과정의 미디어교육 관련 성취기준 분석 연구, 새국어교육, 125, 39-68.
- 정상섭(2005), 공감적 듣기의 듣기교육적 수용 연구, 한국초등국어교육, 28, 277-305.

- 정현선(2014), 복합양식 문식성 교육의 의의와 방법, 우리말교육현장연구, 8(2), 61-93.
- 정현선, 김아미, 박유신, 전경란, 이지선, 노자연(2016), 핵심역량 중심의 미디어 리터러시 교육 내용 체계화 연구, 학습자중심교과교육연구, 16(11), 211-238.
- 조성민, 구남욱, 김현정, 이소연, 이인화(2019), OECD 국제 학업성취도 평가 연구: PISA 2018 결과 보고서, 연구보고서(한국교육과정평가원 RRE 2019-11).
- 조재운(2007), 국어교육: 말하기 평가의 요소 설정 연구-델파이 기법을 이용하여, 새국어교육, 75, 337-358.
- 통계청(2020), 2019년 생활시간조사 결과, 통계청 보도자료(보도일시: 2020.7.30.).
- 한국언론진흥재단(2020), 2020 언론수용자 조사, 한국언론진흥재단.
- 홍은실, 장민정, 오문경, 유하라, 현원숙, 김경희(2018), 역량 기반 학문 목적 한국어 듣기 교육 과정연구, 화법연구, 36, 213-239.
- 황연주(2001), 영상 정보화 시대에 대처하는 미술교육에서의 비주얼 리터러시 교육, 미술교육논총, 12, 137-156.
- Gilster, P. (1997). Digital literacy. New York: Wiley Computer Pub. 김정래 역(1999), 디지털 리터러시, 해냄출판사.
- Hobbs. R. (2010). Digital and media literacy: A plan of action. Washington, D. C.: The Aspen Institute.
- OECD (2011). PISA 2009 Results: Students on Line: Digital Technologies and Performance (Volume VI).
- OECD (2019). PISA 2018 Assessment and Analytical Framework. OECD Publishing.
- OECD (2021). The Assessment Frameworks for Cycle 2 of the Programme for the International Assessment of Adult Competencies. OECD Publishing.
<https://doi.org/10.1787/4bc2342d-en>.
- Ravichandiran, S. (2021). Getting Started with Google BERT. Packt Publishing. 전희원, 정승환, 김형준 역(2021), 구글 BERT의 정석, 한빛미디어.
- Werlich, E. (1976). A Text Grammar of English. Heidelberg: Quelle and Meyer.
- Wilkinson, A. (1965). Spoken English. Birmingham. WM: University of Birmingham Educational Review.
- Yen, W. (1981). Using simulation results to choose a latent trait model. Applied Psychological Measurement, 5, 245-262.

<Abstract>

Fundamental Research on the Improvement of the Survey on Korean Language Proficiency for the general public

The purpose of this study is to develop a proposal to upgrade the survey by evaluating the performance of the 1st and 2nd survey, to develop survey tools in the listening and speaking sections, and to search how to utilize the survey results. To this end, four detailed research objectives were set: 'performance evaluation and status analysis, development of tools in the listening and speaking sections, elaboration of survey tools, and proposal of implementation of the main survey'. With these four objectives, we tried to specify measures to upgrade the survey on Korean language proficiency for the general public by section and year, to establish the scope and concept of Korean language proficiency by section according to the changed Korean language use environment, and to lay the theoretical and methodological foundation for the survey.

The 3rd survey on the Korean language proficiency for the general public is a large-scale national survey located in a series of continuity and stages following the 1st (2013) and 2nd (2018) survey, and this study is a fundamental research and a preliminary survey for the 3rd survey. In other words, further from the first survey that established a regular survey system and the second survey that attempted to develop and stabilize survey tools, it is a study to establish a proposal to upgrade the annual survey by language proficiency section, which focuses on the first section of the divided survey, 'listening and speaking'.

First, for establishing a proposal to upgrade the survey on Korean language proficiency by section and year, the concept of Korean language proficiency was established by analyzing the environment and status of Korean language, and domestic and foreign survey cases related to native language proficiency, listening and speaking proficiency, and digital literacy were reviewed to derive implications for evaluation sections / factors, and survey subjects / methods. In particular, digital literacy is emphasized to reflect the changing literacy environment in various surveys, including the PIAAC and PISA. Therefore, in this study, it was suggested that digital literacy-related elements should be composed of one content element in the survey or designed as a separate survey section in the future, and that the possibility of multilateral interpretation of the survey results should be secured through a background variable survey.

Based on the results of the above analysis on performance and status, the 'listening and speaking' survey tool was developed as follows. First of all, the listening and speaking proficiency survey was designed to strengthen the practicality of spoken communication and the need for cooperative communication, while applying the existing

framework for listening and speaking survey to maintain consistency with the 2018 survey. The listening and speaking survey section was largely divided into 'listening', 'listening and speaking', and 'speaking' sections, and detailed survey sections and elements were set according to the characteristics of each section. Specifically, in the case of the listening section, in addition to the existing factual, reasoning, and critical listening, sympathetic listening was added to the mid-range evaluation items. In the case of speaking, it was revised and supplemented by excluding 'requesting speech' in the existing 'persuading speech (arguing speech, requesting speech)' and adding 'presentation' to 'explanatory speech (explaining experiences or concepts)', taking into account the proportion, social importance of discourse, and changes in the media environment. Finally, in the listening and speaking section, to reflect the integration of language and the dynamics of spoken language discourse, the mid-range section 'expression of friendship' was newly established, and the sub sections 'considerate speech' and 'negotiation' were established, focusing on intimate conversations that often appear in daily conversations.

Considering the number, time, and allocation of items by section of the "2018 Korean Language Proficiency Survey for the General Public", the survey tool of the next year's listening and speaking section will be 15 listening items (cross-linking 2, new 13), 2 listening and speaking items (new 2), and 3 speaking items (cross-linking 1, new 2). Accordingly, the first preliminary survey tool was developed a double number of items to be used in the main survey, including cross-linking items, and based on the results of the first preliminary survey analysis, 1.5 multiple items were selected, corrected, and used as a second preliminary survey tool. After that, based on the results of the second preliminary survey, 1.2 multiple items were finalized. The items developed and selected in this way consisted of (㉠) and (㉡) type test sheets and were used for the first and second preliminary surveys of 480 men and women in their 20s~60s (1st: 240 people, 2nd: 240 people) living in Seoul, Incheon, and Gyeonggi. In the listening section composed of multiple choice items, response data were collected by entering the answer directly into the CAPI system by touching a specific choice on the tablet PC screen, and in the listening/speaking, and speaking sections composed of performance-type items, the data were collected by recording responses.

Each in the first and second preliminary surveys, 3,600 cases (1,440 cases of listening and speaking, 2,160 cases of speaking) were scored, and three Korean language education majors were designed to jointly score each item to secure reliability among the scorers. In addition, web-based training (scorer workshop) was conducted once each in the first and second preliminary surveys to meet the level of use of the scale. Item difficulty and discrimination by section were analyzed by applying a two-parameter item response theory model in the listening section (multiple choice), and in the case of the listening/speaking, and speaking sections (performance-type), the item quality was determined by using logit scale value of the Many-Facet Rasch Model analysis result. In addition, the item discrimination was calculated by applying a two-parameter generalized partial credit model

according to the item response theory among the Facets measurement results. As a result of the reliability analysis between the scorers, in both the first and second preliminary surveys, the interclass correlation in the listening and speaking section was a good level of 0.7 or higher, and in the speaking section was an excellent level of 0.8 or higher.

Looking at the results of the first preliminary survey (September to October), the score distribution was close to the normal distribution in all three sections. However, the listening, listening, and speaking sections showed a slightly negative skewness to the right, and the speaking section showed a slightly positive skewness to the left. Most of the items in the listening section tended to be good in difficulty and discrimination, but some of the items showed low discrimination with less than .35, or low difficulty with less than -2.0, so 1.5 multiple items were selected by modifying or deleting those items. Overall, the application of the scoring scale was found to be stable in the listening and speaking sections. However, in the case of type (ㄴ), the use of the scoring scale was relatively inconsistent, and the item difficulty and discrimination were lower than other items. In addition, considering that there was an adjustment to the scoring element of 'content selection' in the process of the scorer training, the item was excluded from the second preliminary survey items. In the case of the speaking section, the (ㄱ) type question 3, which had a reversal phenomenon between discrimination and difficulty, was excluded from the 1.5-multiple items. In this process, the possibility that the topic of discussion (physical punishment for educational purposes) of the item could cause a problem that is disadvantageous to a specific age group or occupational group was also considered.

Next, looking at the results of the second preliminary survey (November to December), the score distribution was close to the normal distribution in all three sections, as in the results of the first preliminary survey. The listening section showed a slightly negative skewness to the right, but the normality was satisfied with a significance probability of .064. The listening and speaking sections also showed negative skewness, but it was close to the normal distribution, and the speaking section showed a rather pointed positive skewness with a slight bias to the left. In terms of item difficulty and discrimination by section, most of the listening items were good, but some difficult items showed high discrimination and difficulty. If there are no specific problems with discrimination and difficulty, 1.2-multiple items which were 15 main items and 3 preliminary items (18 in total) were selected, focusing on composing questions with various difficulties. Overall, the listening and speaking sections showed low difficulty and good discrimination. The difficulty parameter of the test paper (ㄴ) type was closer to the average than the (ㄱ) type, so the two items that were composed of (ㄴ) type were selected as the final 1.2-multiple items. In addition, in the speaking section, the discrimination and difficulty level of question 2 were good in both types (ㄱ) and (ㄴ), so it was decided to include both two types in 1.2-multiple items.

Based on the above findings, several points of discussion proposed in this study are as

follows. First, it is related to the establishment of the main survey proposal in the listening and speaking section. In terms of survey tool configuration, the size of a single survey tool can be constructed similar to that of a preliminary survey in this study, but further discussion is needed on whether to deploy a single survey tool or a dual survey tool. In addition, further discussions are needed on how to conduct the 3rd and 4th-year surveys, the survey time considering the age group, the examination scope of surveyees by section, and whether to use the artificial intelligence diagnosis system. In terms of the number of surveyees and methods of the survey, it was proposed to reduce the number of samples in consideration of the fact that the survey was conducted in each section of the language proficiency, not the entire language proficiency. In addition, in order to improve the survey environment and participation of the surveyees, it was proposed to switch from the existing 1:1 visit survey method to the visit group survey, and it was mentioned that appropriate incentives were needed. In terms of grading the achievement level, it is necessary to attempt to calculate the equivalence and cut score with the previous test based on cross-linking items by using the Angoff method applied to the first survey in 2013.

Second, it is related to the direction of regularization and establishment of improvement plans of the 5-year periodical Korean language proficiency survey. This study 1) proposed the 'annual survey proposal for each section' of Korean language proficiency based on the divided survey. Furthermore, 2) in the process of establishing a proposal to upgrade Korean language proficiency survey by section and year, we tried to devise a link with the task of 'development of a Korean language proficiency diagnosis system using artificial intelligence' as part of the mid- to long-term plan. In addition, 3) an effective policy establishment proposal was devised through follow-up studies to enable in-depth investigation and reflux policies for each subdivided area, and 4) analysis and interpretation of survey results linked to demographic variables and in-depth variables for each section were proposed.

Key-words: Korean language proficiency, Korean language proficiency survey, Korean language proficiency survey for the general public, listening proficiency, speaking proficiency, digital literacy

[부록] 2023년 국민의 국어능력 실태 본 조사 시행을 위한 최종 문항

1) 듣기 영역

2022 국민의 국어능력 실태 조사		1차 예비 조사		L-1(공통)	2차 예비 조사		L-1(공통)	
평가 문항								
자료 출처								
평가 요소 텍스트 정보		비판적 듣기(내용의 평가)_설득적 담화						
정답								
1차 예비 조사 결과		변별도	3.733	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)
		곤란도	-.184	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)
		문항 반응	① 8.8%		② 21.3%		③ 13.3%	
2차 예비 조사 결과		변별도	.342	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)
		곤란도	-1.805	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)
		문항 반응	① 8.3%		② 10.4%		③ 16.7%	

2022 국민의 국어능력 실태 조사		1차 예비 조사	L-2(공통)	2차 예비 조사	L-2(공통)
평가 문항					
자료 출처					
평가 요소 텍스트 정보	추론적 듣기(적용)-설득적 담화				
정답					

1차 예비 조사 결과	변별도	.288	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)							
	곤란도	-3.503	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)							
	문항 반응	<table><tr><td>①</td><td>②</td><td>③</td><td>④</td></tr><tr><td>5.8%</td><td>7.1%</td><td>72.9%</td><td>14.2%</td></tr></table>						①	②	③	④	5.8%	7.1%	72.9%
①	②	③	④											
5.8%	7.1%	72.9%	14.2%											
2차 예비 조사 결과	변별도	.763	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)							
	곤란도	-1.681	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)							
	문항 반응	<table><tr><td>①</td><td>②</td><td>③</td><td>④</td></tr><tr><td>4.6</td><td>11.7</td><td>75.8</td><td>7.9</td></tr></table>						①	②	③	④	4.6	11.7	75.8
①	②	③	④											
4.6	11.7	75.8	7.9											

2022 국민의 국어능력 실태 조사		1차 예비 조사		L-3(가)		2차 예비 조사		L-3(공통)	
평가 문항									
자료 출처									
평가 요소 텍스트 정보		사실적 듣기(지시의 이행 능력) / 지시(설명)							
정답									
1차 예비 조사 결과		변별도	1.603	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)	
		곤란도	-2.099	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)	
		문항 반응	① 0%		② 91.7%		③ 7.5%		④ 0.8%
2차 예비 조사 결과		변별도	3.821	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)	
		곤란도	-2.315	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)	
		문항 반응	① 1.7%		② 98.3%		③ 0%		④ 0%

2022 국민의 국어능력 실태 조사		1차 예비 조사	L-4(나)	2차 예비 조사	L-4(공통)
평가 문항					
자료 출처					
평가 요소 텍스트 정보	사실적 듣기(세부 정보의 파악) / 면담				
정답					

1차 예비 조사 결과	변별도	1.813	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)							
	곤란도	-1.186	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)							
	문항 반응	<table><tr><td>①</td><td>②</td><td>③</td><td>④</td></tr><tr><td>5.0%</td><td>4.2%</td><td>10.0%</td><td>80.8%</td></tr></table>						①	②	③	④	5.0%	4.2%	10.0%
①	②	③	④											
5.0%	4.2%	10.0%	80.8%											
2차 예비 조사 결과	변별도	1.182	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)							
	곤란도	-1.481	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)							
	문항 반응	<table><tr><td>①</td><td>②</td><td>③</td><td>④</td></tr><tr><td>5.8%</td><td>5.8%</td><td>9.2%</td><td>79.2%</td></tr></table>						①	②	③	④	5.8%	5.8%	9.2%
①	②	③	④											
5.8%	5.8%	9.2%	79.2%											

2022 국민의 국어능력 실태 조사		1차 예비 조사		L-5(나)	2차 예비 조사		L-5(나)
평가 문항							
자료 출처							
평가 요소 텍스트 정보	사실적 듣기(세부 정보의 파악) / 뉴스(설득)						
정답							
1차 예비 조사 결과	변별도	1.593	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)
	곤란도	-.658	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)
	문항 반응	① 4.2%		② 5.8%		③ 67.5%	
2차 예비 조사 결과	변별도	1.330	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)
	곤란도	-.647	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)
	문항 반응	① 4.2%		② 5.8%		③ 67.5%	

	문항 반응	①	②	③	④
		2.5%	7.5%	64.2%	25.8%

2022 국민의 국어능력 실태 조사		1차 예비 조사		L-6(나)	2차 예비 조사		L-6(나)
평가 문항							
자료 출처							
평가 요소 텍스트 정보	사실적 듣기(핵심 정보의 파악) / 강연(설득)						
정답							
1차 예비 조사 결과	변별도	1.426	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)
	곤란도	-.456	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)
	문항 반응	① 61.7%	② 9.2%	③ 7.5%	④ 21.7%		
2차 예비 조사 결과	변별도	1.824	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)
	곤란도	-.857	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)

	문항 반응	①	②	③	④
		71.7%	11.7%	2.5%	14.2%

2022 국민의 국어능력 실태 조사		1차 예비 조사	L-7(나)	2차 예비 조사	L-7(공통)									
평가 문항														
자료 출처														
평가 요소 텍스트 정보	사실적 듣기(핵심 정보의 파악) / 대화(설명)													
정답														
1차 예비 조사 결과	변별도	3.660	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)							
	곤란도	-1.620	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)							
	문항 반응	<table> <tr> <td>①</td> <td>②</td> <td>③</td> <td>④</td> </tr> <tr> <td>6.7%</td> <td>93.3%</td> <td>0%</td> <td>0%</td> </tr> </table>						①	②	③	④	6.7%	93.3%	0%
①	②	③	④											
6.7%	93.3%	0%	0%											
2차 예비 조사 결과	변별도	7.238	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)							
	곤란도	-4.185	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)							
	문항 반응	<table> <tr> <td>①</td> <td>②</td> <td>③</td> <td>④</td> </tr> <tr> <td>0%</td> <td>100.0%</td> <td>0%</td> <td>0%</td> </tr> </table>						①	②	③	④	0%	100.0%	0%
①	②	③	④											
0%	100.0%	0%	0%											

2022 국민의 국어능력 실태 조사		1차 예비 조사		L-8(나)		2차 예비 조사		L-8(공통)															
평가 문항																							
자료 출처																							
평가 요소 텍스트 정보		추론적 듣기(내용의 추리) / 토론(설득)																					
정답																							
1차 예비 조사 결과	<table><tr><td>변별도</td><td>.957</td><td>거의 없음 (0.00 ~ 0.35)</td><td>낮음 (0.35 ~ 0.65)</td><td>적절함 (0.65 ~ 1.35)</td><td>높음 (1.35 ~ 1.70)</td><td>매우 높음 (1.70 이상)</td></tr><tr><td>곤란도</td><td>.602</td><td>매우 쉬움 (-2.0 이하)</td><td>쉬움 (~ -0.5)</td><td>중간 (~0.5)</td><td>어려움 (~2.0)</td><td>매우 어려움 (2.0 이상)</td></tr></table>									변별도	.957	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)	곤란도	.602	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)
	변별도	.957	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)																
	곤란도	.602	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)																
문항 반응	① 38.3%		② 12.5%		③ 37.5%		④ 11.7%																
2차 예비 조사 결과	<table><tr><td>변별도</td><td>.126</td><td>거의 없음 (0.00 ~ 0.35)</td><td>낮음 (0.35 ~ 0.65)</td><td>적절함 (0.65 ~ 1.35)</td><td>높음 (1.35 ~ 1.70)</td><td>매우 높음 (1.70 이상)</td></tr><tr><td>곤란도</td><td>6.411</td><td>매우 쉬움 (-2.0 이하)</td><td>쉬움 (~ -0.5)</td><td>중간 (~0.5)</td><td>어려움 (~2.0)</td><td>매우 어려움 (2.0 이상)</td></tr></table>									변별도	.126	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)	곤란도	6.411	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)
	변별도	.126	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)																
곤란도	6.411	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)																	

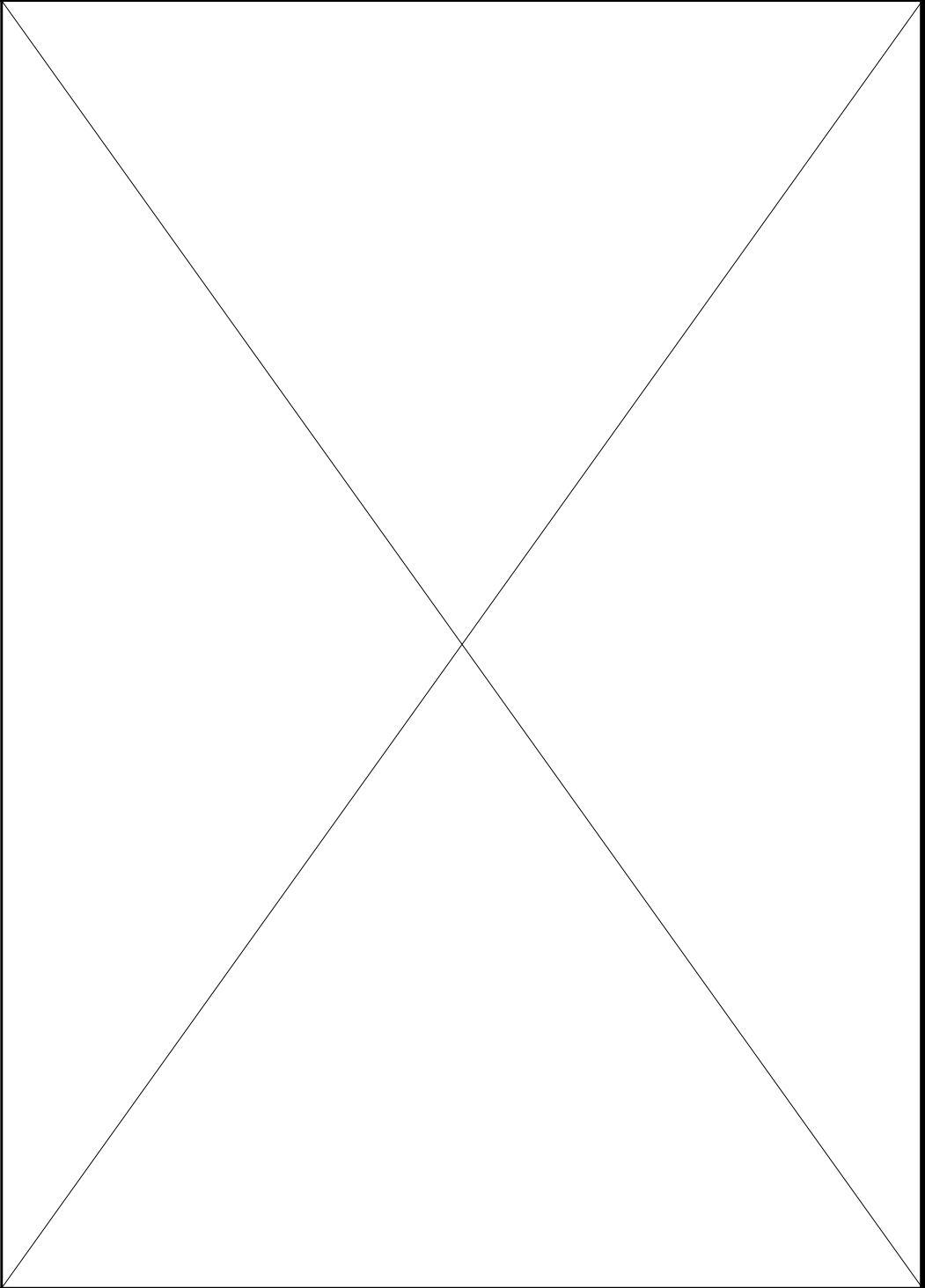
	문항 반응	①	②	③	④
		30.8%	11.7%	42.5%	15.0%

2022 국민의 국어능력 실태 조사		1차 예비 조사		L-9(나)		2차 예비 조사		L-9(나)															
평가 문항																							
자료 출처																							
평가 요소 텍스트 정보		추론적 듣기(심리 또는 태도의 추리) / 대화(설득)																					
정답																							
1차 예비 조사 결과	<table><tr><td>변별도</td><td>2.156</td><td>거의 없음 (0.00 ~ 0.35)</td><td>낮음 (0.35 ~ 0.65)</td><td>적절함 (0.65 ~ 1.35)</td><td>높음 (1.35 ~ 1.70)</td><td>매우 높음 (1.70 이상)</td></tr><tr><td>곤란도</td><td>-1.684</td><td>매우 쉬움 (-2.0 이하)</td><td>쉬움 (~ -0.5)</td><td>중간 (~0.5)</td><td>어려움 (~2.0)</td><td>매우 어려움 (2.0 이상)</td></tr></table>									변별도	2.156	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)	곤란도	-1.684	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)
	변별도	2.156	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)																
	곤란도	-1.684	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)																
문항 반응	<table><tr><td>①</td><td>②</td><td>③</td><td>④</td></tr><tr><td>4.2%</td><td>90.8%</td><td>2.5%</td><td>2.5%</td></tr></table>								①	②	③	④	4.2%	90.8%	2.5%	2.5%							
①	②	③	④																				
4.2%	90.8%	2.5%	2.5%																				

2차 예비 조사 결과	변별도	2.876	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)
	곤란도	-.844	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)
	문항 반응	①		②		③	
		8.3%		74.2%		9.2%	
		④					
						8.3%	

2022 국민의 국어능력 실태 조사	1차 예비 조사	L-10(가)	2차 예비 조사	L-10(가)
평가 문항				

자료 출처							
평가 요소 텍스트 정보	추론적 듣기(심리 또는 태도의 추리)/ 설명(면담)						
정답							
1차 예비 조사 결과	변별도	.703	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)
	곤란도	-2.836	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)
	문항 반응	① 3.3%		② 85.8%		③ 6.7%	
2차 예비 조사 결과	변별도	1.464	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)
	곤란도	.543	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)
	문항 반응	① 3.3%		② 37.5%		③ 38.3%	

2022 국민의 국어능력 실태 조사		1차 예비 조사	L-11(가)	2차 예비 조사	L-11(가)
평가 문항					
자료 출처					
평가 요소 텍스트 정보					
정답					

1차 예비 조사 결과	변별도	1.166	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)							
	곤란도	-2.169	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)							
	문항 반응	<table><tr><td>①</td><td>②</td><td>③</td><td>④</td></tr><tr><td>1.7%</td><td>4.2%</td><td>5.8%</td><td>88.3%</td></tr></table>						①	②	③	④	1.7%	4.2%	5.8%
①	②	③	④											
1.7%	4.2%	5.8%	88.3%											
2차 예비 조사 결과	변별도	1.182	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)							
	곤란도	-1.194	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)							
	문항 반응	<table><tr><td>①</td><td>②</td><td>③</td><td>④</td></tr><tr><td>11.7%</td><td>76.7%</td><td>5.8%</td><td>5.8%</td></tr></table>						①	②	③	④	11.7%	76.7%	5.8%
①	②	③	④											
11.7%	76.7%	5.8%	5.8%											

2022 국민의 국어능력 실태 조사		1차 예비 조사	L-12(가)	2차 예비 조사	L-12(공통)
평가 문항					
자료 출처					
평가 요소 텍스트 정보	비판적 듣기(내용의 평가)/ 설득(토론)				

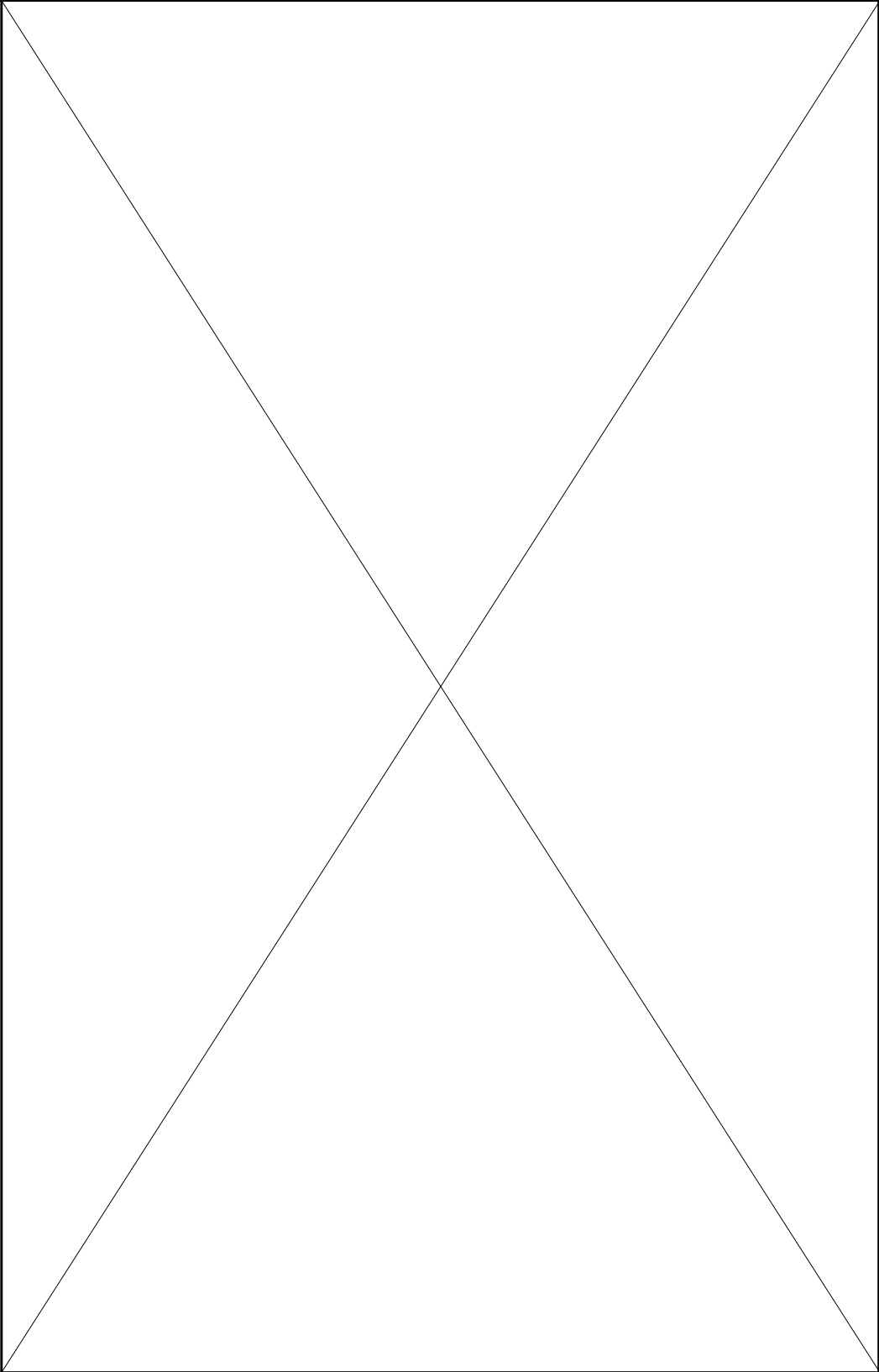
정답							
1차 예비 조사 결과	변별도	.961	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)
	곤란도	-.754	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)
	문항 반응						
		① 13.3%	② 20.0%	③ 2.5%	④ 64.2%		
2차 예비 조사 결과	변별도	.817	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)
	곤란도	-.383	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)
	문항 반응						
		① 15.8%	② 20.0%	③ 8.3%	④ 55.8%		

2022 국민의 국어능력 실태 조사		1차 예비 조사	L-13(나)	2차 예비 조사	L-13(나)
평가 문항					
자료 출처					
평가 요소 텍스트 정보	비판적 듣기(형식의 평가) / 강연(설득)				
정답					

1차 예비 조사 결과	변별도	.983	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)							
	곤란도	-0.529	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)							
	문항 반응	<table><tr><td>①</td><td>②</td><td>③</td><td>④</td></tr><tr><td>15.0%</td><td>14.2%</td><td>60.8%</td><td>10.0%</td></tr></table>						①	②	③	④	15.0%	14.2%	60.8%
①	②	③	④											
15.0%	14.2%	60.8%	10.0%											
2차 예비 조사 결과	변별도	.795	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)							
	곤란도	-.149	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)							
	문항 반응	<table><tr><td>①</td><td>②</td><td>③</td><td>④</td></tr><tr><td>6.7%</td><td>14.2%</td><td>51.7%</td><td>27.5%</td></tr></table>						①	②	③	④	6.7%	14.2%	51.7%
①	②	③	④											
6.7%	14.2%	51.7%	27.5%											

2022 국민의 국어능력 실태 조사		1차 예비 조사		L-14(가)	2차 예비 조사		L-14(가)	
평가 문항								
자료 출처								
평가 요소 텍스트 정보	공감적 듣기(상대방의 입장이나 처지 파악)/ 친교(대화)							
정답								
1차 예비 조사 결과	변별도		1.166	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)
	곤란도		-2.644	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)
	문항 반응	① 1.7%		② 2.5%		③ 3.3%		④ 92.5%

2차 예비 조사 결과	변별도	2.621	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)
	곤란도	-1.438	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)
	문항 반응						
		① 0.8%	② 4.2%	③ 90.0%	④ 5.0%		

2022 국민의 국어능력 실태 조사	1차 예비 조사	L-15(가)	2차 예비 조사	L-15(공통)
평가 문항				
자료 출처				

평가 요소 텍스트 정보	공감적 듣기(상대의 입장이나 처지 파악)/ 설명(대화)							
정답								
1차 예비 조사 결과	변별도	1.027	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)	
	곤란도	-.148	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)	
	문항 반응	①		②		③		④
		5.0%		52.5%		5.0%		37.5%
2차 예비 조사 결과	변별도	1.038	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)	
	곤란도	-.373	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)	
	문항 반응	①		②		③		④
		4.2%		56.7%		10.8%		28.3%

2022 국민의 국어능력 실태 조사		1차 예비 조사		L-6(가)	2차 예비 조사		L-6(가)														
평가 문항																					
자료 출처																					
평가 요소 텍스트 정보	사실적 듣기(핵심 정보의 파악) / 뉴스(설명)																				
정답																					
1차 예비 조사 결과	<table><tr><td>변별도</td><td>1.230</td><td>거의 없음 (0.00 ~ 0.35)</td><td>낮음 (0.35 ~ 0.65)</td><td>적절함 (0.65 ~ 1.35)</td><td>높음 (1.35 ~ 1.70)</td><td>매우 높음 (1.70 이상)</td></tr><tr><td>곤란도</td><td>-.642</td><td>매우 쉬움 (-2.0 이하)</td><td>쉬움 (~ -0.5)</td><td>중간 (~0.5)</td><td>어려움 (~2.0)</td><td>매우 어려움 (2.0 이상)</td></tr></table>							변별도	1.230	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)	곤란도	-.642	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)
	변별도	1.230	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)														
	곤란도	-.642	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)														
문항 반응	<table><tr><td>①</td><td>②</td><td>③</td><td>④</td></tr><tr><td>14.2%</td><td>18.3%</td><td>3.3%</td><td>64.2%</td></tr></table>						①	②	③	④	14.2%	18.3%	3.3%	64.2%							
①	②	③	④																		
14.2%	18.3%	3.3%	64.2%																		

2차 예비 조사 결과	변별도	.884	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)
	곤란도	-.795	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)
	문항 반응	①					
		7.5%					
	문항 반응	②					
		25.8%					
	문항 반응	③					
		0.8%					
	문항 반응	④					
		65.8%					

2022 국민의 국어능력 실태 조사		1차 예비 조사		L-9(가)	2차 예비 조사		L-9(가)															
평가 문항																						
자료 출처																						
평가 요소 텍스트 정보	추론적 듣기(심리 또는 태도의 추리)/ 친교(대화)																					
정답																						
1차 예비 조사 결과	<table><tr><td>변별도</td><td>.955</td><td>거의 없음 (0.00 ~ 0.35)</td><td>낮음 (0.35 ~ 0.65)</td><td>적절함 (0.65 ~ 1.35)</td><td>높음 (1.35 ~ 1.70)</td><td>매우 높음 (1.70 이상)</td></tr><tr><td>곤란도</td><td>-2.401</td><td>매우 쉬움 (-2.0 이하)</td><td>쉬움 (~ -0.5)</td><td>중간 (~0.5)</td><td>어려움 (~2.0)</td><td>매우 어려움 (2.0 이상)</td></tr></table>								변별도	.955	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)	곤란도	-2.401	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)
	변별도	.955	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)															
	곤란도	-2.401	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)															
문항 반응	① 3.3%		② 0.8%		③ 8.3%		④ 87.5%															

2차 예비 조사 결과	변별도	1.914	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)
	곤란도	-.519	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)
	문항 반응						
		①	②	③	④		
		6.7%	7.5%	19.2%	66.7%		

2022 국민의 국어능력 실태 조사		1차 예비 조사	L-10(나)	2차 예비 조사	L-10(나)
평가 문항					
자료 출처					
평가 요소 텍스트 정보	추론적 듣기(심리 또는 태도의 추리) / 토론(설득)				
정답					

1차 예비 조사 결과	<table><tr><td>변별도</td><td>1.405</td><td>거의 없음 (0.00 ~ 0.35)</td><td>낮음 (0.35 ~ 0.65)</td><td>적절함 (0.65 ~ 1.35)</td><td>높음 (1.35 ~ 1.70)</td><td>매우 높음 (1.70 이상)</td></tr><tr><td>곤란도</td><td>-.702</td><td>매우 쉬움 (-2.0 이하)</td><td>쉬움 (~ -0.5)</td><td>중간 (~0.5)</td><td>어려움 (~2.0)</td><td>매우 어려움 (2.0 이상)</td></tr></table>							변별도	1.405	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)	곤란도	-.702	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)
	변별도	1.405	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)														
	곤란도	-.702	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)														
문항 반응	<table><tr><td>①</td><td>②</td><td>③</td><td>④</td></tr><tr><td>8.3%</td><td>67.5%</td><td>14.2%</td><td>10.0%</td></tr></table>						①	②	③	④	8.3%	67.5%	14.2%	10.0%							
①	②	③	④																		
8.3%	67.5%	14.2%	10.0%																		

2차 예비 조사 결과	<table><tr><td>변별도</td><td>2.789</td><td>거의 없음 (0.00 ~ 0.35)</td><td>낮음 (0.35 ~ 0.65)</td><td>적절함 (0.65 ~ 1.35)</td><td>높음 (1.35 ~ 1.70)</td><td>매우 높음 (1.70 이상)</td></tr><tr><td>곤란도</td><td>-.788</td><td>매우 쉬움 (-2.0 이하)</td><td>쉬움 (~ -0.5)</td><td>중간 (~0.5)</td><td>어려움 (~2.0)</td><td>매우 어려움 (2.0 이상)</td></tr></table>							변별도	2.789	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)	곤란도	-.788	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)
	변별도	2.789	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)														
	곤란도	-.788	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)														
문항 반응	<table><tr><td>①</td><td>②</td><td>③</td><td>④</td></tr><tr><td>5.0%</td><td>72.5%</td><td>14.2%</td><td>8.3%</td></tr></table>						①	②	③	④	5.0%	72.5%	14.2%	8.3%							
①	②	③	④																		
5.0%	72.5%	14.2%	8.3%																		

2) 듣기·말하기 영역

2022 국민의 국어능력 실태 조사		1차 예비 조사		LS-1(나)	2차 예비 조사		LS-1(나)																																									
평가 문항																																																
자료 출처																																																
평가 요소		배려하며 말하기 / 일대일 대화(공적 상황)																																														
해설		<table><tr><th colspan="2" rowspan="2">평가 항목</th><th rowspan="2">채점 요소</th><th colspan="3">채점 척도(점수)</th><th rowspan="2">점수 비중</th></tr><tr><th>미흡</th><th>보통</th><th>우수</th></tr><tr><td rowspan="2">내용 구성</td><td>내용 선정</td><td colspan="2">- 문제에서 요구하는 응답을 한다. (사과) - 잘못된 내용에 대해 언급하고, 상대의 기분을 헤아려 말한다.</td><td>1</td><td>2</td><td>3</td><td>30%</td></tr><tr><td>조직</td><td colspan="2">- 응답의 흐름이 사과하는 데 적합하다. - 발화의 시작과 마무리가 자연스럽다.</td><td>1</td><td>2</td><td>3</td><td>30%</td></tr><tr><td rowspan="2">표현 및 전달</td><td>표현</td><td colspan="2">- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.</td><td>1</td><td>2</td><td>3</td><td>20%</td></tr><tr><td>전달</td><td colspan="2">- 발음, 속도, 어조 등이 자연스럽다(유창성). - 발음이 명료하다(정확성).</td><td>1</td><td>2</td><td>3</td><td>20%</td></tr></table>							평가 항목		채점 요소	채점 척도(점수)			점수 비중	미흡	보통	우수	내용 구성	내용 선정	- 문제에서 요구하는 응답을 한다. (사과) - 잘못된 내용에 대해 언급하고, 상대의 기분을 헤아려 말한다.		1	2	3	30%	조직	- 응답의 흐름이 사과하는 데 적합하다. - 발화의 시작과 마무리가 자연스럽다.		1	2	3	30%	표현 및 전달	표현	- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.		1	2	3	20%	전달	- 발음, 속도, 어조 등이 자연스럽다(유창성). - 발음이 명료하다(정확성).		1	2	3	20%
평가 항목		채점 요소	채점 척도(점수)			점수 비중																																										
			미흡	보통	우수																																											
내용 구성	내용 선정	- 문제에서 요구하는 응답을 한다. (사과) - 잘못된 내용에 대해 언급하고, 상대의 기분을 헤아려 말한다.		1	2	3	30%																																									
	조직	- 응답의 흐름이 사과하는 데 적합하다. - 발화의 시작과 마무리가 자연스럽다.		1	2	3	30%																																									
표현 및 전달	표현	- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.		1	2	3	20%																																									
	전달	- 발음, 속도, 어조 등이 자연스럽다(유창성). - 발음이 명료하다(정확성).		1	2	3	20%																																									
1차 예비 조사 결과		<table><tr><td>변별도</td><td>1.09</td><td>거의 없음 (0.00 ~ 0.35)</td><td>낮음 (0.35 ~ 0.65)</td><td>적절함 (0.65 ~ 1.35)</td><td>높음 (1.35 ~ 1.70)</td><td>매우 높음 (1.70 이상)</td></tr><tr><td>곤란도</td><td>.640</td><td>매우 쉬움 (-2.0 이하)</td><td>쉬움 (~ -0.5)</td><td>중간 (~0.5)</td><td>어려움 (~2.0)</td><td>매우 어려움 (2.0 이상)</td></tr></table>							변별도	1.09	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)	곤란도	.640	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)																										
변별도	1.09	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)																																										
곤란도	.640	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)																																										

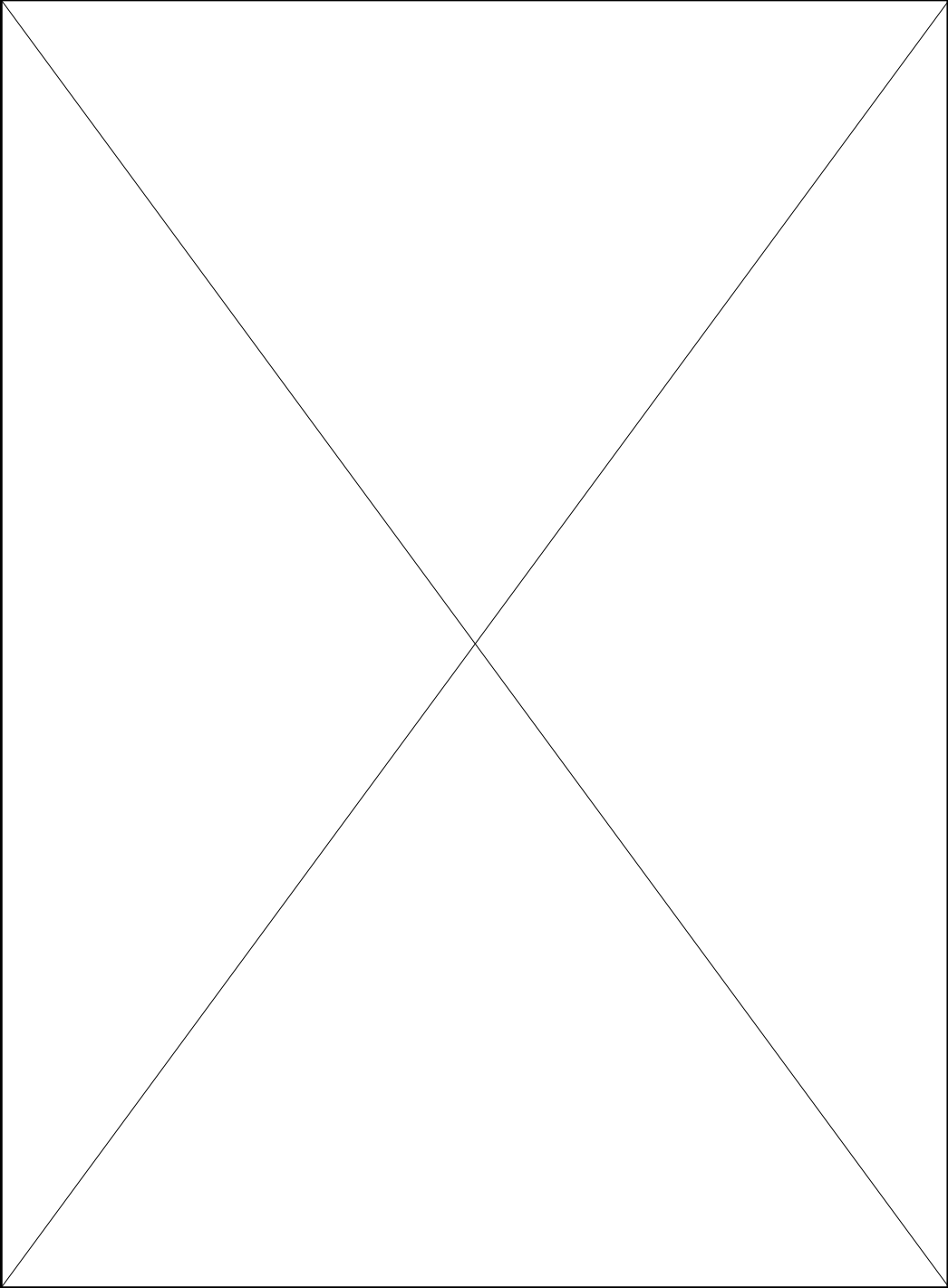
2차 예비 조사 결과	변별도	내용 선정	1.075	조직 영역에서 변별도가 매우 높은 수준으로 나타남. 내용 선정 영역에서 변별도가 적절한 수준으로 나타남. 표현, 전달 영역에서는 변별도가 낮은 수준으로 나타남.
		조직	2.638	
		표현	.603	
		전달	.580	
	곤란도	내용 선정	-.758	전 영역에서 곤란도는 쉬운 수준으로 나타남.
		조직	-.905	
		표현	-.883	
		전달	-.968	

2022 국민의 국어능력 실태 조사		1차 예비 조사		LS-2(가)	2차 예비 조사		LS-2(공통)																																									
평가 문항																																																
자료 출처																																																
평가 요소		상호교섭하기 / 일대일 대화(개인적 상황)																																														
해설		<table><tr><th colspan="2" rowspan="2">평가 항목</th><th rowspan="2">채점 요소</th><th colspan="3">채점 척도(점수)</th><th rowspan="2">점수 비중</th></tr><tr><th>미흡</th><th>보통</th><th>우수</th></tr><tr><td rowspan="2">내용 구성</td><td>내용 선정</td><td colspan="2">- 문제에 부합하는 적절한 응답을 한다. - 문제에 대한 설명과 해결을 위한 반응이 잘 드러난다(적극성).</td><td>1</td><td>2</td><td>3</td><td>30%</td></tr><tr><td>조직</td><td colspan="2">- 응답이 이전 발화의 흐름에 잘 대응된다. - 발화의 시작과 마무리가 자연스럽다.</td><td>1</td><td>2</td><td>3</td><td>30%</td></tr><tr><td rowspan="2">표현 및 전달</td><td>표현</td><td colspan="2">- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.</td><td>1</td><td>2</td><td>3</td><td>20%</td></tr><tr><td>전달</td><td colspan="2">- 발음, 속도, 어조 등이 자연스럽다(유창성). - 발음이 명료하다(정확성).</td><td>1</td><td>2</td><td>3</td><td>20%</td></tr></table>							평가 항목		채점 요소	채점 척도(점수)			점수 비중	미흡	보통	우수	내용 구성	내용 선정	- 문제에 부합하는 적절한 응답을 한다. - 문제에 대한 설명과 해결을 위한 반응이 잘 드러난다(적극성).		1	2	3	30%	조직	- 응답이 이전 발화의 흐름에 잘 대응된다. - 발화의 시작과 마무리가 자연스럽다.		1	2	3	30%	표현 및 전달	표현	- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.		1	2	3	20%	전달	- 발음, 속도, 어조 등이 자연스럽다(유창성). - 발음이 명료하다(정확성).		1	2	3	20%
평가 항목		채점 요소	채점 척도(점수)			점수 비중																																										
			미흡	보통	우수																																											
내용 구성	내용 선정	- 문제에 부합하는 적절한 응답을 한다. - 문제에 대한 설명과 해결을 위한 반응이 잘 드러난다(적극성).		1	2	3	30%																																									
	조직	- 응답이 이전 발화의 흐름에 잘 대응된다. - 발화의 시작과 마무리가 자연스럽다.		1	2	3	30%																																									
표현 및 전달	표현	- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.		1	2	3	20%																																									
	전달	- 발음, 속도, 어조 등이 자연스럽다(유창성). - 발음이 명료하다(정확성).		1	2	3	20%																																									
1차 예비 조사 결과		<table><tr><td>변별도</td><td>1.01</td><td>거의 없음 (0.00 ~ 0.35)</td><td>낮음 (0.35 ~ 0.65)</td><td>적절함 (0.65 ~ 1.35)</td><td>높음 (1.35 ~ 1.70)</td><td>매우 높음 (1.70 이상)</td></tr><tr><td>곤란도</td><td>.370</td><td>매우 쉬움 (-2.0 이하)</td><td>쉬움 (~ -0.5)</td><td>중간 (~0.5)</td><td>어려움 (~2.0)</td><td>매우 어려움 (2.0 이상)</td></tr></table>							변별도	1.01	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)	곤란도	.370	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)																										
변별도	1.01	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)																																										
곤란도	.370	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)																																										

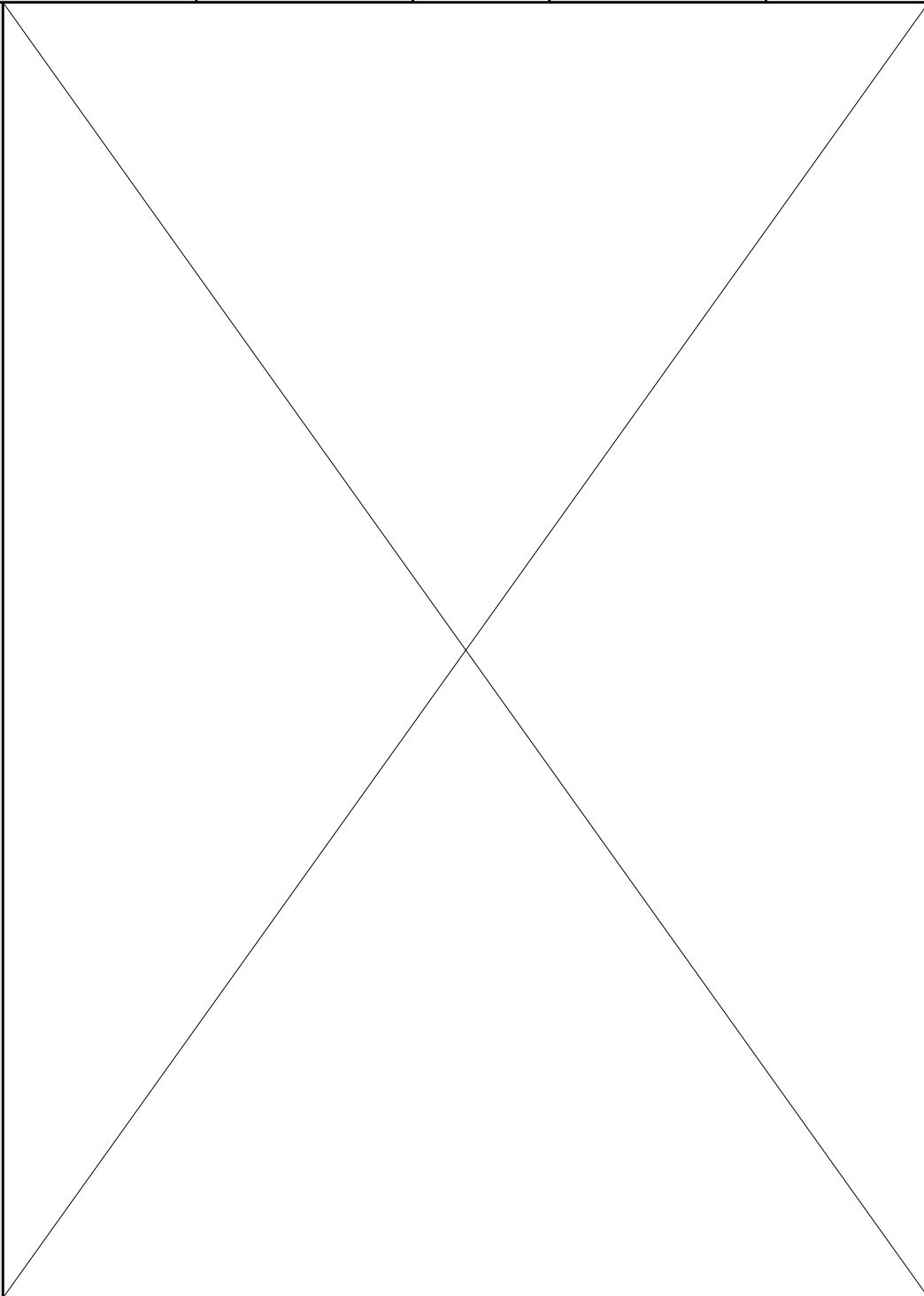
2차 예비 조사 결과	변별도	내용 선정	.471	내용 선정, 조직 영역에서 변별도가 낮은 수준으로 나타남. 표현과 전달 영역의 변별도는 거의 없는 것으로 나타남.
		조직	.446	
		표현	.150	
		전달	.054	
	곤란도	내용 선정	.949	내용 선정 영역에서 곤란도는 어려운 수준으로 나타남. 조직 영역에서 곤란도는 쉬운 수준을, 전달, 표현 영역에서 곤란도는 매우 쉬운 수준을 보임.
		조직	-1.366	
		표현	-7.913	
		전달	-11.296	

3) 말하기 영역

2022 국민의 국어능력 실태 조사			1차 예비 조사		S-1(공통)		2차 예비 조사		S-1(공통)		
평가 문항											
자료 출처											
평가 요소			설명적 말하기 - 경험 또는 개념을 설명하는 말 / 일대일 대화(개인적 상황)								
해설			평가 항목		채점 요소	채점 척도(점수)					점수 비중
						매우 미흡	미흡	보통	우수	매우 우수	
			내용 구성	내용 선정	- 질문에서 요구하는 대답을 한다. - 청자가 이해하기 쉽게 구체적으로 설명한다.	3	6	9	12	15	30%
				조직	- 내용의 흐름이 설명에 적합하다. - 시작과 마무리가 자연스럽다.	3	6	9	12	15	30%
			표현 및 전달	표현	- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.	2	4	6	8	10	20%
전달	- 발음, 속도, 어조가 적절하다. - 적절한 시간을 지켜 말한다.	2		4	6	8	10	20%			
1차 예비 조사 결과			변별도	.980	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)		
			곤란도	-.030	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)		
2차 예비 조사 결과			변별도	내용 선정	3.538	내용 선정 영역에서 변별도가 매우 높은 수준으로, 조직 영역에서 변별도가 높은 수준으로 나타남. 전달 영역에서 변별도가 적절한 수준으로 나타남. 표현 영역의 변별도는 낮은 수준으로 나타남.					
				조직	1.692						
				표현	.399						
				전달	.880						
			곤란도	내용 선정	.050	전달 영역에서 곤란도가 어려운 수준으로 나타남. 내용 선정에서 곤란도는 중간 수준, 조직, 표현, 전달 영역에서 곤란도는 다소 쉬운 수준을 보임.					
				조직	-.136						
				표현	-.366						
				전달	1.245						

2022 국민의 국어능력 실태 조사	1차 예비 조사	S-2(가)	2차 예비 조사	S-2(가)
평가 문항				
자료 출처				
평가 요소				
해설	설명적 말하기 - 프레젠테이션 / 발표(공적 상황)			

평가 항목		채점 요소	채점 척도(점수)					점수 비중
			매우 미흡	미흡	보통	우수	매우 우수	
내용 구성	내용 선정	- 슬라이드의 내용을 충분히 설명한다. - 복잡한 내용을 간명하게 정리하여 말한다.	1	2	3	4	5	30%
	조직	- 발표의 흐름이 자연스럽다. - 발화의 시작과 마무리가 효과적이다.	1	2	3	4	5	30%
표현 및 전달	표현	- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.	1	2	3	4	5	20%
	전달	- 발음, 속도, 어조 등이 자연스럽다(유창성). - 발음이 명료하다(정확성).	1	2	3	4	5	20%
1차 예비 조사 결과								
		변별도	1.160	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)
		곤란도	-.250	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)
2차 예비 조사 결과		변별도	내용 선정	1.096	조직 영역에서 변별도가 매우 높은 수준으로, 표현 영역에서 변별도가 높은 수준으로 나타남. 내용 선정, 전달 영역에서는 변별도가 적절한 수준을 보임.			
			조직	2.856				
			표현	1.401				
			전달	1.106				
		곤란도	내용 선정	.257	전 영역에서 곤란도는 중간 수준을 보임.			
			조직	.279				
			표현	.085				
			전달	.162				

2022 국민의 국어능력 실태 조사	1차 예비 조사	S-2(나)	2차 예비 조사	S-2(나)
평가 문항				
자료 출처				
평가 요소	설명적 말하기 - 프레젠테이션 / 발표(공적 상황)			
해설				

평가 항목	평가 항목		채점 요소	채점 척도(점수)					점수 비중
				매우 미흡	미흡	보통	우수	매우 우수	
	내용 구성	내용 선정	- 슬라이드의 내용을 충분히 설명한다. - 복잡한 내용을 간명하게 정리하여 말한다.	1	2	3	4	5	30%
조직		- 발표의 흐름이 자연스럽다. - 발화의 시작과 마무리가 효과적이다.	1	2	3	4	5	30%	
표현 및 전달	표현	- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.	1	2	3	4	5	20%	
	전달	- 발음, 속도, 어조 등이 자연스럽다(유창성). - 발음이 명료하다(정확성).	1	2	3	4	5	20%	

1차 예비 조사 결과	변별도	1.110	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)
	곤란도	.050	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)

2차 예비 조사 결과	변별도	내용 선정	2.025	내용 선정, 조직 영역에서 변별도가 매우 높은 수준으로 나타남. 표현, 전달 영역에서 변별도가 높은 수준을 보임.
		조직	3.831	
		표현	1.501	
		전달	1.629	
	곤란도	내용 선정	.465	전 영역에서 곤란도는 중간 수준을 보임.
		조직	.350	
		표현	.054	
		전달	.101	

2022 국민의 국어능력 실태 조사			1차 예비 조사		S-3(나)		2차 예비 조사		S-3(공통)			
평가 문항												
자료 출처												
평가 요소			설득적 말하기 - 주장하는 말 / 토론(공적 상황)									
해설			평가 항목		채점 요소		채점 척도(점수)					점수 비중
						매우 미흡	미흡	보통	우수	매우 우수		
			내용 구성	내용 선정	- 문제에서 요구하는 응답을 한다. - 문제 상황에 대한 설명과 주장의 근거가 잘 드러난다.	1	2	3	4	5	30%	
				조직	- 내용의 흐름이 논증에 적합하다. - 발화의 시작과 마무리가 유기적이다.	1	2	3	4	5	30%	
			표현 및 전달	표현	- 적절한 어휘를 사용한다. - 어법에 맞게 말한다.	1	2	3	4	5	20%	
				전달	- 발음, 속도, 어조 등이 자연스럽다. - 적절한 시간을 지켜 말한다.*	1	2	3	4	5	20%	
*10초 미만:매우 미흡, 30초:미흡, 30초 이상은 시간에 따른 감점 없음.												
1차 예비 조사 결과			변별도	.740	거의 없음 (0.00 ~ 0.35)	낮음 (0.35 ~ 0.65)	적절함 (0.65 ~ 1.35)	높음 (1.35 ~ 1.70)	매우 높음 (1.70 이상)			
			곤란도	.060	매우 쉬움 (-2.0 이하)	쉬움 (~ -0.5)	중간 (~0.5)	어려움 (~2.0)	매우 어려움 (2.0 이상)			

2차 예비 조사 결과	변별도	내용 선정	2.773	내용 선정, 조직 영역에서 변별도가 매우 높은 수준으로 나타남. 표현 영역에서 변별도가 높은 수준을, 전달 영역에서 변별도가 적절한 수준을 보임.
		조직	4.051	
		표현	1.560	
		전달	1.072	
	곤란도	내용 선정	.211	전달 영역에서 곤란도가 어려운 수준으로 나타남. 내용 선정, 조직, 표현 영역에서 곤란도는 중간 수준을 보임.
		조직	.340	
		표현	.162	
		전달	.582	

<기획·연구>

국립국어원 이기연 학예연구사

<연구 참여자>

연구책임자 구본관(서울대학교 국어교육과 교수)

공동연구원 민병곤(서울대학교 국어교육과 교수)

이관희(서울교육대학교 국어교육과 교수)

김승현(홍익대학교 국어교육과 교수)

박성석(춘천교육대학교 국어교육과 교수)

이성준(서울대학교 국어교육과 겸임 교수)

최선희(서울대학교 BK21 선임 연구원)

연구보조원 강지영(경상국립대학교 국어교육과 강사)

김예원(서울대학교 국어교육과 대학원 박사 과정)

이수진(서울대학교 국어교육연구소 연구실장)

나상수(서울대학교 국어교육과 조교)

조사 수행 KANTAR PUBLIC

발행인: 국립국어원장

발행처: 국립국어원

서울시 강서구 금남화로 154

전화 02-2669-9775, 전송 02-2669-9727

인쇄일: 2022년 12월 16일

발행일: 2022년 12월 16일

인 쇄: 가람문화사

※ 이 보고서는 국립국어원의 용역비로 수행한 ‘국민의 국어능력 실태 조사 개선 기초 연구’ 사업의 결과물을 발간한 것입니다.