

ISSN: 2800-051X

HankookResearch

한국리서치 주간리포트

여론속의 여輿論論

특별기획: 현 정당지지분포와 2024 총선 구도에 대한 이해
전국지표조사(NBS)와 갤럽조사의 차이는 왜 생길까?
- 정치 성향과 워딩 효과

2023. 11. 10.

담당자 정한울 전문위원
전화 | 02-3014-1057
e-mail | hw.jeong@hrc.co.kr

박정석 수석연구원
전화 | 02-3014-1017
e-mail | jspark@hrc.co.kr

[담담한 선거연구③] 현 정당지지분포와 2024 총선 구도에 대한 이해 전국지표조사(NBS)와 갤럽조사의 차이는 왜 생길까 : 정치 성향과 워딩 효과

2023.11.08

정한울(한국리서치 리서치 디자이너)

박정석(한국리서치 수석연구원)

1. 선거여론조사 방법 논쟁 1라운드: “전화면접조사” vs “ARS”

여론조사가 단순한 경주마 보도의 소재를 넘어 유권자의 후보 선택의 기준으로 활용되거나, 주요 정당의 공직자 후보 선출 제도의 기준이나 주요 국정과 정책 결정의 근거로 활용되는 등 그 활용범위가 확장되고 관련 보도도 증가하고 있다(구본상·박원호 2022; 정한울 2016).¹⁾ 이 과정에서 선거여론조사 방법에는 다음과 같은 변화들이 생겼다. 첫째, 2010년대로 접어들면서 가구전화 조사(유선)를 대신하여 휴대전화 조사(무선)가 보편화되었다. 둘째, 표본추출 방법도 가구조사가 의존해온 KT 등재번호부 기반의 할당표집(quota sampling)에서 임의번호추출(Random Digit Dialing) 시대를 거쳐 통신사에서 받은 휴대전화 가상/안심번호(virtual numbers)를 활용한 층화확률(stratified random sampling) 방법이 확산되고 있다. 셋째, 면접원이 조사를 진행하는 고가의 면접원 조사가 줄어드는 대신 자동응답조사(ARS) 비중이 급격하게 높아지면서 어떤 방법이 더 신뢰할만하고 정확한지 논란이 커져 왔다(중앙선거여론조사심의위원회 2022a, 48).²⁾

실제로 전화면접조사와 ARS 조사 간에는 국정 평가, 정당지지율, 선거 구도 및 투표 선호와 관련하여 ‘들쭉날쭉’하게 상반된 결과를 보여주면서 조사 방법에 따라 우열이 뒤바뀌는 사례가 반복되어 온 것이 사실이다. 정치권과 언론에서는 이러한 우열의 차이를 특정 조사 방법의 정치적 편향으로 ‘공식화’하는 경향이 나타난 것도 사실이다. 즉 최근 언론보도에서는 “전화면접=국민의힘에 유리, ARS=민주당에 유리”라는 식의 공식이 등장하기도 하고 각각 조사가 특정 정파의 이해관계를 반영하며 여론몰이하는 것이 아닌가 하는 의문까지 제기되고 있다. 이를 근거로 자신에게 유리한 조사 결과는 옹호하고, 반대로 정치적 이해관계에 부합하지 않는 조사에 대해서는 폄하하거나 선거여론조사에 대한 규제를 강화하려는 움직임으로 이어지고 있다.³⁾

① 쟁점1 : 조사방법의 정치적 편파성?

정권교체 후 공식 변화: 前 “전화면접은李, ARS는尹” → 後 “전화면접은국힘, ARS는민주”

그러나 정권교체 전후 공식이 완전히 뒤바뀌었다는 점에서 특정 조사와 특정 정파의 성향을 연결하는 것은 선부르다. [표1]을 보면 불과 1년 전 치러진 대선에서는 반대로 “전화면접=이재명/민주당, ARS=윤석열/국민의힘”이라는 정반대의 공식이 언론을 통해 유포된 바 있다. 즉 전화면접이 ‘보수 혹은 진보에 유리하다’, 반

- 1) MBC. “이재오 특임장관, 개헌 필요성 거듭 역설”(2015/05/12); 경향신문. “기초단체 무공천 파동 ‘민의’는 기반 당했다”(2014/04/12); 경기신문. “정부, 여론조사 활용 주요 정책 추진에 사회적 혼란 지적.”(2022/09/02); 오마이뉴스. “이번 대선 여론조사, 직전 대선 여론조사 횡수 추월했다” (2022/01/17); 주간조선. “이종섭 국방 “BTS 병역, 여론조사로 신속 결론”(2022/08/31); 조선일보. “진짜 민심 맞습니까… 가짜뉴스 같은 여론조사” (2023/05/10) 한겨레. “9가지 방식 여론조사 쏟아져 나와…전세계에 이런 경우 없어”(2022/03/16)
- 2) 조사 방법별 조사현황을 보면 2022년 대선 여론조사 총 774건 중 ARS 단일 조사가 56.6%로 반수를 넘었고 전화면접 단일조사가 28.8%, 전화면접과 ARS의 혼합조사가 7.2%, 기타(인터넷, 앱) 조사가 7.4%로 나타났다. 3달 후 진행된 지방선거에서는 총 3,588건 중 ARS 단일조사가 무려 94.7%를 차지했고, 전화면접 단일조사가 2.0%, 혼합조사가 1.6%, 기타 조사가 1.8%에 그칠 정도로 저가의 ARS 조사에 의존하는 비중이 급격히 커지고 있다(중앙선거여론조사심의위원회. 2022a, 43-44). 립가중의 문제점에 대한 간략한 소개는 구본상·박원호(2022, 471-72) 참조.
- 3) 뉴시스 “輿, ‘불량 여론조사’ 뿌리 뽑아…여론조사 법 정비 추진” (2022/10/22); 동아일보 “넌뛰기 여론조사, ‘심판의 날’ 온다” (2022/02/22); 중앙일보 “‘못 믿을 여론조사, 차라리 발표 금지’ 꼭 따져봐야 할 이 숫자” (2022/10.28); 한겨레 “깜깜이 선거 속 들쭉날쭉 여론조사, 유권자는 혼란스럽다”(2020/04/11)

대로 'ARS 조사가 진보 혹은 보수에 유리하다'라는 식의 공식은 특정 시점에 따라 상이하게 나타날 수 있는 현상일 수 있다는 점에서 성급한 일반화라 볼 수 있다. 분명한 것은 전화면접과 ARS 조사 결과 사이에 뚜렷한 응답 차이가 체계적으로 나타난다는 점이며, 동시에 다른 결과가 공표되면서 유권자들과 이해관계가 걸린 정치권에 혼선을 빚고 있다는 점이다. 문제는 각 정파에서 자신에게 유리한 조사 결과만을 선택적으로 수용하면서 자신에게 유리한 방법을 옹호하고, 불리한 결과가 집중되는 방법에 대해서는 신뢰할 수 없다는 태도를 견지하면서 쉽게 해소되지 못하고 있다(박종희 2013; 구분상·박원호 2022).⁴⁾

전화면접조사를 옹호하는 쪽에서는 전화면접조사의 양호한 품질지표(응답률, 가중배율)를 근거로 방법론적 신뢰성과 함께 ARS 조사는 극단적 입장을 가진 강경한 유권자층이 집중된다고 비판하는 입장이다. 반면, ARS 조사의 옹호론은 실제 투표는 태도가 분명한 고관여층이 주도하고, 저관여층은 투표에 참여하지 않기 때문에 선거 예측의 정확성에서는 ARS가 우월하다는 입장으로 맞서는 경향을 보여왔다. 전자가 조사방법의 신뢰성과 대표성 논쟁으로 이어진다면, 후자는 선거결과의 예측력을 중심으로 한 정확성(accuracy) 논쟁으로 변했다.

[표1] 대선 전후 1년 만에 뒤바뀐 조사편향 공식

| 대선 이전 공식 | 대선 이후 공식 |
|--|--|
| “전화면접은 이재명, ARS는 윤석열에 유리” | “전화면접은 국민의힘, ARS는 민주당에 유리” |
| <ul style="list-style-type: none"> ▪ “들쭉날쭉 여론조사…전화면접엔 이재명·ARS 윤석열 이유는” 뉴스1(2021/06/05) ▪ “전화면접은 이재명, ARS는 윤석열...이유 있는 여론조사” 시사저널(2021/09/04) ▪ “지지율·당선 가능성 들쭉날쭉…ARS는 尹·전화면접은 李 유리” 매일경제(2022/02/28) 외 | <ul style="list-style-type: none"> ▪ “29%? 44%?...민주당 지지율 '여론조사 미스터리' 한국경제 (2023/07/31) ▪ “어떤 여론조사 믿어야 하나... 면접방식은 국민의힘, 자동응답방식은 민주당” 내일신문 (2023/08/02) ▪ “'사이 민주당' 얼마나 될까...전화면접은 국힘, ARS는 민주 우위” 중앙일보(2023/09/20) 외 |

② 쟁점2 : 품질지표(응답률과 가중배율)로 본 신뢰성과 대표성 논쟁

사실, 조사의 품질과 방법론적 신뢰성의 경우 중앙선거여론조사심의위원회(이하 '여심위')에서 발간해온 선거여론조사 백서들이나 등록된 선거여론조사 결과들에 대한 학계의 연구결과들을 보면 여론조사 품질 지표(높은 응답률, 낮은 가중배율, 림가중 아닌 셀 가중방식의 적용)에서 전화면접조사가 ARS 조사 대비 우수한 질을 보인다는 것은 논란의 여지가 없다(구분상·박원호 2022; 조진만 외 2021; 박인호 외 2019; 중앙선거여론조사심의위원회 2022a, 2020, 2018, 2017, 2016).

그러나 응답률의 차이가 조사 품질 혹은 조사 결과의 정확성/편향성에 미치는 영향에 대해서는 논쟁이 존재한다. 조사 결과와의 차이를 기준으로 보면 높은 응답률이나 낮은 가중배율 조사가 더 정확한(편향성 지표 A) 데이터를 제공하고, 반대로 낮은 응답률, 높은 가중배율 조사가 편향이 크다고 단정할 수 없다는 입장도 있다(이소영 외 2017; AAPOR 2023; Groves and Peytcheva 2008). 응답률 그 자체가 정확성과 신뢰성을 직접적으로 보장하는 것은 아니지만, 낮은 응답률과 높은 가중배율을 보여주는 ARS 조사에서 젊은 층과 여성 비율 등이 충분히 대표되지 않고 투표 미결정자/부동층이 표본에서 배제됨으로써 각 조사의 응답자 간 성향 차이로 이어진다는 반박도 제기된다. 강성 태도 층의 여론이 과대 대표되고, 결과적으로 조사 방법 간 응답 편향이 발생하는 요인이 된다는 것이다(고길곤·김대중 2018; 구분상·박원호 2022; 최중호 외 2022).

4) 관련 보도로는 JTBC “이재명 “ARS 조사 못 믿어”...JTBC 전화면접도 '박빙'” (2022/05/24); YTN. “이준석·김기현 “전례 없는 여론조사 방식은 혼란...분열 우려” (2021/10.26); 동아일보. “윤호중 “부정확한 여론조사, 국민 혼란스럽게 해”(2022/05/25); 서울경제. “여론조사 의미없다” 尹...김종민 “대중 영합 없인 국정 동력 찾기 어려워” (2022/07/06); 오마이뉴스. “‘별로 의미 없다’더니... 대통령실 “여론조사는 민심 지표” (2022/08/05); 오마이뉴스 “홍준표 “배현진, 전희경과 쌍벽... 여론조사 못 믿어”(2018/05/13); 조선일보 “이재명 “여론조사 다 틀립니다” 송영길 “여러분 여론조사 안 믿죠?”(2022/05/27);

[표2] 2022년도 양대 선거에서 공표된 선거여론조사의 응답률을 보면 평균적으로 전화면접조사가 ARS에 비해, RDD 조사 대비 가상번호를 사용한 조사일수록 일관되게 높은 응답률을 보여주고 있다. 현재 한국 언론에서 보도되는 협조율 기준 응답률 평균(통화 후 조사 완료율)이 지난 대선에선 전화면접 가상번호 21.9% vs ARS 가상번호 8.8%로 뚜렷한 차이를 보여준다. 지방선거에서도 전화면접 가상번호 19.5% vs ARS 가상번호 10.4%로 협조율 기준 응답률에서 전화면접 가상번호 조사의 품질이 높다는 것을 알 수 있다. 다만 ARS 조사가 협조율은 낮은 대신 접촉률(발신한 적격번호 중 통화가 이루어진 비율)이 높아 협조율과 접촉률의 곱으로 구한 미국여론조사협회(AAPOR)의 응답률(RR) 기준에서는 그 격차가 상대적으로 작다. 또한 지난 2022년 대선 이후 지방선거에서 양 방법 간 응답률 격차는 크지만 ARS의 응답률 상승으로 격차는 줄고 있다.

[표2] 2022년도 제20대 대선 및 제8회 지방선거 여론조사 방법/표본추출 별 평균 응답률 비교⁵⁾⁶⁾

| (1) 제20대 대선 | | 전화면접 | | | | ARS | | | |
|--------------|--------------|------|--------|--------|------|------|--------|--------|-------|
| 전체 | | 가상번호 | 무선 RDD | 유선 RDD | 소계 | 가상번호 | 무선 RDD | 유선 RDD | 소계 |
| 건 | 2,353 | 323 | 297 | 262 | 882 | 382 | 537 | 261 | 1,424 |
| 응답률 | 협조율(%) | 21.9 | 18.1 | 11.1 | 17.4 | 8.8 | 5.3 | 3.4 | 5.6 |
| | 접촉률(%) | 28.3 | 21.8 | 30.5 | 26.8 | 47.8 | 27.6 | 41.6 | 38.0 |
| | RR(AAPOR, %) | 6.1 | 3.7 | 3.3 | 4.5 | 3.8 | 1.5 | 1.4 | 2.1 |
| (2) 제8회 지방선거 | | 전화면접 | | | | ARS | | | |
| 전체 | | 가상번호 | 무선 RDD | 유선 RDD | 소계 | 가상번호 | 무선 RDD | 유선 RDD | 소계 |
| 건 | 2,966 | 396 | 15 | 128 | 539 | 1426 | 65 | 929 | 2,420 |
| 응답률 | 협조율(%) | 19.5 | 12.7 | 12.2 | 17.7 | 10.4 | 3.6 | 4.0 | 7.7 |
| | 접촉률(%) | 26.2 | 30.7 | 24.6 | 26.6 | 40.4 | 22.2 | 40.0 | 40.3 |
| | RR(AAPOR, %) | 5.1 | 3.9 | 3.0 | 4.7 | 4.2 | 0.8 | 1.6 | 3.1 |

출처: 중앙선거여론조사심의위원회. 2022. 「2022년 양대선거 선거여론조사 백서」를 필자가 재정리

[표3] 가중배율을 봐도 ‘셀 가중’을 주로 사용하는 전화면접조사의 가중배율 값이 ‘1’에 근접할 뿐 아니라 1로부터 편차가 가장 낮다. 현재 한국의 선거여론조사는 성*연령*지역별 대표성을 확보하기 위해 모집단의 층(strata)별 인구구성비에 비례하게 표본 할당 목표를 설정하고, 과대/과소 표집 된 층은 사후층화가중(post-stratification weight)으로 조정한다. 층별 목표 할당 수에 정확히 부합하면 가중값은 1(이상적 값)이다. 1보다 크면 목표 할당 수를 채우지 못해 배가시켰음을 의미하고, 1보다 작은 가중치는 과다 표집되어 축소했음을 의미한다. 물론 “가중값이 1에 수렴한다고 해서 반드시 표본의 대표성을 확보한다고 할 수는 없다.”(조진만 외 2021), 다만 가중값이 1을 초과하면 할수록 “추정량과 분산의 편향에 영향을 줄 수 있으며 오히려 오차를 증가”시킬 수 있다. 가중배율 문제는 특히 여성, 20대에 집중되고, ARS 조사, 유선 조사에서 취약한 것으로 나타났다(Groves et al. 2004; 구분상·박원호 2022).⁷⁾

- 5) 전체 사례 수는 혼합조사의 경우 1개의 조사에 사용된 2개 이상의 조사 방법을 각각 산정한 수치. 전체 대선 조사 건수는 2,353건은 표에 포함된 전화면접조사(881건) 및 ARS 방법(1,424건) 외 기타(패널 등) 47건의 합이며, 전체 지방선거 여론조사 건수 2,966건은 전화면접조사(539건), ARS 조사(2,420건)와 기타 7건을 합한 값이다.
- 6) 여심위가 발간한 백서에 따르면 대선 여론조사 전화면접조사-유선 조사 건수는 응답률(협조율), 접촉률의 경우 262건이나 AAPOR 기준 응답률(RR)의 경우 261건으로 차이가 있다. 따라서 응답률, 접촉률 총건수는 882건, RR 총건수는 881건이다. 지방선거 통계에서도 전화면접조사 건수가 조사 방법별 응답률(협조율) 등록현황표에서는 540건(유선 128건, 무선 412건)이지만, 표본추출별 응답률(협조율) 등록현황표에서는 539건(유선 RDD 128건, 무선 RDD 15건, 가상번호 396건)으로 차이가 난다(중앙선거여론조사심의위원회 2022, 70p).
- 7) 여심위는 <선거여론조사 기준>을 통해 성, 연령, 지역별로 0.7~1.5를 벗어난 조사는 공표를 금지하고 있다(2014년 제정, 제5조 2017년 3월 0.5~2.0 →2019년 7월 개정 0.7~1.5, 여심위 2022b). 2019년 기준 강화 전엔 30대도 가중배율 문제가 컸지만, 현재는 부분적으로 해소되는 추세이나, 20대와 여성 가중배율 문제는 여전히 문제가 심각하다(조진만 외 2021).

[표3] 제20대 대선 및 제8회 지방선거 여론조사 방법별 층별 평균 가중배율 비교

| | 제20대 대선 | | | | 제8회 지선 | | | |
|-------|---------|-------|-------|-------|--------|-------|-------|-------|
| | 소계 | 전화면접 | ARS | 혼합조사 | 소계 | 전화면접 | ARS | 혼합조사 |
| 건수 | 1384 | 422 | 702 | 222 | 1895 | 373 | 1473 | 44 |
| 남 | 0.90 | 0.97 | 0.88 | 0.80 | 0.94 | 0.97 | 0.93 | 0.94 |
| 여 | 1.13 | 1.02 | 1.15 | 1.30 | 1.08 | 1.02 | 1.10 | 1.06 |
| 20대 | 1.19 | 1.08 | 1.21 | 1.33 | 1.18 | 1.08 | 1.21 | 1.12 |
| 30대 | 1.08 | 1.05 | 1.06 | 1.22 | 1.10 | 1.09 | 1.10 | 1.09 |
| 40대 | 0.99 | 1.01 | 0.98 | 1.00 | 1.00 | 1.04 | 0.99 | 0.98 |
| 50대 | 0.90 | 0.96 | 0.89 | 0.81 | 0.90 | 0.95 | 0.89 | 0.91 |
| 60대+ | 0.95 | 0.94 | 0.95 | 0.95 | 0.95 | 0.94 | 0.95 | 1.00 |
| 지역 최소 | 0.87 | 1.04 | 1.15 | 1.19 | 1.11 | 1.04 | 1.13 | 1.11 |
| 지역 최대 | 1.12 | 0.94 | 0.86 | 0.78 | 0.90 | 0.96 | 0.89 | 0.86 |
| 단순 평균 | 1.01 | 1.00 | 1.01 | 1.04 | 1.02 | 1.01 | 1.02 | 1.01 |
| 표준편차 | 0.118 | 0.051 | 0.132 | 0.222 | 0.103 | 0.057 | 0.117 | 0.093 |

출처: 중앙선거여론조사심의위원회. 2022. 「2022년 양대선거 선거여론조사 백서」를 필자가 재정리

③ 쟁점3 : 꺼지지 않는 논란, 정확성 논쟁 ‘고 관여층 조사가 더 정확할까?’

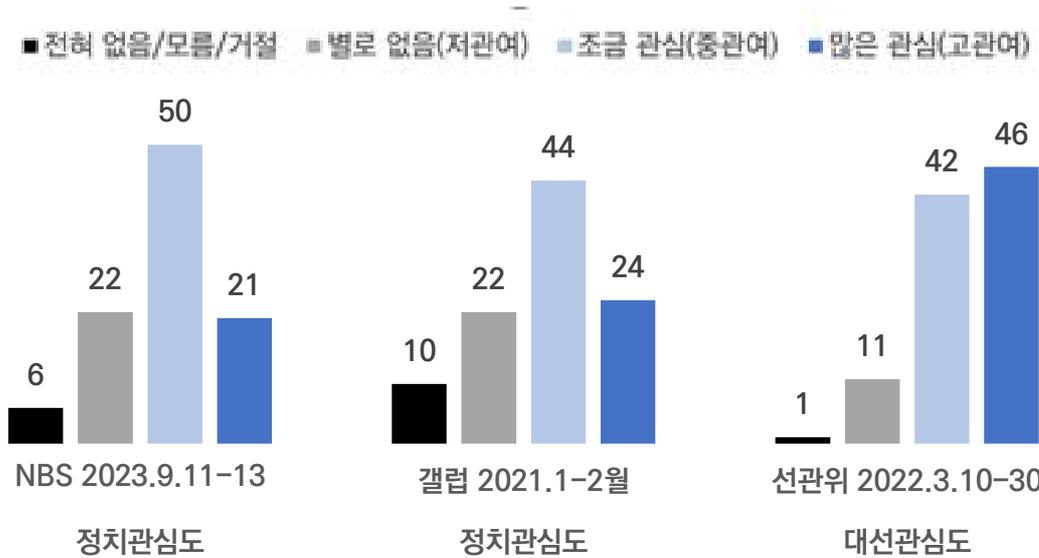
품질 지표에 대한 논란은 사실상 일단락되는 분위기지만, 정확성 논란은 잦아들지 않고 있다. ARS 조사를 옹호의 입장에서는 정치적 저 관심(여)층은 투표장에 가지 않고, 고 관심(여)층이 투표장에 가기 때문에 고 관심층 조사에 가까운 ARS 조사가 보다 정확하게 실제 투표를 예측한다고 주장하기도 한다. 이는 실증연구를 통해 지속적으로 검증됐다. 한국통계학회의 2019년 실험 연구(박인호 외 2019)를 비롯하여 여러 실증연구를 통해 응답률이 높은 조사(주로 전화면접조사)는 고 관여층에 비해 정치적 성향이 약하고 유동적인 중도층이나 무당파, 태도 변동 층(미결정, 스윙보터 등) 비중이 높은 반면, 반대로 응답률이 낮은 조사는 극단적 태도 층이 과대 대표되고 중/저 관여층이 과소 대표되는 경향이 반복적으로 확인됐다. 그런데도 실제 선거 여론을 주도하는 것은 고 관여층이며 저 관여층은 실제 투표에 참여하지 않고, 자신의 뚜렷한 선호 없이 정치권과 언론의 동원에 의해 좌우되는 집단이기 때문에 고 관여층 여론을 대표하는 ARS가 선거 예측력에서는 더 우월하다는 주장도 제기되고 있다(구본상 2017; 최종호 외 2022).

고 관여층 단독으로 선거를 결정할 수 없다

첫째, 고 관여층은 전체 여론을 단독으로 좌우할 정도의 크기가 아니다. 고 관여층의 크기는 또한 선거 시기 여부나 정치적 상황에 따라 상당히 유동적이다. 우선, 고 관여층의 비중을 보면 선거 시기와 비 선거 시기에 차이가 있다. 비 선거 시기인 최근의 NBS 조사 9월 2주 정치관심도 조사 결과를 보면 관심이 ‘매우 많다’ 21%, ‘약간 있다’ 51%, ‘별로 없다’ 22%, ‘전혀 없다/모름’이 7%다. 한국갤럽의 2021년도 1-2월 통합자료에서도 ‘매우 관심이 있다’는 24%, ‘약간 있다’는 44%, ‘별로 없다’는 22%, ‘전혀 없다/모름/응답 거절’이 10% 수준이다.⁸⁾ 2022년 3월 대선 직후 중앙선거관리위원회의 선거 직후 ‘유권자 의식조사’ 결과를 보면 0.73%p 박빙의 대결을 펼쳤던 상황이었음에도 ‘매우 관심이 많다’는 과반에 못 미치는 46%, ‘약간 있다’는 42%, ‘별로 없다’는 11%, ‘전혀 없다’는 1%였다. 고 관심층 여론만으로 전체 여론의 판세를 판단하기에는 고 관심/고 관여층의 크기가 크지 않다(그림1).

8) 정치 관여도를 ‘관심도’를 통해 측정할 경우 보통 4점 척도(1. 매우 관심 2. 약간 관심 3. 별로 관심 없다 4. 전혀 관심없다)를 사용한다. 보통 “매우 관심이 많다”는 응답을 ‘고 관심/고 관여층’, ‘약간 관심이 있다’를 ‘약 관심층/중 관여층’, ‘별로 없다’를 ‘저 관심/저 관여층’ 및 ‘전혀 없다’는 응답을 ‘비 관심층/비 관여층’으로 해석할 수 있다.

[그림1] 2023년 9월 2주 정치관심도와 2022년 대선 시기 선거관심도(%)



자료: 정치관심도(NBS, 2023년 9월 2주), 대선 관심도(중앙선거관리위원회 제22대 대선 유권자 의식조사 2022년 3월),

중 관여층(약 관심층)에 대한 오해: (1) 투표 참여율과 (2) 고 관여층과 정치 성향의 차이

관심도를 기준으로 분석 시 '1. 매우 관심있다'는 '고 관여층', 3. 별로 없다'를 '저 관여층', '4. 전혀 없다'를 '비 관여층'으로 분류하는 것은 큰 문제가 없다. 문제는 '2. 약간/대체로/조금 관심있다'를 분류 시 조심할 필요가 있다. 그동안 때로는 고 관여층과 함께 정치 관심이 있는 '관심층(1+2번)' vs '비 관심층(3+4번)'으로 묶어 비교하기도 하고 갤럽의 정치관심도 분석처럼 이들을 '약 관심층'으로 분류하여 독립적으로 분석하기도 하지만 대체로 정치참여가 약한 '저 관여층의 일부(2+3+4번)'으로 분류하는 경향이 있다(한국갤럽 2021).

본 보고서는 '2. 약간/대체로/조금 관심있다' 층을 다른 집단과 묶어 분석하기 보다는 독립적으로 분석해야 하며 특히 저 관여층으로 통합하는 것은 이들 집단의 상당한 참여 성향을 왜곡할 수 있다고 본다. '저 관여층'은 '기권을 많이 하고 독자적 선호 없이 동원되는 무태도층'으로 이해됐는데 '2. 약간/대체로/조금 관심있다'는 층은 적극적으로 투표에 참여하고 독자적인 정치적 성향을 가지고 있다는 점에서 '저 관여층'과 구별되는 특성이 분명하다. 그런 의미에서 본 보고서는 2번 응답층을 고 관여층, 저 관여층과 구별하기 위해 잠정적으로 '중 관여층(middle 혹은 moderate)'으로 분류하고자 한다.

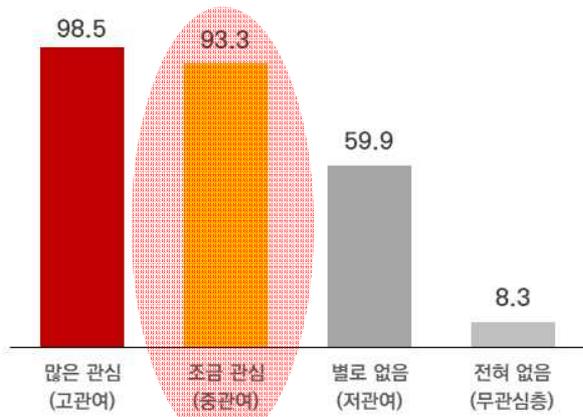
실제로 중앙선거관리위원회의 유권자 조사를 보면 2022년 대선에서의 선거 관심 유형별 투표 참여 여부 질문에 고 관여층('매우 관심이 많다'라는 투표율 98.5%) 못지않게 중 관여층('약간 관심이 있다'=갤럽기준 약 관심층의 대선투표율)은 93.3%로 높은 투표 참여율을 보였다. '별로 관심이 없다'고 답한 저 관여층의 투표율은 59.9%로 상대적으로 낮았고, '전혀 관심이 없다'고 답한 비 관심층에서는 8.3%만이 투표를 했다고 답했다. 기권 성향은 '4. 비 관심층'과 '3. 저 관여층 중 일부'에 해당하는 고정관념인 셈이다(그림2).

중 관여층(약 관심층)은 이념 분포에서도 고 관여층 혹은 저 관여층과 구별된다. NBS의 최근 조사 결과(2023년 9월 2주 조사)를 봐도 '매우 관심(고 관여)'층은 중도층이 적고(22%), 보수(40%)와 진보(32%) 등 이념 성향이 다수를 점하는 '양봉형 분포(polarization)'를 보이는 반면, '중 관여층(약 관심층)'은 중도 정체성 집단이 가장 많고(38%), 진보(27%)-보수(28%)성향이 균형을 이루는 정규분포를 보인다. '저 관여층(비 관심층)'은 중도/보수성향과 함께 응답을 유보한 무성향층이 적지 않다(그림3). 한국갤럽의 2021년 정치관심도 분

9) 대선 관심도 조사는 중앙선거관리위원회 「제20대 대통령선거 유권자 의식조사」 제3차 조사 보고서(한국갤럽, 1,512명, TAPI, 3월 10~30일). 보통 선거/정치 관심도 문항은 '사회적으로 바람직한 답변'을 하는 대표적인 문항으로서, 실제 투표율보다 선거전 '적극적 투표 의향'이나 사후 투표율 조사에 과대표되는 것이 사실이나, 일반 여론조사에 비해 중앙선거관리위원회에서 선거 전후 조사하는 「유권자 의식조사」에서는 이러한 격차가 상대적으로 크지 않아 투표 참여/참여 의향 분석에서는 가장 신뢰할만한 자료로 볼 수 있다.

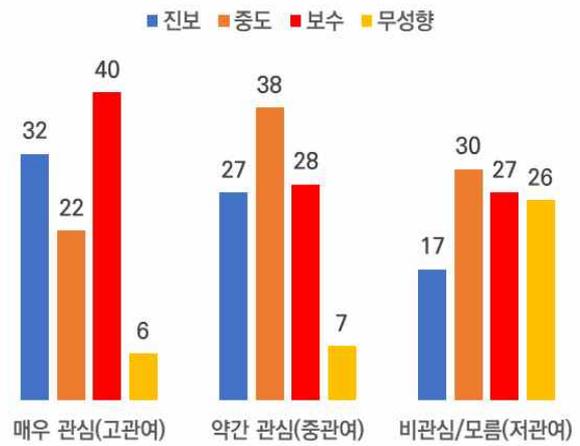
석에 따르면 2번 '약 관심층'은 정치적 선호에서 '고 관심/고 관여층'이나 '저 관여/비 관여층'과 뚜렷한 선호 차이를 갖고 있다(한국갤럽 2021). 1번 고 관심층(기준 변수) 대비 '약 관심층'과 '저 관심/비 관심층'의 양당 지지 성향을 로지스틱 회귀분석으로 분석한 결과 중 관여층(약 관심층)은 고 관심층 대비 국민의힘을 덜 지지하고, 더불어민주당을 더 선호하고, 비 관심층(저 관여+비 관여)에 비해서는 민주당/국민의힘에 대한 지지가 강하다는 독자적 정치 성향이 확인되었다.

[그림2] 선거 관심 유형별 투표율 차이(%)¹⁰⁾



자료: 중앙선거관리위원회 유권자의식조사 2022.3

[그림3] 정치 관심 유형별 이념 성향 차이(%)



자료: NBS 정기지표조사(2023년 9월2주 조사)

스윙보터(중/저 관여층)가 선거를 결정한다.

더구나 선거 경쟁의 과정과 결과를 좌우하는 태도 변동 층(스윙보터)나 당파적 태도가 상대적으로 약한 탈정파적 유권자층의 분석을 위해서라도 중 관여층(=약 관심층) 및 저 관여층의 대표성도 중요하다.

젤러의 RA 모델(Recept-Accept)은 태도 변화를 새로운 메시지에 대한 노출 확률(Reception)과 노출된 정보의 수용 확률(Acceptance)의 곱으로 설명한다. '고 관여층(high awarness)'은 새로운 메시지를 수신할 확률(0.9)은 높으나, 기존의 정치적 성향(political predispostion=이념성향/정당 당파성)이 강한 집단이기 때문에 기존 성향과 충돌하는 새로운 정보를 '수용(accept)'할 확률(0.1)이 낮아 태도변화로 이어질 가능성(0.09)이 낮아진다. '저관여층(low awarness)'의 경우 접수된 메시지에 대해서는 정치적 성향이 약해 수용할 확률(0.9)은 높지만, 새로운 정보 취득량이 부족(0.1)하여 태도변화 확률(0.09)은 고관여층처럼 낮다. 반면 '중 관여층(moderate awarness)'은 정보 취득 확률이 중간 수준(0.5), 새로운 메시지를 수용할 확률도 중간 수준(0.5)으로 곱하면 태도변화 확률이 0.25로 고관여, 저관여층보다 높다는 것이다(표3).

[표3] 젤러의 정보의 노출과 정치성향의 상호작용이 태도변화에 미치는 영향 (Zaller, 1992, p123)

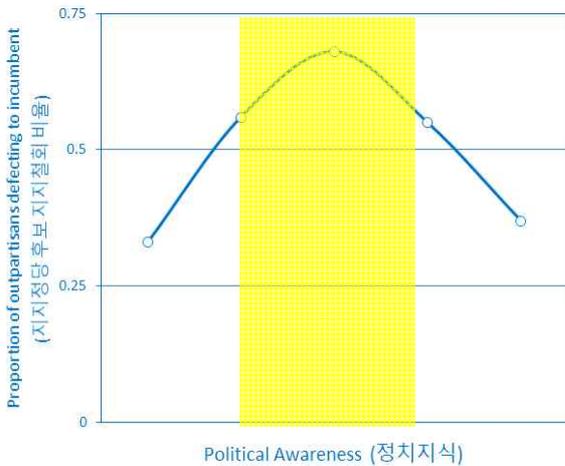
| | 정치적 관여도 (Level of awareness) | | |
|---|------------------------------|-----------|---------|
| | 저(Low) | 중(middle) | 고(high) |
| 확률(메시지 수신) Prob(Reception) | .10 | .50 | .90 |
| 확률(수용 수신) Prob(Accept Reception) | .90 | .50 | .10 |
| 태도 변화(수신×수용) Change(Reception×Acceptance) | .09 | .25 | .09 |

[그림4]는 미국 선거에서 중간 관여층에서 태도변화(유권자가 자신이 지지하던 정당의 후보가 아닌 상대 정당의 현직자 지지로 이탈하는 비율)가 높음을 보여주는 대표 사례이다(Zaller 1992, 20). 이는 한국의 선거

10) 보고서에는 관심도별 투표율 자료는 없고 투표유무별 관심도 <표1>에 제시되어, 이를 필자가 역산하여 관심도별 투표율을 계산한 수치다. 중앙선거관리위원회, 「제20대 대통령선거에 관한 유권자 의식조사」(2022. p318)

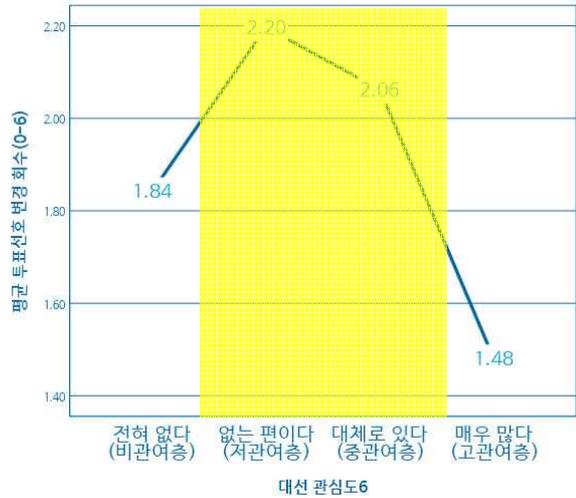
에서도 마찬가지다. 2012년 대선에서 6차례 진행한 패널조사에서 대선 지지후보를 교체한 횟수를 선거관여수 준별로 평균낸 결과 고관여층은 1.48회로 지지후보 교체경험이 가장 낮고, 그 다음이 선거에 전혀 관심이 없다는 비관여층에서 1.84회로 낮았다. 중관여층은 평균 2.06회, 관심이 별로 없다고 답한 저관여층에서 지지후보 교체회수는 2.20회로 가장 높게 나타나 중관여/저관여층이 투표 선택이 변동하는 핵심 스윙보터층임이 확인된다. 이는 통계적으로 유의한 차이였다([그림5], ANOVA test $p < 0.01^{**}$).¹¹⁾

[그림4] 관여도와 비지지정당 후보로 이탈율(미국)



출처: Zaller (1992, p20, 1978 NES Data, 필자가 재 도식화)

[그림5] 선거관심 유형 별 지지후보 교체 회수(한국)



정한울(2016, EAI·한국리서치 2012 패널조사)

정확성 논쟁에서 유의할 점: 편향척도 A에 기반한 정확성 논쟁과 경주마 조사의 한계

2014년 여심위의 선거여론조사 등록 제도 시행 이후 역대 선거 과정에서 이미 조사 방법에 따라 조사 결과의 차이가 뚜렷하게 나타나면서 정확성(편향성) 논란은 꺼지지 않고 혼란은 커지고 있다. 정확성 논쟁은 실제 투표 결과와 각 기간의 여론조사 결과와의 차이를 통해 측정된 결과(소위 편향척도 A)를 중심으로 이루어져 왔다.

그러나 최종 투표 결과의 차이에 기반한 정확성 평가는 “선거전이나 선거운동 기간 전체가 아니라 선거 직전의 여론조사”, 특히 출구조사와 선거 당일(혹은 선거 임박한 시점)의 여론조사만을 대상으로 삼아야 하는 한계가 발생한다(박종희 2013, 2021). 나아가 선거가 임박할수록 여론조사와 실제 결과가 근접할 것이라는 가정 자체가 “지나치게 강한 가정”이라는 비판도 제기된다. 실제로 분석 기간을 조정하면 조사별 정확성에 관한 판단 자체가 근본적으로 바뀌기도 한다. 이는 (1) ‘선거여론조사(투표자+비투표자 전체 유권자 대상 조사) ≠ 예측조사(투표자 대상/투표자 추정)’ (2) D-6 공표금지 조항으로 인해 선거 임박한 시점의 투표일 임박 결정 층(late deciders)의 배제 (3) 선거 기간 동안 국면변동이 잦은 역동적인 선거 시 기존의 정확성 측정지표(편향척도)에만 의존해서는 안 된다는 주장이 설득력을 갖게 한다(구본상·박원호 2022).

조사 정확성이나 선거 예측력의 개선을 위해서는 전화면접조사와 ARS 방법 간의 차이에 대한 논쟁 외에 조사 ‘응답자’와 ‘비응답자’간 성향 차이에 대한 검증이 더욱 중요하다는 주장(장덕현 외 2014; 장덕현 2021)이나, 조사기관 혹은 의뢰기관의 편향성 문제도 주목할 주제이다(박종희 2021). 지난 대선 전후로 한국에서도 본격적으로 시도되기 시작한 선거여론조사 결과들에 대한 ‘메타분석 방법/선거조사 자료 결합 방법(poll aggregation)’도 향후 정확성 논쟁과 관련하여 주목할 만한 새로운 실험이다(MBC·박종희 교수연구팀 “여론 조사를 조사하다”, SBS·Underscore 팀 “메타·S : Poliscore”). 선거조사 결합 방법은 “다양한 방법으로 이

11) 이러한 비선형관계(nonlinear model)는 이후 다른 연구에서도 확인된다(Goidel et al. 1997)

루어진 조사에 담긴 정보(information)는 반영하면서 각 방법이 가진 다양한 편향(bias)은 상쇄되어” 추세 파악에 유용하다는 평가가 나오고 있다(구본상·박원호 2022 ; Deane et al 2019).¹²⁾

본 보고서는 ‘예측 정확성’ 중심으로 논쟁이 진행되면서 선거여론조사 관련 논의가 지나치게 ‘방법론’적 논의에 집중된 결과, 내용상으로는 ‘경주마 조사(horse race polls)’라는 전통적인 선거여론조사에 대한 비판에 대해서는 논의가 진전되지 못하고 있음을 강조하고 싶다. 선거여론조사는 ‘경주마 게임’의 도우미에 불과하고, 정작 선거 과정에서 객관적 민의 수렴(현직자 및 후보 캠페인 평가, 정책적 수요와 요구사항 파악 및 향후 국정 및 정책 방향 수립에 반영 등)을 통해 민주주의의 성숙에 기여한다는 본연의 기대와는 멀어지고 있다는 것이다(정한울 2019).

12) 물론 이에 대한 반론도 제기된다. 자료 결합으로 다양한 자료 수집 방법에 따라 여론의 변동 추이를 비교, 종합해볼 수 있다는 장점이 있지만 “모든 재료가 똑같이 영양가 있는지 알기 어렵고 사실상 “누가 앞서고 뒤서는지 소위 ‘경주마(horse race)’ 범주” 이상의 분석적 의미를 갖기 어렵다는 우려도 제기된다(Deane et al. 2019).

2. 조사 방법 논쟁 2라운드 : NBS-갤럽 조사 사이의 불일치

① 또 다른 혼란의 시작 : NBS-갤럽 결과의 차이

해프닝으로 끝날 뻔한 혼란

최근 한국 선거여론조사의 또 다른 혼란의 요인은 유사 시점의 동일 조사 방법 내에서도 다른 결과가 나타나고 있다는 점이다. 예전에도 종종 유사 시점의 동일 조사 방법의 조사 결과가 불일치하여 관심을 받은 적이 있었다. 특히 2020년 7월부터 NBS 조사가 발표한 이래 종종 같은 전화면접조사를 사용하는 한국갤럽의 정기조사 결과와 일치하지 않는 결과들이 종종 발표되면서 언론에 주목받기도 했다.¹³⁾

- 21년 8월 국정 평가 긍정 비율: NBS (12일) 43% vs 갤럽 (13일) 36%
- 21년 11월 대선 양자: NBS (18일) 李 35%, 尹 36% vs 갤럽(19일) 李 31%, 尹 42%

그러나 면접조사와 ARS 조사 방법, 같은 ARS 조사 방법 간의 결과 차이가 두드러졌기 때문에 상대적으로 같은 전화면접 조사 간 차이에 관한 관심은 집중되지 않았다. 더구나 2022년 대선 직전 마지막 D-6인 2월 28일-3월 2일에 실시한 양 기관의 최종 조사 결과(3일 발표)를 보면 국정 긍정 평가 비율 NBS(46%)/갤럽(45%), 양당 정당 지지 NBS(35%:35%), 갤럽 (38%:38%)로 동률을 기록했고, 대선 지지도에서 NBS(李 40%: 尹 40%: 安 9%)와 갤럽(李 38%: 尹 39%: 安 12%)이 한 기관의 조사라 보일 정도로 유사한 결과를 보여주면서 양 기관의 차이 논란은 해프닝으로 끝나는 듯 했다.

- 22년 3월 국정 평가 긍정 비율: NBS (3일) 46% vs 갤럽 (3일) 45%
정당 지지 : NBS(3일) 민주 35%, 국민 35% vs 갤럽(3일) 민주 38%, 국민의힘 38%
대선 지지 : NBS(3일) 이 40%, 윤 40%, 안 9% vs. 갤럽(3일) 이 38%, 윤 39%, 안 12%

윤 정부 국정 평가에서는 유의한 차이가 없다

더구나 윤석열 정부 시기(22년 5월 2주부터 2023년 10월 2주)에 한정하여 양 기관의 조사 결과를 보면 국정 평가에서는 응답 비율의 변화 추이나 응답 비율 간 유의미한 차이를 발견하기 어렵다. NBS 조사는 22년 5월 3주부터 23년 10월 2주까지 총 37차례에 걸쳐 4점 척도 (매우 잘함/대체로 잘함/대체로 못함/매우 못함-재질문 안 함)로 질문하여 긍정/부정 평가로 분류하고, 갤럽조사는 22년 5월 2주부터 23년 10월 2주까지 2점 척도(잘했다, 잘못했다)로 총 66차례 조사를 진행했다.¹⁴⁾

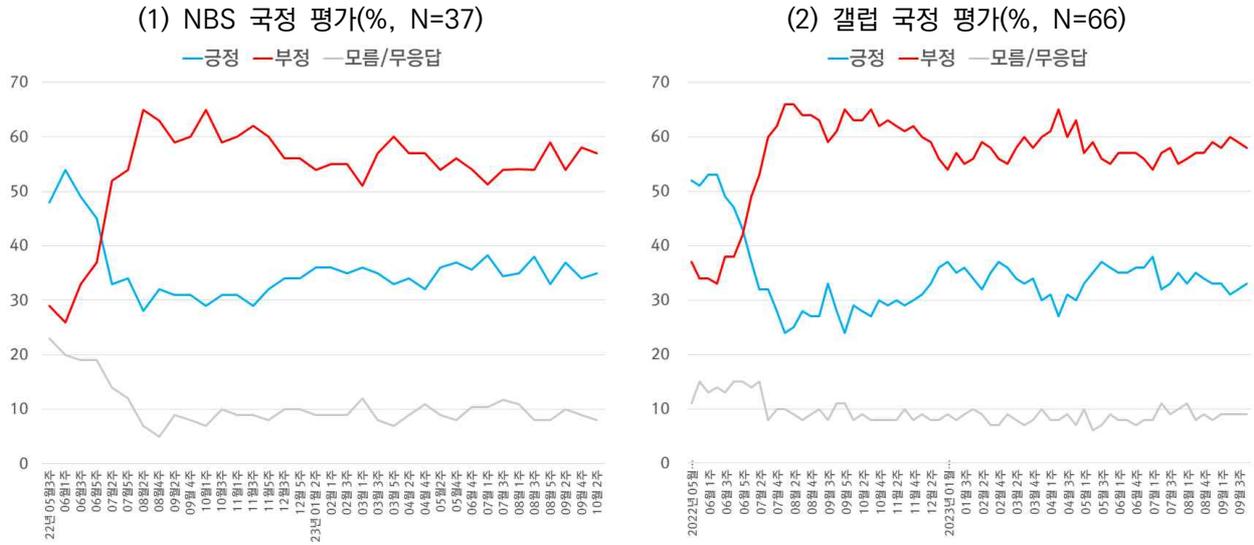
변화추이를 보면 양 조사 모두 윤석열 대통령 취임 후 지방선거 전까지 긍정 평가가 우세했지만, 22년 하반기 이후에는 양 기관의 조사에서 긍정 평가가 30%대, 부정 평가가 55~60%대에서 고착되는 양상을 보여주고 있다(그림6). NBS 조사는 격주 조사, 갤럽조사는 매주 조사인 점을 고려하면 양 조사 간 차이를 엄밀히 살펴보기 위해 동일 조사 시점의 결과만으로 비교해보았다. 양 기관의 긍·부정 평가 비율은 그 신뢰구간이 겹치는 것으로 볼 때 통계적으로 유의미한 차이가 없다. NBS 긍정 평가 평균은 35.8%, 갤럽은 34.0%(차이 +1.8%p), 부정 평가 평균은 NBS 53.5%, 갤럽 56.5%(차이 -3.0%p) 였다(그림7, 부록-표1).¹⁵⁾

13) JTBC. “[여론 읽어주는 기자] 자고 나면 널뛰는 ‘스카이콩콩’ 여론조사 왜?” (2021/11/24); 서울신문. “유권자는 헛갈린다... 들쭉날쭉 여론조사” (2022/01/12); 세계일보. “같은 문항 가지고도 ‘들쭉날쭉’... 여론조사 이대로 괜찮나” (2023/05/23); 조선일보. “여기는 18%, 저기는 30%... 여론조사 더워 먹었나”(2021/08/16); “[여론&정치] 못 믿을 정당 지지를 조사” (2022/07/16); “추석 민심 보면 안다? ‘널뛰기 조사’로 총선 예측 안갯속” (2023/09/30)

14) 사실 과거에 간혹 발견되던 전화면접조사 간 차이는 주로 대통령 긍·부정 평가 비율에 발견되곤 했다. 2점 척도로 측정되는 갤럽조사 결과는 대체로 4점 척도의 다른 전화면접조사에 비해 상대적으로 국정 긍정 평가가 낮게 나타나는 경향이 있었는데 지난 대선 전후부터 현재까지 이러한 척도 간 차이 효과는 크게 눈에 띄지 않는다. 또한 NBS는 재질문을 안 하지만, 갤럽은 1차 질문 시 무응답자에게 재질문하는 차이도 있다.

15) 윤 대통령 취임 이후인 2022년 5월 2주 퍼터 10월 2주까지 양 기관이 동일시점의 조사(같은 주에 실시, NBS는 월,

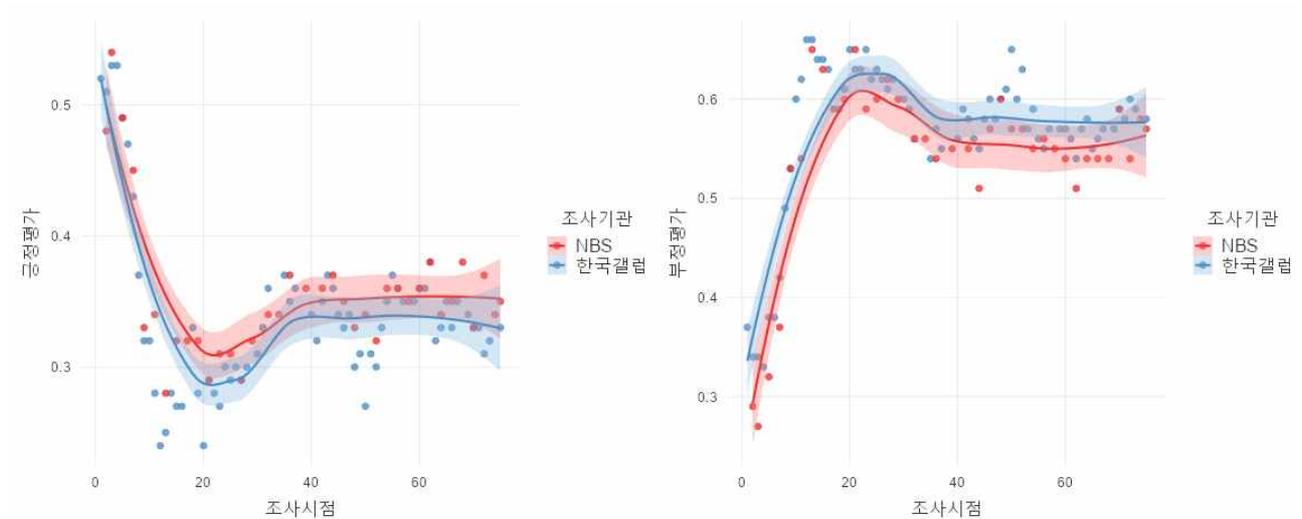
[그림6] 윤 정부 시기 NBS 조사와 갤럽 조사 변동 추이



자료: NBS 전국지표조사 리포트(2022.5월2주~2023.10월2주); 한국갤럽 데일리오피니언(2022.5월2주~2023.10월2주)

[그림7] 동일시점 긍정 평가: NBS/갤럽(총 N=60, 각 N=30)

(1) 긍정평가 비율 차이 (2) 부정평가 비율 차이



주: 회귀선은 LOESS method로 산출, 그래프는 JAMOVİ program 활용함(Jamovi project 2023).

② NBS와 갤럽 차이가 발생하는 지점

그러나 유권자 정치지형 분석의 핵심지표인 정당 지지율과 선거구도 예측의 주요 지표 중의 하나인 차기 총선 구도 문항을 보면 체감적으로 작지 않은 차이가 일관되게 발표되면서 혼란이 생기고 있다.

정당 지지율/무당파 비율이 다르다

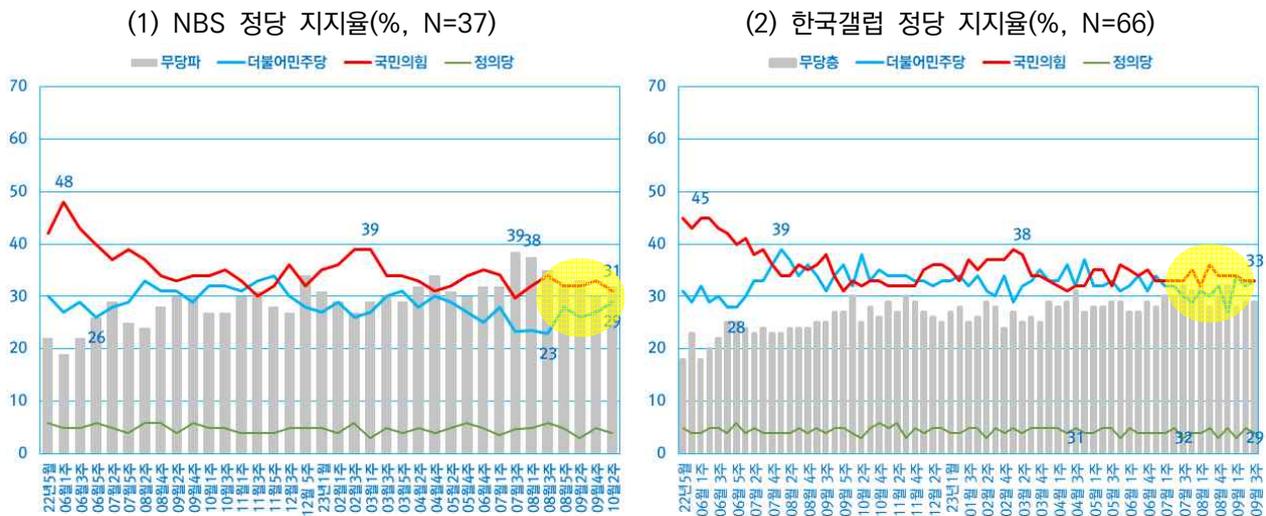
정당지지율 변화의 추세 자체로는 양 기관의 발표에 큰 차이가 발견되지 않는다. NBS 조사에선 윤 대통

화,수 조사 목요일 발표, 한국갤럽은 화,수,목 조사 금요일 발표)를 발표한 총 30회 조사 결과를 비교한다. 조사시점은 1주차(22년 5월2주)~20주차(22년 9월5주)~40주차(23년 2월1주)~60차(23년 6월4주)~75차(23년 10월2주)임.

령 임기 초에 여야 정당 지지율은 비슷한 수준에서 출발했지만, 6월 지방선거를 거치며 더불어민주당 지지율은 30%에 못 미치는 수준에 머문 반면, 국민의힘 지지율이 48%까지 상승하며 ‘허니문 효과’와 더불어민주당의 ‘검수완박’ 역풍의 반사이익을 누렸다. 갤럽조사에서도 5월 국민의힘 지지율은 45% 내외, 더불어민주당 지지율도 28%까지 하락하면서 두 조사 간 큰 차이를 체감하기 어렵다. 지방선거 이후 이준석 대표 징계(22년 7월)를 둘러싼 내분과 이태원 참사 등의 악재와 함께 지속적으로 하락하고, 2022년 하반기 대통령의 “4대 개혁 드라이브”와 함께 대통령과 국민의힘 지지율이 동반 회복세를 기록하면서 2023년 2~3월에는 39%대까지 회복하는 듯했다. 그러나 이후 2023년 김기현 대표체제 등장(3월) 이후 30% 초반대의 지지율에 고착된 양상이다. 갤럽조사에서도 국민의힘 지지율에서는 NBS 조사와 큰 차이가 확인되지 않는다.

반면 더불어민주당 지지율은 NBS 조사에서는 이태원 참사 등 여당의 악재가 집중되었던 2022년 하반기에 양당 지지율 격차가 줄었지만, 2023년 5월 이후에는 대체로 20% 중 후반대에서 하락하면서 국민의힘 지지율에 뒤처지는 국면이 유지된다. 2023년 하반기에는 국민의힘 지지율 정체와 더불어민주당 지지율 하락세로 인해 무당파 비율이 38~39%대까지 증가하는 양상을 보여준다. 그러나 8월 이후 무당파가 감소하면서 민주당 지지율이 23%(8월 3주)에서 29%(10월 2주)로 상승하며 양당 지지율이 팽팽하게 경합하는 국면으로 변화하였다. 반면 갤럽의 정당 지지율 변화추이를 보면 더불어민주당 지지율의 경우 NBS 조사보다 상대적으로 높게 나타나면서 국민의힘과 오차범위에서 경합하는 한편(2022년 하반기부터 대부분 30% 초반대), 무당파의 비율이 상대적으로 낮게 나타나는 것으로 보인다(그림8).

[그림8] 윤 정부 시기 정당 지지율 차이(%)



자료: NBS 전국지표조사 리포트(2022년 5월 2주~2023년 10월 2주); 한국갤럽 데일리오피니언(2022년 5월 2주~2023년 10월 2주)

동일 시점의 조사 차수만 뽑아서 정당지지율을 비교해보면 국민의힘 지지율에서는 유의한 차이가 발견되지 않으나 (1) 더불어민주당 (2) 제3정당(정의당+기타정당) (3) 무당파(없음/모름/무응답) 비율에서 양 기관의 차이가 오차범위를 넘어선 차이로 벌어지고 있다. 동일 조사 시점별로 윤 정부 초기부터 10월 2주까지 발표된 결과를 보면 다음과 같은 특징이 발견된다.(그림9)

첫째, 시기별로 차이는 있지만, 대체로 NBS 조사(붉은 점/선)는 더불어민주당 지지율에서는 갤럽조사(파란 점/선)보다 낮게 나타나고, 반대로 제3정당(정의당+기타정당) 비율과 무당파 비율(없음/모름/무응답)에서는 높게 나타난다. 같은 시점에 조사한 30개 조사의 국민의힘 지지율 평균은 양 기관이 같지만(NBS 35.2%, 갤럽 35.2%), 더불어민주당 지지율은 NBS 조사가 평균 3.53%p 가량 낮고 이는 통계적으로 유의한 수치로 나타난다(독립표본 T 검정). 반면 제3정당과 무당파 비율 평균에서는 NBS 조사가 각각 6.6%, 29.0%로 갤럽의 5.2%, 27.1%보다 높고 이 차이 또한 통계적으로 유의한 차이로 나타난다. 상대적으로 강한 관여층이 많은 더

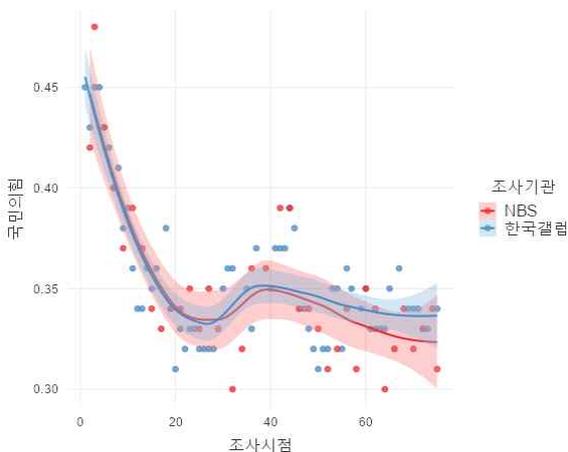
불어민주당 지지층은 갤럽에서 높게 나오고, 대신 NBS 조사에서는 상대적으로 무당파/소수정당 지지층이 높게 나오는 셈이다(각 평균 비율 차이에 대한 독립 표본 T-검증, 부록-표2 참조).

둘째, 양 기관의 차이가 일괄적으로 나타나는 것이 아니라 특정 시점에 집중되고 있으며, 최근에는 그 격차가 감소하는 특징을 보여준다. 40차 조사 시점(23년 2월) 이전까지는 정당 지지율에서 큰 차이가 발생하지 않았지만, 그 이후부터 2023년 7월경까지 양 기관의 정당 지지율 차이가 두드러지게 나타난다(통계적으로 유의한 차이). 그러나 다행히 2023년 8~9월을 거치면서 양 기관간 더불어민주당, 제3정당, 무당파 비율의 격차는 다시 좁혀지는 경향을 보인다. 이를 NBS 조사 결과 중심으로 요약하면 아래의 세 국면으로 요약할 수 있을 듯하다. 앞으로 양 기관의 차이가 어떻게 변화할지 지속적으로 관찰할 필요가 있다. 대체로 민주당과 무당파/제3당 지지 변동의 시점별 변동의 민감성이 NBS 조사가 큰 것으로 보인다(그림9).¹⁶⁾

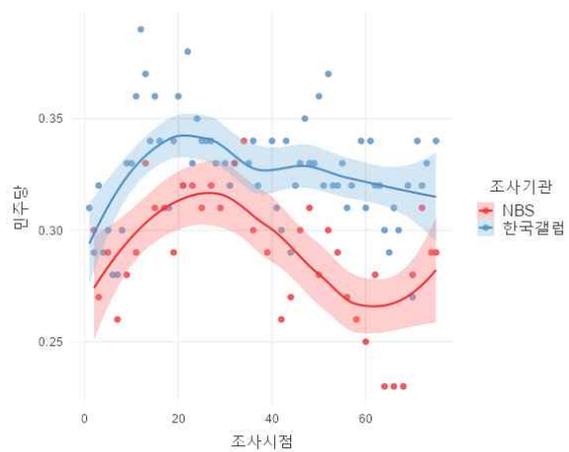
- 제1국면(22년 5월 1차~22년 12월 34차) : 국민의힘 하락, 더불어민주당 회복, 무당파 증가
- 제2국면(23년 1월 35차~23년 6월 61차) : 국민의힘 정체, 더불어민주당 하락, 무당파 증가
- 제3국면(23년 7월 62차~23년 10월 75차) : 국민의힘 정체, 더불어민주당 회복, 무당파 정체

[그림9] 동일 시점 정당 지지율 차이 : NBS/갤럽(총 N=60, 각=30)

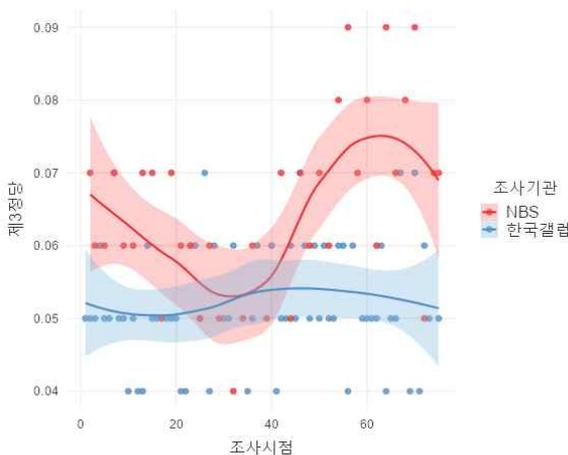
(1) 국민의힘 지지율



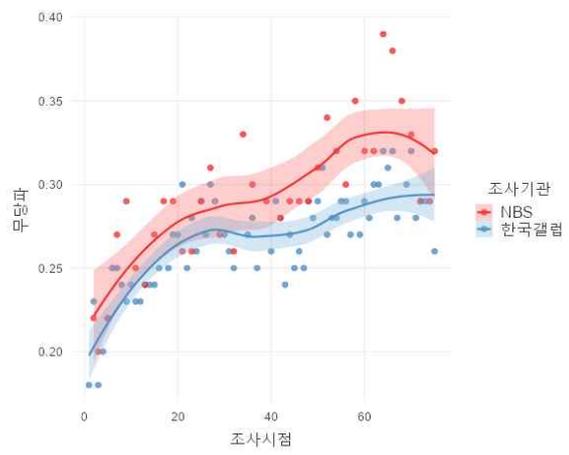
(2) 더불어민주당 지지율



(3) 제3정당(정의/기타) 지지율



(4) 무당파 비율



주: 회귀선은 LOESS method로 산출, 그래프는 JAMOV program 활용함(Jamovi project 2023).

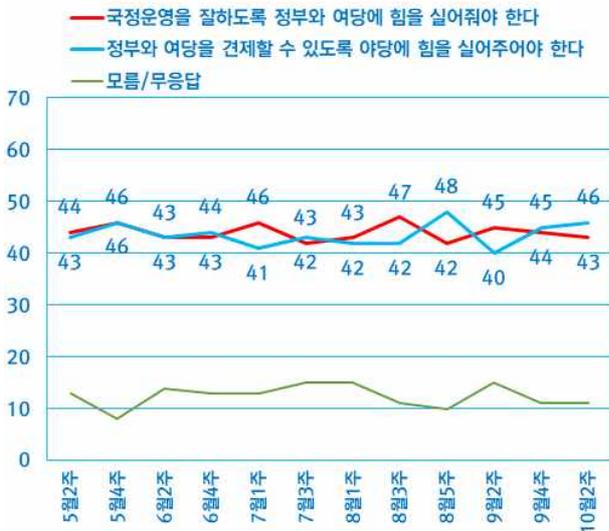
16) 양 기관조사 결과의 차이가 나타나는 시기(양 기관 회귀선과 산점도가 겹치지 않는 구간 발생)는 대체로 40차수(23년 2월)부터 60차수 대(23년 6월)까지 집중되고 7~8월 이후에는 국민의힘을 제외한 더불어민주당, 제3정당, 무당파에 대한 양 기관 추세선 신뢰구간(confidence interval)이 겹치면서 차이가 줄어드는 경향을 보인다.

총선 구도에 대한 상이한 결과

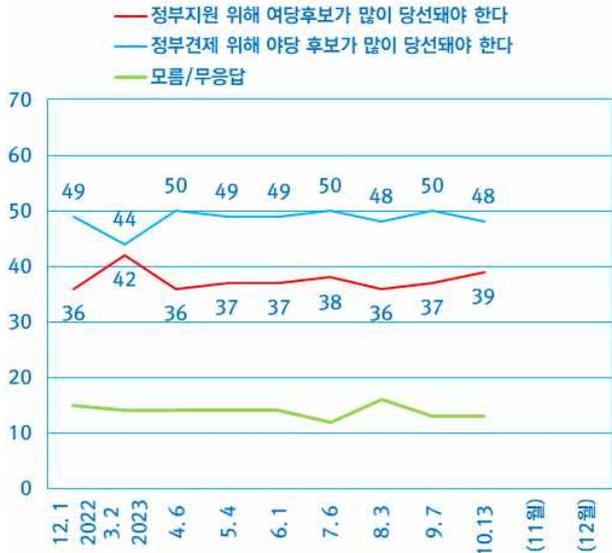
차기 총선 구도 관련 문항에서는 정당지지율보다 양 기관의 조사 결과 차이가 뚜렷하게 나타난다. 차기 총선에서 유권자들의 투표 선호를 엿보기 위한 총선 구도 문항에서 NBS는 2023년 5월부터 격주로 조사에 포함되고 있는데 ‘국정 안정-여당 지원론’과 ‘정부·여당 견제-야당 지원론’의 비율이 큰 변동 없이 오차범위 내에서 경합 국면으로 나타난다. 갤럽조사에서는 2023년 4월 이후 조사에서는 안정적으로 ‘정부를 지원하기 위한 여당 승리 기대’와 ‘정부 견제를 위해 야당 승리 기대’ 중 ‘정부를 견제하기 위해 야당 후보가 많이 당선돼야 한다’는 견제론이 12~14%p 우위로 나타난다. 양 조사 응답 간 차이도 차이지만 최근 변화의 방향에서도 미묘한 차이가 있다. 갤럽조사에서는 양 응답 간 차이가 유지되거나 줄어드는 추세이나 NBS 조사에서는 정부·여당 지원론이 정채되면서 견제론이 늘어나는 조짐이 보인다. 상당한 인식의 혼란이 불가피한 상황이다(그림10).

[그림10] 제22대 총선 인식조사 결과 차이(%)

(1) NBS



(2) 한국갤럽



자료: NBS 전국지표조사 리포트(2023년 5월 2주~2023년 10월 2주); 한국갤럽 데일리오피니언(2022년 12월~2023년 10월 2주)

NBS와 개별기관 간 차이도 혼선

최근에는 NBS와 NBS 참여 기관 개별 조사 간 주요 정치 지표와 총선 여론에서도 뚜렷한 차이가 나타나면서 혼선이 가중되기도 한다. 실제로 [표4]를 보면, 추석 특집 언론사 조사(9월 3주)나 비슷한 시기에 발표된 NBS(9월 3주), 갤럽조사(9월 2주)는 물론 NBS 참여 조사기관이 개별적으로 언론사 의뢰로 진행한 조사에서 문항에 따라 상당한 차이들이 발견된다(표4).

국정 평가에서 이들 조사 결과간의 차이를 발견하기 어려웠으나 역시 정당지지율에서는 한 주 전에 조사했던 갤럽조사와 NBS 조사의 국정 지지율이 오차 범위 밖에서 우세했고, 코리아리서치(MBC)/한국리서치 조사(KBS)에서는 양당 지지율이 오차범위 내 경합 구도였고, 엠브레인 조사(YTN)에서는 더불어민주당 지지율이 오차범위 밖 우위를 보였다. 무당파 비율도 갤럽/NBS 조사에서는 34%, 31%로 30%를 훌쩍 넘었지만, 코리아리서치/한국리서치 조사에서는 23%로 차이가 났다. 총선 구도 문항도 NBS는 ‘안정론’이 44%, ‘견제론’이 45%로 팽팽하지만, 나머지 조사에서는 대체로 ‘견제론’이 오차범위 밖에서 우위를 보여주고 있다(표4).¹⁷⁾

17) 무당파 비율에서는 엠브레인 조사가 36%로 가장 많았지만, 다른 조사들이 1차로 질문할 때 없거나 모르겠다고 답한 응답층에 2차 재질문을 하였지만, 엠브레인은 재질문을 하지 않아 다른 조사들과 함께 비교 가능한 조사 결과는 아니다.

[표4] 추석 민심 여론조사에서 나타난 NBS 조사/NBS 소속 기관 조사 간 응답 차이(%)

| 발표일 | 조사 기관 | 의뢰 기관 | 응답률 | 국정 평가 | | 지지 정당 | | | 총선 구도 | | 문항 순서(번호) | | | 문항 수 | | |
|------|--------|-------|------|-------|----|-------|----|----|-------|----|-----------|----|----|-------|-------|-------|
| | | | | 긍정 | 부정 | 민주 | 국힘 | 무당 | 안정 | 건제 | 국정 | 정당 | 총선 | 본 문항수 | 할당/배경 | 총 문항수 |
| 9/22 | 한국갤럽 | 자체 | 13.4 | 32 | 59 | 27 | 35 | 34 | - | - | 1 | 3 | 4 | 12.5 | 6 | 18.5 |
| 9/28 | NBS | 자체 | 20.0 | 34 | 58 | 27 | 33 | 31 | 44 | 45 | 1 | 6 | 7 | 12 | 8 | 20 |
| 9/27 | 코리아리서치 | MBC | 12.4 | 36 | 59 | 35 | 34 | 23 | 39 | 53 | 1 | 3 | 4 | 14 | 6 | 20 |
| 9/28 | 한국리서치 | KBS | 13.8 | 35 | 59 | 36 | 33 | 23 | 39 | 52 | 1 | 15 | 7 | 16 | 7 | 23 |
| 9/28 | 엠브레인 | YTN | 11.4 | 34 | 60 | 34 | 27 | 36 | 35 | 48 | 7 | 1 | 4 | 11 | 5 | 16 |

자료: 중앙선거여론조사심의위원회(여론조사 등록자료)

3. 차이가 발생하는 이유는? 유력 가설들

면접조사 간 차이는 왜 나타날까? 기존 가설과 새로운 가설

기존의 유력 가설 : 유무선 RDD 대 가상번호

과거에는 전화면접조사 간 차이가 발생할 때 크게 양 조사 간의 조사 방법(유무선 비율) 및 표본추출 방법의 차이(가상번호 층화표집 vs RDD 확률 표집)가 응답자 정치 성향 분포의 차이를 낳는다는 해석이 주를 이루어 왔다.¹⁸⁾

실제로 2023년 6월까지 [표5]와 같이 NBS 조사와 갤럽조사가 동일한 전화면접조사를 채택했지만, 표본 추출 방법이나 가구전화 포함 여부 등에서 작지 않은 차이가 있었다. NBS 조사는 2020년 7월 첫 조사 이후 줄곧 ① 100% 휴대전화로 ② 이동통신 3사가 제공하는 지역*성*연령 집단별 가상번호를 (목표 표본수의 20배수를) 받아 조사를 진행하는 ‘층화확률표집(stratified random sampling)’ 방법을 적용했지만, 한국갤럽은 2023년 7월 조사 방법을 변경하기 전까지 ① 휴대전화와 가구전화 혼합(유선 비율 5~15%)¹⁹⁾ ② 임의전화걸기 확률 표집(Random digit dialing) 방법을 적용해왔다. 1장에서 살펴본 대로 가상번호 사용 여부/유무선 여부는 응답률에 영향을 미치고, 응답률이 높을수록 중/저 관여층(인구학적으로는 젊은 세대/여성, 정치적으로는 중도/무당파) 등의 비당파적 유권자 표본 비중이 높아져 조사 결과의 차이에 영향을 미친다.

[표5] 한국갤럽 조사 방법의 변화와 NBS 조사 방법

| | 한국 갤럽(조사 방법 변경 전후) | | NBS |
|-------|-----------------------------------|--------------|---------------|
| 공표 개시 | 2012년 1월~2023년 6월 | 2023.7월 이후 | 2020년 7월 이후 |
| 조사 방법 | 유무선 RDD 전화면접 (유선5~15%) | 무선 가상번호 전화면접 | 무선 가상번호 전화면접 |
| 표본 추출 | RDD 확률 표집(할당?) | 층화확률표집 | 층화확률표집 |
| 조사 기간 | 매주 5일(12년)→4일(13년)→3일(14년 이후) | | 격주 (대선 기간 매주) |
| 표본 수 | 1,500명→1,200명(13년)→1,000명(14년 이후) | | 1,000명 내외 |

자료: NBS 전국지표조사 홈페이지(<http://nbsurvey.kr/>); 한국갤럽 홈페이지(<https://www.gallup.co.kr/gallupdb/report.asp>)

18) 한겨레21. “여론조사 회사가 반성문을 쓴 이유는” (2020/09/11); 한국일보. “‘단 하루 차이인데’... 널뛰는 이재명·윤석열 지지율, 왜?” (2021/11/23). 앞서 소개한 2021년 11월 상반된 대선 후보 지지율 조사가 발표되었을 때 직전 조사 대비 NBS 조사에서는 “NBS 조사에서 진보 유권자 비율이, 갤럽 조사에서는 보수 유권자 비율이 우연히 반대 방향으로 변동하며 큰 차이를 보인 것”이라는 당시 한국통계학회장이었던 김영원 숙명여대 교수의 진단이었다(한국일보. “단 하루 차이인데... 널 뛰는 이재명·윤석열 지지율, 왜?” (2021/11/23)).

19) 한국갤럽 데일리 오피니언 정기조사의 경우 매주 유무선 혼합조사로 진행해왔는데 초기에는 가구전화(유선) 비중을 15% 내외에서 시작해서 조사 방법 변경 전에는 5%까지 낮춰왔다. 또한 조사기관과 표본 수의 경우 조사 시작 첫해인 2012년에는 조사 기간 5일(표본 수 1,500명), 2013년에는 4일(1,200명)로 진행했고 2014년부터 이후부터는 3일(1,000명) 조사 일정을 지켜왔다. NBS는 2020년 첫 조사부터 조사 기간 3일, 표본 수 1,000명 기준으로 조사한다.

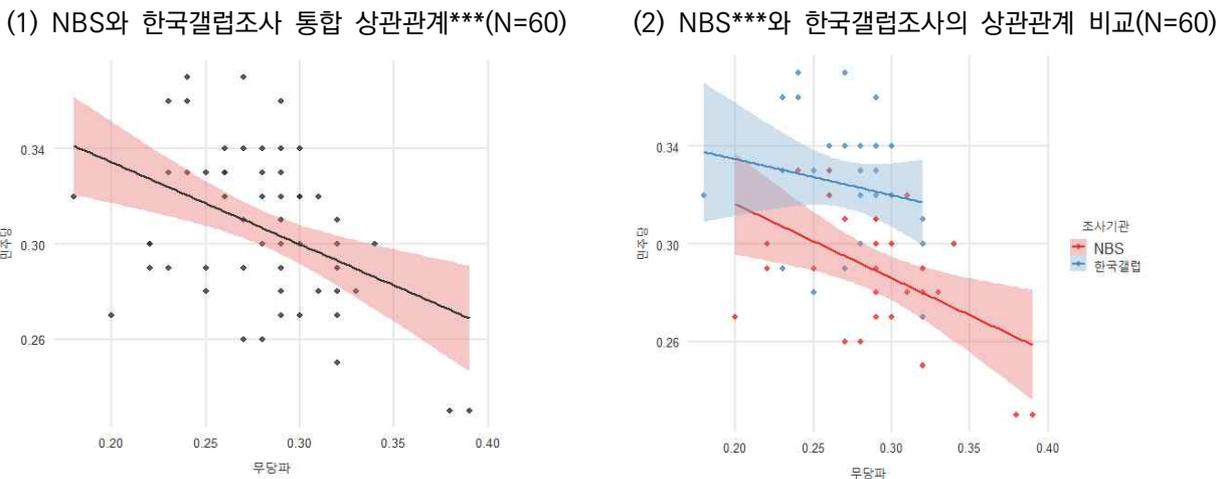
더욱 엄밀한 분석과 검증을 필요로 하는 현상이지만, 전화면접조사 기준으로 윤석열 정부 초기는 더불어민주당 지지율의 변동이 무당파 규모의 변동으로 이어지는 경향이 정당 지지율 변화를 만들어는 주된 요인이었다면(필자의 표현으로는 이탈 민주 층이 집중적으로 발생하는 시기), 이후 22년 하반기부터 23년 7월까지의 윤석열 정부와 국민의힘에 대한 실망으로 45~48%까지 육박했던 국민의힘 지지율이 하락한 것으로부터 무당파의 증가가 유발되었음을 시사하는 결과다. 같은 방식으로 해석하자면 2023년 7월 이후로는 국민의힘 지지율이 고착된 가운데 민주당에 대한 정당 태도 변동이 무당파 비율과 연동되는 시기라고 볼 수 있다. [그림 9-(2)]의 변화추이를 NBS 조사 기준으로 해석 보자면 정권 초기 민주당 지지율과의 상관관계가 주도했던 시기에는 주로 민주당 지지율의 하락이 무당파 상승으로 이어졌다면, 최근의 상관관계는 최고조에 다다랐던 무당파가 감소하면서 민주당의 지지율 회복이 이루어지는 상반된 상황일 수 있다는 점을 주목할 필요가 있다.

새로운 가설들 : (1) 정치 성향 분포의 차이 유발요인(응답률/조사길이/시점 효과-정치 상황)과 (2) 워딩 효과

본 보고서에서 주목하는 NBS와 한국갤럽 조사만 보더라도 역시 윤석열 정부 초기(22년 5월 2주부터 23년 10월 2주까지)의 조사 결과를 통합해 ‘무당파’ 비율과 ‘더불어민주당’ 지지율 간의 상관관계를 보면 역시 무당파의 비율이 높은 결과에서는 민주당 지지율이 낮고, 반대로 무당파 비율이 낮은 시점에는 민주당 지지율이 높은 경향성이 뚜렷하다. 양 기관 결과만을 놓고 보면 특히 NBS 조사에서 무당파 비율과 민주당 비율 간의 상관관계가 강하게 나타나고, 갤럽조사에서는 상대적으로 상관관계가 미약한 것으로 나타난다(그림12).

결국 현재 조사 결과의 혼선이 주로 더불어민주당 지지율과 무당파 비율과의 상관관계에 집중되어 있다. 그렇다면 현재 조사 방법 간 차이를 만들어내는 요인으로 ‘무당파’ 비율과 정당 지지층(특히 더불어민주당 지지율)에 영향을 미칠 요인에 주목해야 한다. 그 중 무엇보다 동일한 전화면접조사 내 응답률에 영향을 미칠 요인들을 살펴보는 것이 순서일 듯하다. 결국 응답률이 높아야 ‘무당파/중·저 관여층’ 표본이 제대로 표집될 것이기 때문이다. 이 외에도 최근 과거에 주목하지 않았던 요인들에 대한 흥미로운 가설들이 제기되고 있다. 추석 특집 조사의 혼선에 대해 (1) 문항 길이, (2) 조사 기간(갤럽/NBS/한국리서치 3일, 코리아리서치/엠브레인 2일), (3) 정당 지지율 질문 시 재질문 여부 등의 요인이 주목받았다(김영원 2023).²¹⁾

[그림12] 동일 시점 무당파 비율과 민주당 비율 간의 상관관계



주: 회귀선은 OLS Regression method, 그래프는 JAMOVI program 활용함(Jamovi project 2023).

21) 한국리서치·KBS 조사 질문지는 본 문항과 배경 문항을 포함하여 23문항으로 다른 조사들이 16~20문항 이내였던 것과 대비하여 길다. 한국리서치가 참여한 NBS 조사 응답률이 20.0%였는데 KBS 조사 응답률은 13.8%에 그쳤다. 문항 길이가 길면 응답률이 하락할 수 있다. [표4]의 문항 수는 단일문항은 1개, 척도형 세트 문항은 문항당 0.5개로 환산. 또한 일부 조사에서는 정당지지율 측정 시 재질문 없이 측정함으로써 상당히 차이나는 무당파 비율이 발표되기도 했다.

최근 문항 순서 효과에 주목하는 논의도 유의미하고 흥미로워 보인다.²²⁾ 앞서 [표4]에 정리한 국정 지지율, 정당지지율, 총선 구도 문항의 문항 번호를 보면 대부분의 전화면접 조사에서 국정 지지율 문항을 1번에 배치하고 정당지지율을 초중반에 묻고(꺾림 3~4번, NBS 3~6번), 그 뒤에 총선 구도를 묻는 순서가 배치되는 경우가 많았다. 그런데 이번 추석 직전 조사에 한국리서치·KBS 조사의 경우 정당지지율이 후반부(15번)에 배치되고, 엠브레인·YTN 조사의 경우 정당지지율 문항을 1번, 총선 구도 문항이 4번, 국정 지지율 문항이 7번에 배치되었다. 특히 ARS 조사에서는 상대적으로 정당지지율 문항을 1번에 배치하는 조사가 많다고 한다. 사실 문항 순서와 관련하여 좋은 자료 수집을 보증하는 절대적 법칙이 있는 것은 아니다(Groves et al. 2004, Stefkovics and Kmetty 2022). 다만 몇몇 질문지 후반부에 배치된 문항은 응답자 집중력이 떨어져 불성실 응답의 가능성이 크고, 선행 문항에 대한 응답에 영향을 받을 수 있어서 엄밀한 응답 분포 추정이 필요한 문항을 앞에 배치하는 것이 좋다. 반대로 초반에 심리적 부담이 크거나 인지적 접화효과가 큰 문항은 배치하지 않는 것이 좋다는 것이 일반적인 기준으로 제시되고 있다. 초반에 부담이 큰 문항을 접하면 중도 탈락 비율을 높여 응답률에 영향을 주거나, 특정 가치와 프레임을 활성화할 수 있기 때문이다. 다만 실제 영향을 실증적으로 검증하기 위해서는 초기 질문이 주는 인지적, 맥락적 효과에 대한 정교한 실험 디자인과 제3변인의 개입에 대한 엄밀한 통제가 필요하다. 실제 해외의 연구 결과를 보면 문항 순서를 실증적으로 지지해주는 연구들과 실제 실험 결과 문항 순서 효과가 실증적으로 검증되지 않는다는 견해도 맞는다. 그러나 질문 순서 효과를 검증할 데이터를 확보하지 못해서 본 보고서에서는 다루지 않는다(심형인 2017; Lee and Kim 2022; Hughes and Jones 2019).²³⁾

본 보고서는 응답자 성향 분포에 영향을 미칠 응답률 요인과 함께 총선 인식 문항을 중심으로 총선 '워딩 효과'를 경험적으로 검토해 보고자 한다. 최근 동일 전화면접 조사간 응답 차이가 선거 예측과 유권자의 투표 행태를 판단하는데 핵심 지표 중의 하나인 '총선 인식 구도' 문항에서 집중적으로 나타나고 있지만, 그 원인에 대해서는 실증적 분석이 제대로 이루어지지 않았던 것이 사실이다.²⁴⁾ 사실 총선 투표행태는 대통령 중심제하에서의 정부·여당에 대한 상별 투표의 의미(소위 중간평가로서의 총선)를 갖는 것으로 인식하는 경우가 대부분이다. 하지만, 차기 대선의 전초전으로서 양 진영의 경쟁이라는 '중앙정치적 요인'과 지역구 투표와 비례투표의 1인 2표제하에서 '지역 정치 요인'이 결합한 총선 투표 구도를 한 문항으로 측정하기는 쉽지 않은 과제다. 대체로 한국에서 총선 투표에는 정부 여당에 대한 회고적 평가(retrospective evaluation) 즉 '정권 심판/견제론'에 대한 찬성 여부가 작동하는 것으로 알려져 있다(강원택 편 2010; 장승진 2016, 2020). 그러나 역설적으로 민주화 이후 한국의 총선에서 정권심판론이 투표행태에는 유의미한 영향을 미치지 않지만, 실제 의석 변화로 이어진 사례는 찾아보기 힘들다(정한울 2013, 2020).

유권자들의 태도가 정부·여당에 대한 찬반이라는 이분법적 선택지로 결정하는 당파적 유권자층만 존재하는 것이 아니라 그에 맞서는 상대 진영에 대한 평가를 종합하여 정부·여당과 야당, 진보 진영과 보수 진영 각각을 평가하는 양면적/상충적 유권자층이 존재하기 때문이다(허석재·정한울 2017; 정한울 2013). 여론조사에서 워딩 효과 혹은 프레임 효과는 특정 대상에 대한 확고한 선호나 태도가 형성되어 있지 않은 유권자층에서

22) 순서 효과를 지적한 글로는 김봉신 “ARS 조사 읽는 법...민주당에 숨겨진 위기: 정당 지지도와 무당층 연구 ① - 문항 순서 효과” 오마이뉴스(2023/08/03) 및 신창운 “여론조사 결과, 그 행간 사이에 숨어있는 민심을 찾아라.” 시사저널(2023/09/11) 등을 참조할 것. 한국리서치 자체 기획조사나 필자가 관여했던 EAI 선거여론조사에서는 정당지지율 문항을 가장 후반부 혹은 배경 문항으로 측정하는 경우가 많았다.

23) 많은 조사에서 가장 정치적 관심이 집중되는 '국정 지지율' 문항을 앞에 배치하는 이유다. 그러나 실제 질문 설계에서는 이러한 가이드 간 상충하는 상황이 자주 발생하며 이는 조사 설계자의 연구 목적에 대한 고려나 조사 철학에 따라 선택의 문제가 발생하게 된다. 정당지지율 문항의 배치 순서도 논쟁 거리다. 선거 행태 연구에서 정당 태도는 핵심 태도 결정요인이기 때문에 선행 문항의 영향을 덜 받는 앞쪽에 배치해야 한다는 주장도 일리 있지만, 사실 지지 정당의 경우 개인의 정치 성향을 노골적으로 묻는 질문이기 때문에 특정 지지 정당이 없는 무당파나 저 관여층에는 상당한 심리적/정치적 부담감을 줄 수 있으며, 정당 태도는 상대적으로 안정적인 태도이기 때문에 선행 문항의 영향력 상대적으로 크지 않기 때문에 후반부에 배치해야 한다는 판단도 가능하다. 최근 문항 순서 효과 등에 관한 실증적 연구들이 나오고 있지만, 선거여론조사에 대한 검증은 미흡한 상황이다(심형인 2017; Lee and Kim 2022).

24) 워딩 효과에 관한 기사로는 박성민. “총선 결과 예측의 3가지 지표” 법률신문 (2023/09/18)을 참조.

나타나는 현상으로서 자신의 “진정한 선호나 의견(true opinion)이 아닌” 조사 질문의 특정 문구나 조사 프레임의 영향으로 특정 방향의 응답을 유도당하거나 오락가락 응답의 원인으로 지적되기도 한다(Zaller 1992). “대중은 비합리적인 존재다”라는 시각을 뒷받침하는 논의에 포함되지만, 양면적/상충적 태도(ambivalent attitudes)를 가진 유권자층의 다차원적, 다면적 태도를 가진 사람이 내적 갈등과 양면적 선호를 가지고 있을 때 조사 질문의 워딩과 프레임 효과가 나타난다는 반박도 있다(Alvarez and Brehm 2002).

본 보고서는 워딩에 따라 일관되지 않은 응답을 하는 워딩 효과를 ‘무식한 유권자(irrational voter)’의 특징으로 이해하기보다는 ‘상충적 유권자(ambivalent voter)’의 ‘태도 갈등(attitude conflict)’을 유발하는 환경과 시점의 결과로 해석하는 측면에서 검증할 것이다(Alvarez and Brehm 2002; Lavine et al. 2012; 정한울 2013). 2017년부터~2020년 총선처럼 대한민국 국민의 대다수가 대통령 탄핵에 대한 찬반을 기준으로 투표하게 되는 상황이나 양면적/상충적 유권자는 감소하고 일방적 태도가 주류 여론으로 등장하는 시기에는 ‘미묘한(nuanced)’ 워딩이나 프레임의 영향은 제한적이다. 그러나 2016년 총선, 2022년 대선처럼 여야 모두에 대한 실망이 공존하고 여야 동시 심판 정서가 확산되어 입장이 확고한 당파적 유권자들 외에 태도 갈등으로 유동적인 유권자층이 확산된 조건에서 워딩이나 프레임 효과(질문 순서도 마찬가지)의 가능성이 커질 수 있다는 것이 본 보고서의 문제의식이다.

4. 실증 분석 : 정당지지도와 총선 인식조사에서 양 기관의 차이를 발생시키는 요인은 무엇인가?

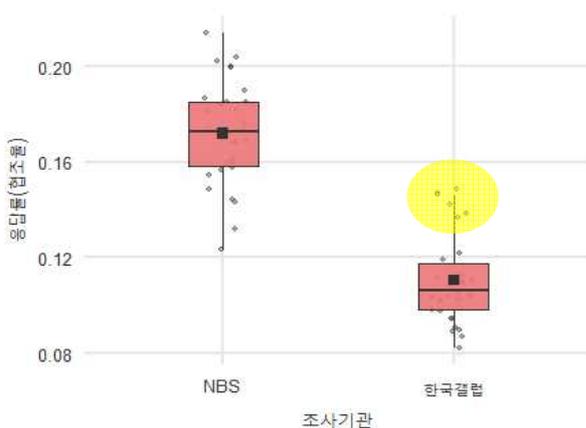
① 양 기관의 정치 성향 분포 차이 : 생각보다 복합적인 요인들이 상호영향을 미치고 있다

조사 방법 전환과 응답률 효과

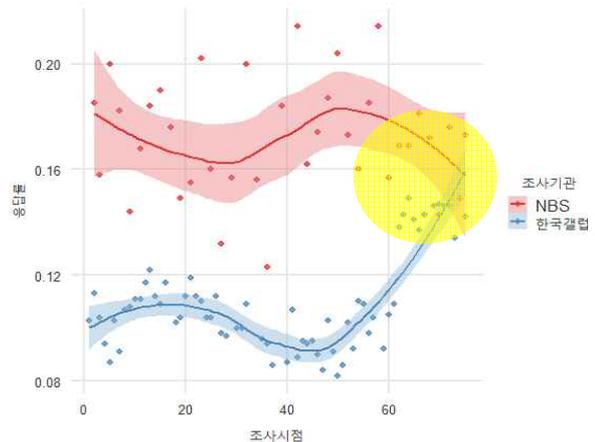
기존 등록된 여론조사들에 대한 메타분석에서 무당파/정당지지율 비율에 영향을 미치는 것으로 검증된 ‘응답률’이 NBS와 갤럽조사에서 어떻게 나타나는지 살펴보자. [그림13]에서 동일 시점의 응답률 평균을 보면 NBS는 30개 차수에서 17.2%지만, 같은 시점 갤럽조사의 응답률 평균은 11.1%로 6.1%p NBS 조사가 높은 것으로 나타났다(독립 T 검증, $p < 0.00$, 그림13-(1)). 시기별로 보면 61차 이전(23년 6월 5주)까지는 NBS 응답률이 갤럽의 응답률을 능가했다. 그러나 갤럽도 가상번호 방식으로 전환한 23년 7월 이후는 응답률이 유의미하게 상승하면서 응답률 차이가 감소하는 경향이 뚜렷하다(그림13-(2) 노란색 원이 7월 이후의 응답률).

[그림13] 동일 차수 NBS와 갤럽의 응답률 비교

(1) 동일 차수 응답률 분포(동일 차수, N=30**)



(2) 조사 시점별 응답률 차이 변화(동일 차수, N=30)

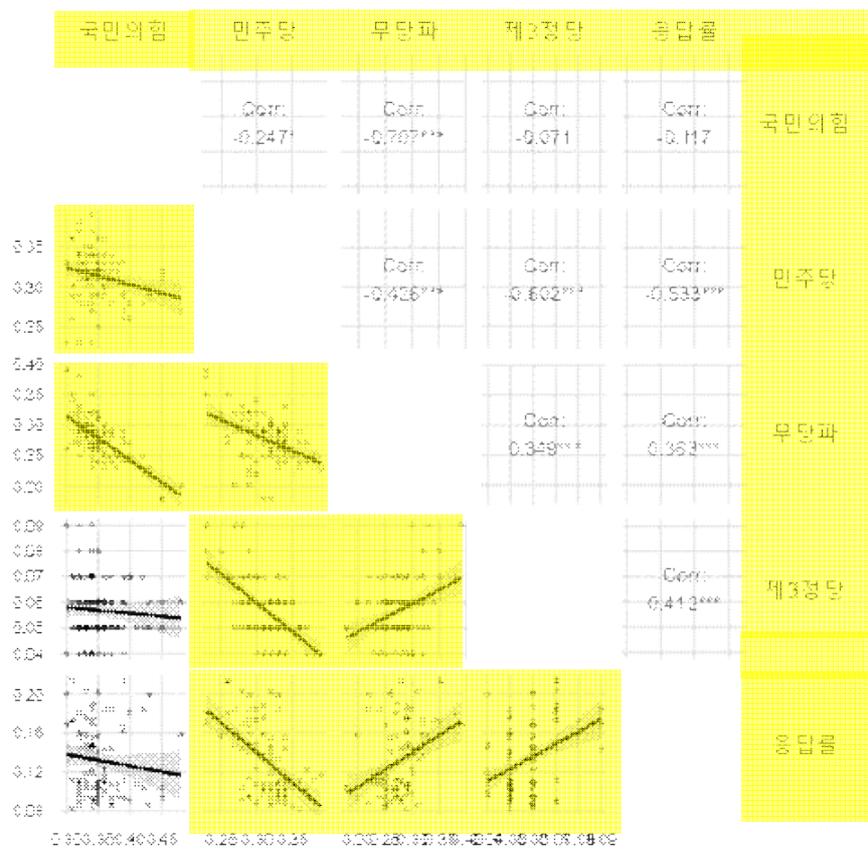


주1: 회귀선은 LOESS method로 산출, 그래프는 JAMOV program 활용함(Jamovi project 2023).
 주2: + (90% 신뢰수준), * (95% 신뢰수준), ** (99% 신뢰수준), *** (99.9% 신뢰수준)에서 통계적으로 유의한 차이

그렇다면 응답률이 선거행태(투표 선호, 투표 참여, 선거 인식 등)에 결정적 영향을 미치는 '정당 태도'에 어떠한 영향을 미칠까? 윤석열 정부 시기 양 기관의 동일 조사 시점 조사 결과에서 발표자들의 응답률과 각 정당에 대한 지지/무당파 비율이 어떠한 관계를 보이는지 살펴본 상관행렬 계수 및 도표가 [그림14]이다.

우선, 응답률의 크기와 국민의힘 지지율 간에는 뚜렷한 상관관계($r=-0.117$, $p>0.1$)가 나타나지 않는다. 앞서 살펴본 대로 응답률 높낮이에 무관하게 국민의힘 지지자들이 안정적으로 표집되고 있음을 시사한다. 다음으로 기존 연구 결과와 비슷하게 응답률은 더불어민주당 지지율, 제3정당(정의당+기타정당), 무당파 비율과 유의미한 상관을 보여주었다. 기존 결과들과 동일하게 응답률이 높은 조사일수록 더불어민주당 지지율이 낮은 경향($r=-0.533^{***}$, $p<0.001$, 음의 상관관계)이 뚜렷하다. 반면 응답률이 높은 조사일수록 무당파 비율과 ($r=+0.363^{***}$, $p<0.001$) 제3정당 비율($r=+0.412^{***}$, $p<0.001$)이 높게 나타나는 경향이 나타난다.

[그림14] 응답률과 각 당 정당 지지율 간 상관관계(NBS N=37, 갤럽 N =66)



주1: 회귀선은 OLS Linear Regression, 그래프는 JAMOVİ program 활용함(Jamovi project 2023).
 주2: + (90% 신뢰수준), * (95% 신뢰수준), ** (99% 신뢰수준), *** (99.9% 신뢰수준)에서 통계적으로 유의한 차이, 노란색 셀

양 기관의 정당정치 성향 분포 차이: 조사기관 효과인가?

다만 이러한 응답률이 조사기관, 조사 시점 등을 통제하더라도 나타나는 효과인지, 아니면 응답률이 높은 NBS와 갤럽 조사의 기관 효과 또는 특정 시점의 영향인지는 추가로 살펴볼 필요가 있다. 우선 기관 효과를 통제해보자. [그림14]에서 조사기관 구분 없이 통합하여 보면 응답률과 민주당 지지율(-)과 무당파/제3정당 지지율(+) 간에 뚜렷한 상관관계가 확인되었지만 [그림15]-(1)과 (2)처럼 조사기관별로 나누어 보면 각 개별기관의 개별 조사 결과에서는 상관관계가 약해지고(직선의 기울기가 [그림14]의 기울기 대비 수평에 가깝다), '조사기관' 효과가 커 보인다. 즉 조사기관에 상관없이 개별 조사에서 '응답률이 높으면 민주당 지지율이 하락하고 무당파/제3당 지지율 상승한다'라는 해석은 타당하지 않다는 것을 의미한다. 엄밀히 얘기하면 양 조사기관

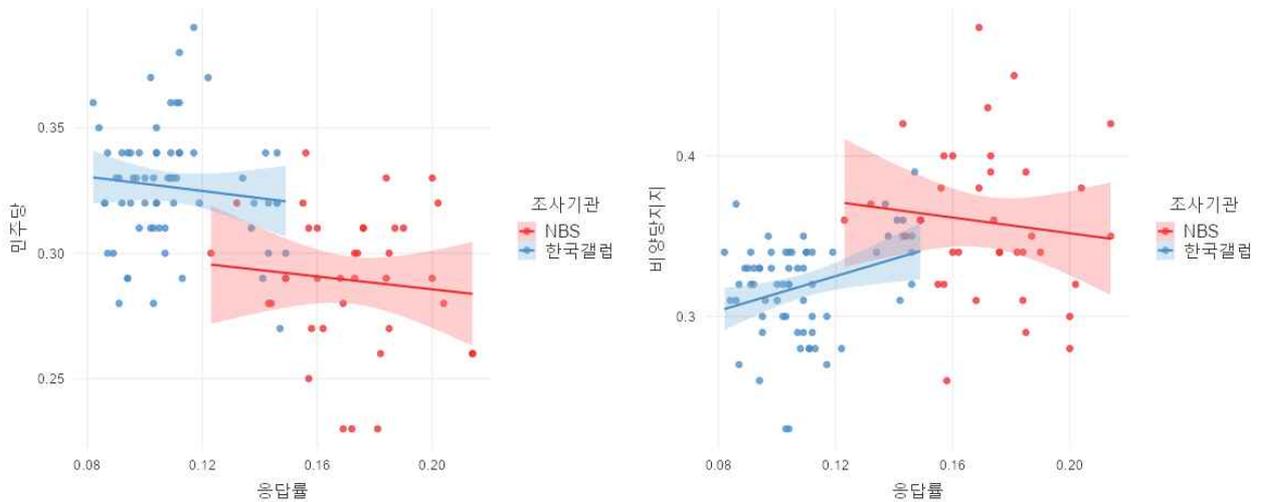
요인을 통제하면 ‘응답률 수준이 높은 기관(NBS, 붉은색)’이 평균적으로 민주당 지지율이 낮고 무당파/제3당 비율이 높지만, ‘응답률 수준이 낮은 기관(갤럽, 파란색)’에서 민주당 지지율이 높고, 무당파/제3정당 비율이 높다고 봐야 한다. 즉, 개별 조사 결과로 보면 응답률의 영향보다 조사기관의 차이로 해석될 수 있으며, 그 기관 특성 중의 하나로 전반적인 응답률의 수준이 고려될 수 있을 것 같다. 이러한 결과는 최근 조사기관 효과를 강조하는 주장을 뒷받침해준다(박종희 2013; 2021).

상관관계만 보면 응답률은 민주당 지지율(+), 무당파/제3정당 지지율(-)에 영향을 미칠 수 있음을 보여주었지만, NBS와 갤럽의 조사기관 효과를 통제하면 응답률 효과는 사라진다. [부록-표5]은 각 정당 및 무당파 비율에 대한 NBS 조사 결과(N=37)와 갤럽 조사 결과(N=66)들의 (1) 조사기관 (0=NBS, 1=갤럽) (2) 조사 방법 (0=유무선 RDD, 1=가상번호 총화표집)을 통제한 후 (3) 조사 시점(0=22년 12월 이전, 1=23년 6월 이전, 2=23년 7월 이후) (4) 응답률이 미치는 효과를 검증한 선형회귀분석 결과이다. 분석 결과를 보면 국민의힘 지지율을 제외한 더불어민주당, 제3당, 무당파 비율에서는 응답률 효과가 통계적으로 유의하지 않은 대신 양 조사기관 차이는 통계적으로 유의했다.

[그림15] 양 기관의 응답률과 민주당/무당파 비율 간 상관관계 차이(N=103)

(1) 더불어민주당 지지율에 미치는 효과

(2) 무당파+제3정당 비율에 미치는 효과



주: 회귀선은 OLS Linear Regression, 그래프는 JAMOVİ program 활용함(Jamovi project 2023).

응답률은 의미 없는 변수인가? 조사기관 및 조사 시점에 따라 응답률 효과가 달라진다

양 기관의 통합 집계자료 결과상으로는 응답률 차이가 정치 성향 분포의 차이와 관계가 있는 것으로 확인되는데, 기관별로 살펴보면 조사기관 효과로 보이고 응답률 차이가 영향이 없는 것으로 나타났다. 그렇다면 개별기관 차원에서는 응답률 효과가 없었다고 속단해도 될까? 더욱 면밀한 판단을 위해서는 혹시 조사 시점별로 다른 양상이 나타나는 것은 아닌지 살펴볼 필요가 있다. [그림]에서 살펴본 것처럼 양 기관의 각 정당 지지율이나 무당파 비율의 차이가 일정한 것이 아니라 조사 시점별로 차이가 나고 있기 때문이다.

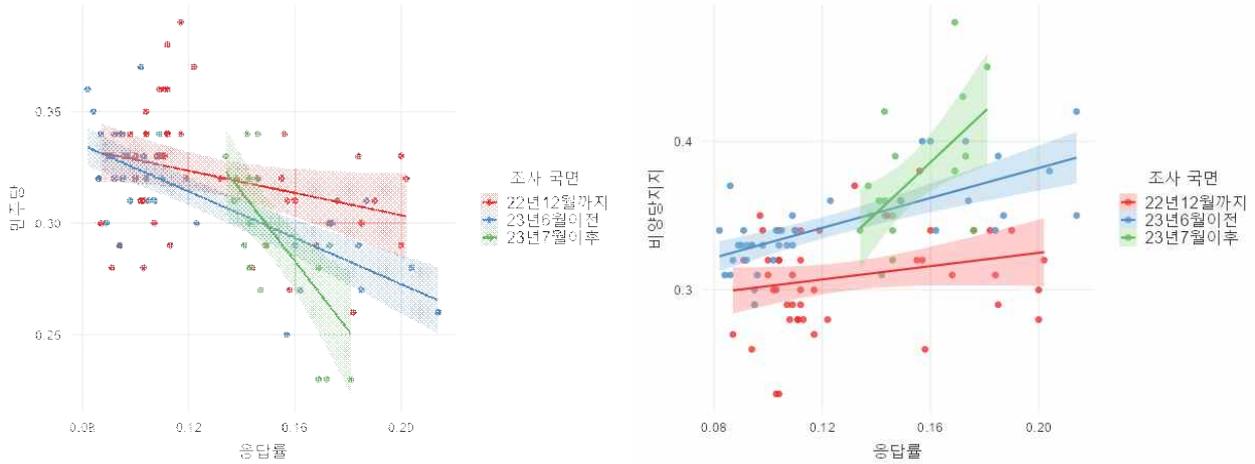
실제로 [그림16]에서 응답률이 더불어민주당/무당파 비율에 미치는 영향을 (1) 제1국면(~22년 12월 이전 (붉은색)) (2) 제2국면(23년 1월~23년 6월(파란색)) (3) 제3국면(23년 7월~10월 (연두색))별로 살펴보면 국민의힘 지지율이 하락하고 더불어민주당/무당파 비율이 하락하던 제1국면 대비 제2국면, 제3국면을 거치면서 응답률의 변화가 더불어민주당, 무당파에 미치는 영향력이 큰 것으로 나타난다. 특히 갤럽의 조사 방법이 가상번호로 바뀌면서 양 기관의 조사 방법이 같아진 23년 7월 이후에 응답률 효과가 뚜렷하게 나타나는 특징을 보여준다. 23년 7월 이후 시점을 보면 조사 방법의 차이로 응답률의 편차(8~22%)가 컸던 제1국면과 제2국면에 비해 응답률 편차(13~18% 수준)가 많이 좁혀졌지만, 오히려 줄어든 응답률 차이에 따라 연동하는 민주

당/제3정당 지지율 변동 폭이 커지고 있다.

[그림16] 조사 시점별 응답률과 민주당/비 양당 지지(무당파+소수정당)와의 상관관계 변화(N=103)

(1) 더불어민주당 지지율에 미치는 효과

(2) 무당파+소수정당 지지율에 미치는 효과



주1: 회귀선은 OLS Linear Regression, 그래프는 JAMOVI program 활용함(Jamovi project 2023).

더불어민주당 지지율/제3당 지지율의 경우 대체로 NBS와 꺾림 간 차이가 20주 차(22년 9월경)까지 유의미해 보이지만, 20~40주 차(22년 연말) 차이가 사라지고 다시 40~60주 차수(23년 상반기)에 차이가 집중적으로 벌어졌다가 7월 이후 감소하는 경향성을 보여주었다. 무당파의 경우 40주 차(22년 연말)까지는 조사 결과 차이가 나타나지 않았지만 역시 40~60주 차(23년 상반기)에 차이가 집중적으로 나타났고, 23년 하반기부터는 감소하는 상황이다. 조사 시점에 따라 응답률이 민주당 지지율이나 무당파/제3정당 지지율에 미치는 효과가 달라지는 점을 고려할 때, 조사기관별 응답률이나 정당 성향 분포의 차이를 설명할 때 조사 시점에 대한 고려가 필요하다는 것을 말해준다. 앞의 [그림15]에서 전 기간으로 보면 조사기관을 구분해서 보면 응답률과 민주당 혹은 무당파/제3정당 지지율 간의 상관관계가 사라지거나 크게 약화되는데 이 역시 조사 시점별로 구분해보면 응답률 효과를 보다 정교하게 이해할 수 있을 듯하다.

조사기관, 조사 시점 효과 상호작용 : 조사 방법, 효과에 대한 성급한 공식화를 조심해야 한다

이러한 조사 기간(period effect)에 따라 응답률 효과나 정당 지지율에 조사 방법이 미치는 영향력이 커질 수 있음을 고려하여 국민의힘, 더불어민주당, 제3정당, 무당파 비율 각각을 종속변수로 하고 이들 비율의 차이에 영향을 미칠 수 있는 요인들에 대한 회귀분석 검증을 수행했다. 설명 변수로 (1) 조사 기관(0=꺾림조사, 1=NBS) (2) 조사 방법(0=RDD, 1=가상번호) (3) 조사 시점 (0=22년 5월~12월, 1=23년 1월~6월, 2=23년 7월~10월) (4) 본 연구의 핵심 변수인 응답률을 주 효과 변수로 포함하고 상호작용 효과를 보기 위해 (5) 조사기관과 응답률의 상호작용 항 (6) 조사 시점과 응답률 상호작용 항을 분석 모델에 포함했다.

[표6]의 분석 결과를 보면 첫째, 응답률은 국민의힘(B=-1.076+), 제3정당 지지율(B=-0.307+)에는 유의한 영향이 확인되지만 본 조사에서 관심을 두고 있는 더불어민주당과 무당파 비율에 대해서는 유의한 효과가 나타나지 않는다. 이는 [그림15]의 결과와 맥을 같이 하는 결과다.²⁵⁾ 둘째, 그렇다고 응답률은 아무런 영향을 미치지 않는다고 속단해서는 안 될 듯하다. 응답률 단독의 영향이 아니라 조사기관, 조사 시점과 상호작용(interaction effect)을 통해 더불어민주당, 제3정당, 무당파 비율의 변화에 응답률이 미치는 영향이 교란되고 있음이 확인된다. 조사기관*응답률의 상호작용 항의 분석 결과를 보면, 꺾림과 달리 NBS 조사에선 응답률이

25) 본 분석모델은 상호작용을 포함한 결과이다. 상호작용항을 포함하지 않고, (1) 조사기관 (2) 조사 방법 (3) 조사 시점 (4) 응답률 변수 만으로 이루어진 회귀분석 결과에서 응답률은 유의하지 않은 것으로 확인되었다(부록-표5 참조).

높으면 민주당 지지율이 하락하는 것으로 나타났다(B=-0.908*). 그리고 조사 시점*응답률의 상호작용 항을 보면 응답률 상승이 민주당 지지율 하락으로 이어지는 효과는 최근 조사 시점의 조사일수록 유의한 것으로 나타났다(제2국면 시점의 응답률 효과 B=-0.452**, 제3국면 시점의 응답률 효과 B=-1.151+). 셋째, 무당파 비율의 경우 조사 시점 이외에 조사 시점 1(22년 5-12월) 대비, 조사 시점 2(23년 상반기), 조사 시점 3(23년 하반기) 조사들에서는 응답률이 높아질수록 무당파 비율이 증가하는 패턴이 확인되었다(조사 시점 2*응답률 항 B=+0.276+, 조사 시점 3*응답률 항 B=+1.280+).

[표6] NBS·갤럽조사 결과로 본 민주당·국민의힘·제3정당·무당파 비율에 대한 영향요인: OLS 회귀분석²⁶⁾

| 예측 모델 (N=103) | 국민의힘 | | 민주당 | | 제3정당 | | 무당파 | | |
|--------------------------|----------------|----------------|-----------------|----------------|-----------------|----------------|-----------------|----------------|---------|
| | 추정값(B) | 표준오차(SE) | 추정값(B) | 표준오차(SE) | 추정값(B) | 표준오차(SE) | 추정값(B) | 표준오차(SE) | |
| (Intercept) | 0.481** | (0.063) | 0.214** | (0.042) | 0.084** | (0.017) | 0.232** | (0.048) | |
| 조사 기관 (1=NBS, 0=갤럽) | -0.157+ | (0.093) | 0.107+ | (0.062) | -0.012 | (0.026) | 0.111 | (0.071) | |
| 조사 방법 (1=가상, 0=RDD) | 0.052 | (0.037) | -0.046+ | (0.025) | -0.008 | (0.010) | 0.005 | (0.029) | |
| 조사 시점 2.(~23년 6월) | -0.047 | (0.029) | 0.054** | (0.019) | -0.011 | (0.008) | -0.001 | (0.022) | |
| 1.(~22년12월) 3.(~23년 10월) | -0.069 | (0.156) | 0.154 | (0.104) | 0.067 | (0.043) | -0.146 | (0.121) | |
| 응답률 | -1.076+ | (0.585) | 0.551 | (0.260) | -0.307+ | (0.161) | 0.163 | (0.451) | |
| 상호 작용 | 조사기관 * 응답률 | 1.002 | (0.607) | -0.908* | (0.405) | 0.282+ | (0.167) | -0.655 | (0.468) |
| | 조사 시점 2 * 응답률 | 0.161 | (0.211) | -0.452** | (0.141) | 0.109+ | (0.058) | 0.276+ | (0.163) |
| | 조사 시점 3 * 응답률 | 0.177 | (0.934) | -1.151+ | (0.624) | -0.327 | (0.258) | 1.280+ | (0.721) |
| 모델 F(df1, df2) | F(8, 94)=2.51* | | F(8, 94)=12.0** | | F(8, 94)=7.67** | | F(8, 94)=13.1** | | |
| Adj R ² | 0.106 | | 0.464 | | 0.343 | | 0.489 | | |

주1: OLS 회귀분석 결과(Jamovi project 2023).

주2: 회귀계수 추정값(B): + (90% 신뢰수준), * (95% 신뢰수준), ** (99% 신뢰수준)에서 통계적으로 유의한 차이

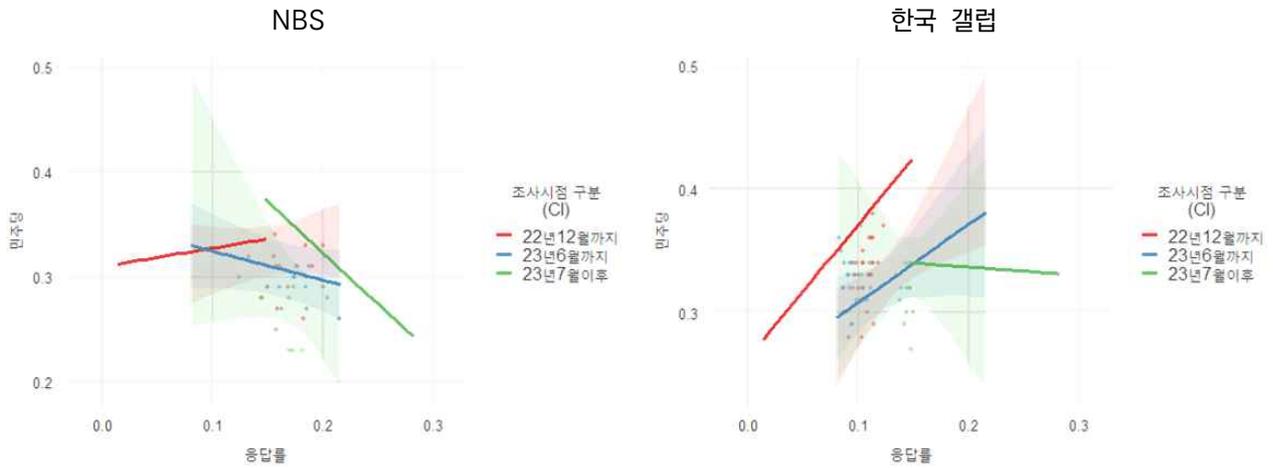
[그림17]은 위의 회귀분석을 통해 도출한 회귀 방정식에 따라 응답률이 조사 시점별(붉은색=조사 시점 1, 파란색=조사 시점 2, 녹색=조사 시점 3) 더불어민주당 지지율에 미치는 영향(직선의 기울기)이 달라지고 있음(상호작용(interaction)이 발생하고 있음)을 직관적으로 보여준다. NBS는 2023년 6월까지의 응답률의 변동이 민주당 지지율에 영향을 미치지 못했지만, 23년 7월 이후에는 응답률의 증가가 민주당 지지율 하락으로 귀결(우하향 그래프)된다. 반면, 갤럽은 반대로 2023년 6월까지의 응답률이 높아질수록 민주당 지지율이 상승함으로써 NBS와는 상반된 효과를 미치고 있다. 그러나 23년 7월 이후에는 갤럽의 경우 응답률의 증가가 민주당 지지율에 미치는 효과는 유의미하지 않았다.

[그림18]는 조사 시점별로 응답률이 무당파 비율에 미치는 영향을 보면, NBS는 조사 시점에 따라 22년 하반기까지는 응답률이 높아지면 무당파 비율이 오히려 낮아지는 경향을 보이다 23년 7월 이후 조사에서는 반대로 응답률의 상승이 무당파 비율의 상승으로 이어지는 패턴이 뚜렷해졌다. 갤럽은 대체로 23년 6월 이전까지는 응답률의 변화가 무당파 비율에 큰 영향을 미치지 못했지만, 23년 7월 이후 조사 시점 3 국면에서는 NBS와 마찬가지로 무당파 비율의 상승으로 이어지는 패턴이 뚜렷해졌다(우상향 그래프).

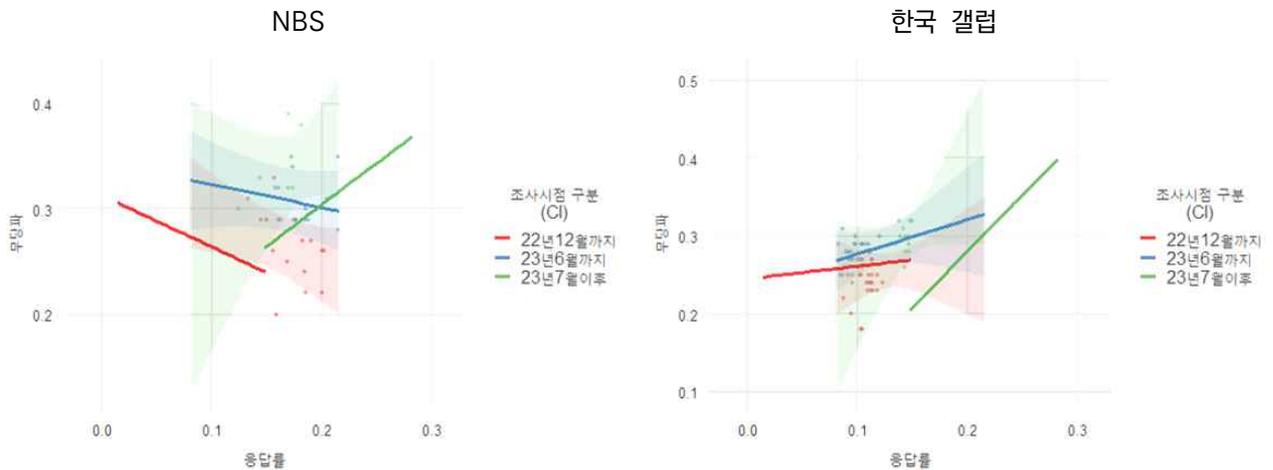
이상의 내용을 종합해보면 응답률이 조사 결과에 미치는 영향은 생각보다 복잡한 요인들이, 복잡한 상호작용을 통해 발생한다고 볼 수 있다. 이는 일반적으로 논의되고 있는 여러 진단이 실제 데이터 분석을 거치지 않고 지나치게 단순화하면서 좀 더 심층적으로 분석해야 할 과제들을 놓치게 할 수 있음을 시사한다. 특히 본 보고서에서 NBS와 갤럽 조사간 차이뿐 아니라 각 기관의 조사 내에서도 조사 시점에 따라 상이한 영향력이 나타날 수 있다는 점을 고려하면, 데이터 분석에 기반한 심층적, 실증적 접근 노력이 필요해 보인다.

26) 데이터 간 자기상관(auto-correlation) 가능성이 있고, 특히 제3국면 하반기의 변동(민주당 지지율 회복, 무당파 비율 하락 조짐)을 보려면 좀 더 세밀한 시계열 분석이 타당하겠으나 여기서는 다른 주요 변수를 통제할 때 응답률의 효과 유무를 간략히 파악하기 위해 간략한 선형회귀분석을 통해 응답률 효과를 검증해보았다. 조사기관, 조사기간, 조사 방법은 이분형 범주(dummy variable)로 전환하여 비교했고, 괄호 안이 준거집단(reference group)이다.

[그림17] 민주당 지지율에 미치는 조사기관과 조사 시점의 상호작용 효과(N=103)



[그림18] 무당파 비율에 미치는 조사기관과 조사 시점의 상호작용 효과(N=103)



그 외 응답률에 영향을 주는 요인 : 문항 길이(survey length) 효과

사실 조사 응답률(혹은 거절률/중도 탈락률)에 영향을 줄 수 있는 요인은 무수히 많다. 조사 방법(survey mode), 재접촉 프로세스, 자료 수집 기간, 면접원 업무량, 조사기관, 의뢰기관, 인센티브 유무, 면접원 숙련도 등 조사기관의 역량과 프로세스와 관련된 요인과 함께 응답자가 조사과정에서 느끼는 응답 부담(질문의 난이도나 심리적, 정치적 민감성)도 응답률 및 거절률에 영향을 미칠 수 있다. 앞서 NBS 조사와 갤럽 정기조사 간 응답률에 영향을 줄 수 있는 요인으로 문항 길이(길수록) 및 조사 기간(짧을수록), 문항 순서 효과(예민한 문항이 앞에 배치될수록) 등은 응답 부담을 높여 응답률을 낮추는 요인들로 알려져 있다. 즉 응답의 인지적, 심리적 부담이 커질수록 중/저 관여층의 무응답률/중단율을 높일 수 있기 때문이다(Groves et al. 2004, 291-2; Peytchev and Peytcheva 2017).

NBS 조사나 한국갤럽의 정기조사는 주기적으로 가장 정확한 조사 결과를 제공한다는 취지로 조사를 진행하기 때문에 다른 언론사나 정당의 기획조사들에 비해 문항 수 관리(총 18~20문항 이내)나 조사일 수(3일), 재접촉(5회) 횟수 등을 엄격하게 유지한다. [표7]을 보면 2023년 8월~10월 2주까지 NBS 및 갤럽조사의 총 문항 수 평균이 각각 19.6개, 17.6개 수준에서 크게 벗어나지 않는다.²⁷⁾ 앞서 살펴본 추석 특집 언론사 의뢰

27) 이들 기관의 여론조사 총 문항 수는 (1) 지역·성·연령 비례 할당을 위한 할당 문항 (2) 본 질문 문항 (3) 분석을 위한 배경 문항의 수 합으로 구성된다. 한 가지 오해하지 말아야 할 것은 전화면접조사가 대체로 15~25문항 내외의 질문길이를 유지하는 반면 ARS 조사는 최대 7~9문항 수준에 그친다. 문항 수로 보면 ARS 조사가 가장 응답률이 높아야 하

조사를 보면 한국리서치·KBS 조사는 문항수 23개, 코리아리서치·MBC 조사는 20개, 엠브레인·YTN 조사는 문항 수는 16개로 문항 수의 차이가 난다. 조사 기간에서도 기관마다 차이가 있었다.

앞 장의 [표4]에서 정리한 것처럼 NBS-갤럽 기관의 문항 순서(국정 평가-정당 지지-총선 지지 순), 문항 길이에서는 큰 차이가 없고, 조사 기간은 3일씩으로 동일하다. 그러나 별도 의뢰를 받아 진행되는 조사(언론이나 정당 등)의 경우 문항 길이나, 조사 기간 등에 작지 않은 차이가 발견된다. 문항 순서 효과나 조사 기간 차이에 대한 검토는 본 보고서의 범위를 넘어서지만 문항 길이 효과에 대해서는 짚고 넘어갈 필요가 있다. 이들 조사들의 문항 길이를 보면 길이가 긴 NBS 조사가 오히려 응답률이 높고, 문항 수 적은 조사에서 낮은 경향도 나타나 문항 길이 효과는 무의미해 보이기도 한다.

[표7] 응답률과 문항 수 및 조사 기간 차이: NBS/한국갤럽(23년 7월 이후 평균), 추석 특집 조사 3개 사 비교

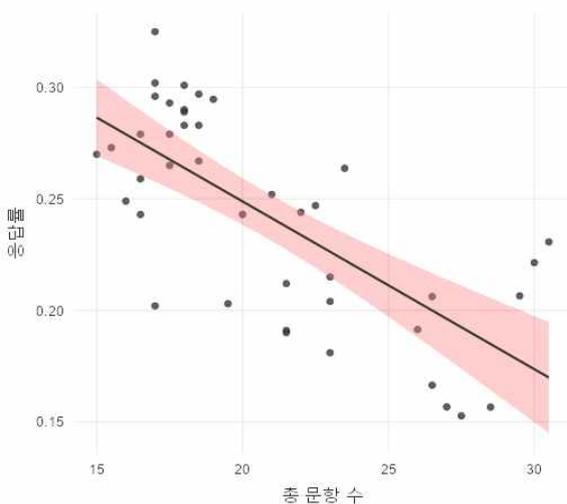
| 조사기관 | 조사 시점 | 응답률 | 본 문항수 | 할당/배경 | 총 문항 수 | 조사 기간 |
|----------|---------------|------|-------|-------|--------|-------|
| NBS(평균) | 23년 8월~10월 2주 | 16.6 | 12.3개 | 7.3개 | 19.6개 | 3일 |
| 한국갤럽(평균) | 23년 8월~10월 2주 | 14.3 | 11.6개 | 6.0개 | 17.6개 | 3일 |
| 한국리서치 | 23년 9월 25~27일 | 13.8 | 16개 | 7개 | 23개 | 3일 |
| 코리아리서치 | 23년 9월 25~26일 | 12.4 | 14개 | 6개 | 20개 | 2일 |
| 엠브레인 | 23년 9월 26~27일 | 11.4 | 11개 | 5개 | 16개 | 2일 |

자료: 중앙선거여론조사심의위원회(여론조사 등록자료)

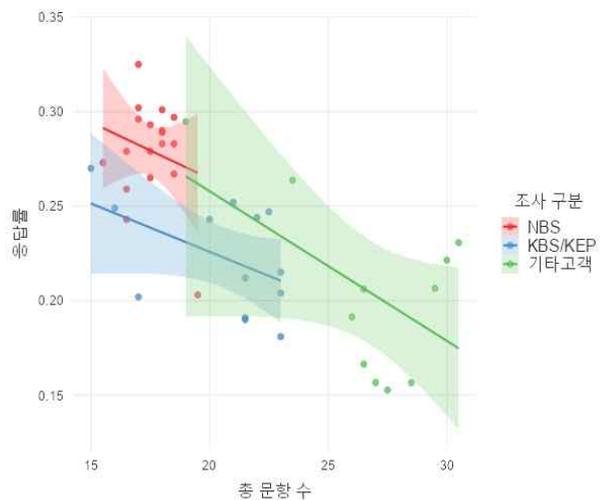
그러나 한국리서치가 지난 2022년 대선 과정에서 실시했던 조사들을 보면 동일한 조사기관에서 동일한 방법(NBS 조사방식, 동일한 조사 시간)을 적용한 조사들(NBS 조사 및 언론/정당 의뢰 조사)의 문항 길이는 확실히 응답률(협조율 기준)과 정당 지지율과 연관을 갖는다. [그림19]에서 총 문항수(본 문항+할당/배경 변인)가 18~20문항을 넘어가면서 응답률 하락 폭이 커지고, 특히 25문항 이상을 넘어가면 응답률이 급락한다(그림19-(1)). 정기조사인 NBS 조사가 대체로 16~19문항 내외, KBS/KEP(방송 3사 선거방송단) 조사의 경우 15~23문항, 정당이나 단체 등 그 밖의 의뢰기관 조사의 경우 19~31문항으로 나타났는데 한국리서치 조사의 경우 문항 길이는 응답률 하락으로 귀결된다는 점은 분명해 보인다(그림19-(2)).

[그림19] 2021~2022년 한국리서치 대선 전화면접조사의 문항 수와 응답률 간의 관계

(1) 전체 문항 수와 응답률 (N=42)



(2) 조사분류별 문항 수와 응답률(N=42)



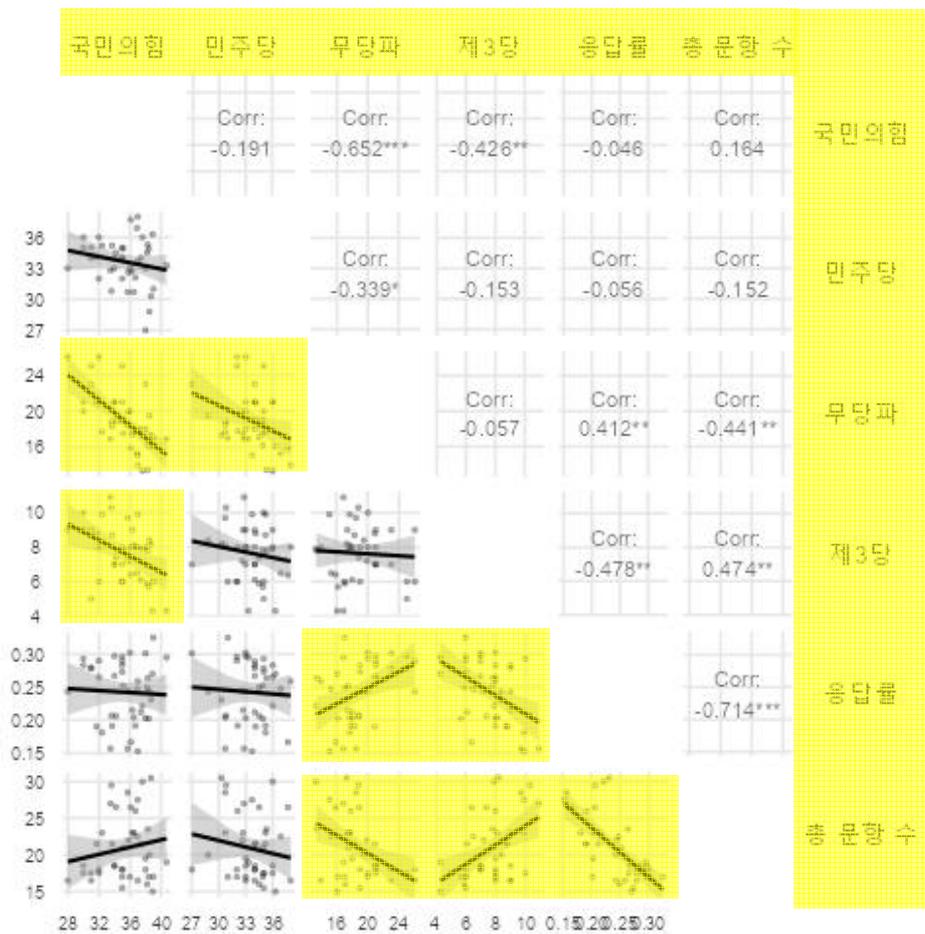
주1: 회귀선은 OLS Linear Regression, 그래프는 JAMOVI program 활용함(Jamovi project 2023).
 자료: 한국리서치 전화면접 선거조사 DB(2021.10~2022.5, N=42)

지만, 자기 응답식 조사의 특성상 중도 탈락 비율이 매우 높아 문항 수를 줄여도 응답률이 높아지지 않는다.

실제로 [그림20]에서 총 문항 수와 응답률의 상관계수는 -0.714로 매우 강한 음의 상관관계가 확인되었고, 이러한 응답률의 변동은 특히 무당파와 제3정당 지지율과 강한 상관관계가 확인되었다. 다만 [그림14]의 현시점과 비교해보면 현재는 높은 응답률이 무당파와 제3당 지지율의 상승과 연결($r > 0$) 되는 데 반해 2022년 대선 당시에는 높은 응답률 조사는 높은 무당파 비율과 연계($r > 0$) 되지만, 제3당 지지율과는 음의 관계($r < 0$)가 나타난다는 점에서 차이가 난다는 점은 차이가 나는 대목이다.²⁸⁾

결국 문항 길이나 조사방식 등의 요인이 정당 태도나 무당파 비율과 상관관계를 보이기는 하지만, 현재 나타나는 응답률과 민주당 지지율 간의 반비례 현상이나, 무당파 비율과 제3정당 비율 변화의 동조화 현상은 정치적, 시기별 상황과 맥락에 따라 달라질 수 있다는 점을 주목할 필요가 있다. 동일 조사 방법을 쓰는 기관 간의 결과 차이, 동일한 조사기관이라도 차이 나는 결과가 발생하는 원인을 탐색하는 데 조사 시점별 분석이 필요한 이유라 할 수 있다.²⁹⁾

[그림20] 2021~2022년 한국리서치 대선 전화면접 조사의 문항 수와 응답률 간의 관계(N=42)



주1: 회귀선은 OLS Linear Regression, 그래프는 JAMOVl program 활용함(Jamovi project 2023).

주2: + (90% 신뢰수준), * (95% 신뢰수준), ** (99% 신뢰수준), *** (99.9% 신뢰수준)에서 통계적으로 유의한 차이, 노란색 셀

28) 그러나 현재는 당시보다 응답률이 큰 폭으로 하락하고 중도 탈락률이 높아지는 상황이라는 점을 고려할 때 문항 수에 대한 보다 보수적인 접근이 필요할 것으로 보인다.

29) 물론 NBS와 한국갤럽의 정기조사는 18~20개 정도의 문항 수를 유지하며 큰 차이를 보이지 않기 때문에 양 조사 간의 차이를 만들어내는 요인으로 보기는 힘들다. 그러나 정기조사 외 언론사나 정당 기획조사 등 많은 문항 수에 대한 조사 수요가 높은 조사들에 대해 문항 수에 따른 응답률 저하 현상이 뚜렷하게 나타나기 때문에 의뢰기관 및 조사기관에서 문항 수에 대한 철저한 관리와 제한이 필요할 것으로 보인다.

② 총선 인식 문항에서 나타나는 워딩 효과에 대한 시론적 탐색

워딩 효과의 가능성: 정당 지지율 차이가 사라진다면 총선 인식 문항 격차는 사라질까?

앞서 살펴본 대로 NBS·한국갤럽 정당 지지율 격차는 유의했고 조사기관의 차이가 조사 시점별 상황에 따라 차이를 보였지만, 최근 양 기관 사이의 조사 방법 일치 후 양 기관의 더불어민주당 지지율과 무당파/소수 정당 지지율 간 격차는 좁혀질 조짐을 보인다. 그렇다면 현 정부 시기 조사에서 정당 지지율보다 더 큰 격차를 보여온 총선 인식 문항도 수렴하는 경향으로 이어질까? 총선 인식 문항에 대한 태도를 결정하는 중요한 예측/설명 지표가 ‘정당지지율’이기 때문에 만약 양 기관의 정당지지율 분포에서 나타난 차이가 좁혀지면 총선 구도 문항에서의 차이도 상당 부분 감소할 것은 분명하다. 그러나 일각에서 제기되듯 ‘워딩 효과(wording effect)’나 그에 따른 질문의 ‘프레이밍 효과(framing effect)’나 ‘질문 순서 효과(question order effect)’ 같은 요인이 영향을 미친다면 정당 성향 분포에서의 차이가 좁혀진다고 해도 양 기관의 총선 인식 문항에서 나타나는 체계적인 차이는 남아 있을 것이다.

‘워딩 효과’란 조사 질문이나 응답 척도에 사용된 “단어”나 “문구” 등의 추가, 부가의 영향으로 (1) 질문의 차원을 확장 혹은 축소하거나(가령 000 대통령의 복지정책에 대한 찬반을 묻는 것과 복지정책 내용에 관한 질문 간 워딩 차이)³⁰⁾ (2) 특정 의제와 이슈의 중요성을 부각(priming issue salience)하거나 (3) 질문이 담고 있는 다양한 프레임 중 특정 “프레임”을 활성화(framing)하거나 (4) 응답자 특유의 “특성”과 “응답 스타일”을 활성화해 상이한 응답을 유발하거나 (5) 인지적, 정서적 부담을 일으켜 사회적으로 바람직한 답변 socially desirable answers)이나 무응답(Non response bias)을 유발하면서 발생하는 현상을 말한다(Goidel et al. 1997; Hughes and Jones 2019; Schrieshem et al. 2010; Zaller 1992, 33-34). 조사 결과 분포 차이를 유발하는 워딩 효과를 “일관되고 안정적인 현상(stable attributes and response style)”인지, 특성의 맥락적 조건과 상호작용 하에서 나타나는 “일시적(temporal or ephemeral) 현상”인지에 대해서는 논쟁이 있어 왔다. 본 보고서는 응답자들의 태도 강도에 따른 현상으로 이해한다는 점에서 후자의 관점을 차용한다(Ponce et al. 2023).

NBS와 갤럽의 총선 인식 문항 워딩 차이

사실 선거의 의미를 생각하거나 자신의 투표 선택을 파악하기 위해 많이 사용하는 문항이 ‘총선 구도 인식’이다. 대체로 ‘정권 안정/지원론’ 대 ‘정권 심판/견제론’에 대한 태도를 측정하는 방식과 워딩은 제각각이었다. 일부에서는 “이번 선거에서 정부·여당(혹은 000 대통령)을 심판해야 한다”, “이번 선거에서 정부를 심판하기 위해 야당을 지지해야 한다” 등의 진술을 주고 이에 대한 동의 여부를 측정하는 ‘정권심판론’에 대한 찬반 질문 방식을 사용하기도 한다. 현재 NBS나 한국갤럽이 사용하는 방식처럼 ‘안정론/지원론’ 진술과 ‘심판론/견제론’ 진술을 주고 택일하게 하는 문항 방식이 있는데, 두 문항 방식 간에는 워딩효과의 측면에서 차이가 있다. 특정 진술을 주고 찬반을 물으면 워딩 효과가 ‘긍정형/부정형’인가에 따라 다른 대답이 나타나고, 선호 강도가 미약한 사람들은 대체로 해당 진술을 ‘승인’하는 워딩 효과가 나타나기 때문에 진술에 대한 찬반보다는 양극단의 입장을 주고 택일하는 후자의 방식이 좋은 것으로 알려져 있다(Groves et al. 2004). 그러나 양극단 택일형의 경우 양쪽의 주장을 치우침 없이 균형 있게 진술을 만드는 것이 쉽지 않은 작업이며, 반대로 지나치게 워딩의 균형을 맞추다 보면 현실의 대립 구도나 쟁점의 맥락을 놓치게 되거나 특히 다차원적인 질문이 포함되는 경우 상충적/양면적 태도 유형을 배제하는 문제가 발생한다(Hughes and Jones 2019).

현재 NBS와 한국갤럽에서 사용하는 총선 문항의 구체적인 질문 워딩과 응답 워딩은 [표8]과 같다. 갤럽 조사는 22년 12월부터, NBS 조사는 23년 5월부터 제22대 총선 인식 문항을 사용하는데, 두 문항 모두 총선에서 소위 정부·여당 중간평가에 대한 태도 문항으로서 일부 워딩 차이에도 불구하고 1번 워딩이 소위 ‘안정

30) “000 대통령의 *** 정책”에 대한 질문은 단서 없는 “*** 정책”에 대한 질문 대비 (1) 000 대통령에 대한 태도 (2) *** 정책에 대한 태도가 혼재(double barreled question)되면서 응답의 방향성과 빈도에 차이를 유발한다(Zaller 1992).

론' 2번 항목이 '견제론/심판론'으로 측정하는 문항으로 인식되었던 것으로 볼 수 있다. 사실 초기에 발표되었을 때는 워딩 차이를 특별히 우려하는 목소리는 거의 찾아볼 수 없었고, 이후 일관되게 양 기관의 차이가 발견되면서 역으로 워딩 효과의 가능성에 주목하는 목소리가 뒤늦게 제기되었다는 점이 이를 방증한다.

[표8] NBS와 한국갤럽의 제22대 총선 인식조사의 질문 워딩 차이

| A그룹 : NBS (22대 총선 인식 구도) | B그룹: 갤럽(제22대 총선 결과 기대) |
|---|--|
| 내년 국회의원 선거에 대한 다음 두 가지 주장 중에 어느 주장에 조금이라도 더 공감하십니까? | 내년 국회의원 선거와 관련하여 다음 중 어느 주장에 더 동의하십니까? |
| 1. 국정운영을 더 잘하도록 정부와 여당에 힘을 실어줘야 한다. | 1. 현 정부를 지원하기 위해 여당 후보가 많이 당선돼야 한다 |
| 2. 정부여당을 견제할 수 있도록 야당에 힘을 실어줘야 한다 | 2. 현 정부를 견제하기 위해 야당 후보가 많이 당선돼야 한다 |
| 9. 응답 유보 | 9. 응답유보 |

그러나 양 기관의 조사 방법상의 차이가 없어지고, 23년 9월~10월을 거치면서 정당 지지 성향 분포 차이도 좁혀질 조짐을 보이고 있지만, 총선 구도 인식조사 결과에서는 여전히 적지 않은 차이들이 발견되는 상황에서 더 이상 워딩 효과에 대해 관성적이고 가설적인 차원의 해석에만 의존할 수는 없다. 이에 따라 한국리서치에서는 자체 웹 서베이 실험 조사(Web survey experiment)를 통해 워딩 효과 유무에 대한 실증적 검증과 발생 원인에 대해 탐색해보기로 한다.

워딩 효과 비교를 위한 웹 서베이 실험 디자인

사실 필자도 내부적으로 양 조사간 워딩 효과의 가능성에 대해 검토했을 때 특별한 워딩 효과 가능성에 주목하지 않았고, 정당 지지율과 비슷하게 양 기관의 조사 방법(가상번호 대 유무선 RDD) 차이로 이해하고자 했다. 그러나 7월 이후 갤럽의 조사 방법 전환에 주목하면서 7월 1주, 8월 3주 두 차례 양 기관의 총선 인식 질문의 워딩 효과 비교를 위한 웹 실험(web-survey experiment)을 기획하였다. 실험을 통해 (1) 과연 총선 인식조사 결과에 강한 영향을 미칠 정치 성향 분포(정당 지지-무당파 비율, 주관적 이념 성향)를 통제할 조건에서도 워딩 효과가 유의미한 영향을 미치는지 (2) 영향을 미친다면 필자의 가정대로 내적 태도 갈등을 가지고 있고(양면적/상충적 태도), 당파적 태도가 약한 중/저 관여층(정치적 태도로 보면 이념적 중도층/무당파층)에서 워딩 효과가 더 영향을 미칠 것인지를 검증하고자 했다.

실험 검증은 워딩 효과 검증을 위한 일반적인 절차에 따라 한국리서치의 MS 패널(Master-sample)을 대상으로 모바일과 이메일을 통해 지역, 성, 연령 인구비례에 맞게 1,000명을 표집한 후 A, B 두 개 실험 집단으로 "무작위 배정(random assignment)"하여 A그룹에는 NBS의 워딩으로, B그룹에는 한국갤럽의 워딩으로 진행한다. "무작위 배정"을 통한 실험군 배정은 실험응답자들의 할당이 우연에 의해 결정됨으로써 실험 집단 간 대상자들 사이에 인위적 차이나 연구자의 주관적 개입을 막고, 검증하려는 실험변수 외의 요인들이 일으키는 내생적인 교란을 통제하는 절차이다(Morton and Williams 2010). 본 연구에서는 웹 조사 시스템(Computer Aided Web Interview system)의 자동화 프로그램을 통해 표본을 두 A, B 실험 집단에 무작위 배정한 후 총선 인식 문항 워딩만 다르게 한 동일 질문지, 동일 질문 순서로 진행했다.

워딩 효과 검증을 위해서는 A, B 그룹에 배정한 응답자들 사이에 응답에 영향을 줄 공변인 간의 균형(balance of the covariate)을 확인해야 한다. 물론 실험군 배치가 무작위 확률 배정이 이루어지면 응답에 영향을 줄 공변인들의 분포가 균형을 이룰 것으로 가정하지만, 무작위 배정에도 불구하고 실제 실험군과 대조군에 배정된 응답자들 사이에 공변인들의 불균형이 발생할 수 있다.³¹⁾ 양 집단에 배정된 응답자들의 공변인 분포의 차이가 날 경우 두 실험-대조집단의 차이를 실험변수(여기서는 워딩 효과)의 영향으로만 해석하기 어

31) 무작위 배정의 결과이기 때문에 이는 우연한 결과로 볼 수 있다.

려워진다. 좀 더 엄밀한 통제를 위해서 공변량의 균형을 위한 매칭 등의 방법을 고려할 수 있으나 본 연구에서는 집단별 관찰값을 통해 검증하기보다는 제3변인들의 영향을 통제하는 회귀분석을 통해 워딩 효과를 검증한다(이상신 외 2023; Morton and Williams 2010).

총선 인식 응답에 영향을 미칠 공변인으로는 선거 투표행태 결정요인(vote determinants)으로 알려진 (1) 정치적 성향 변수(정당 태도, 주관적 이념 성향) (2) 인구학적 통제 변인(성, 연령, 지역, 소득, 학력) 요인을 설정하고 (3) 실험 처치 변수(워딩 효과, 실험 시기) 변인을 포함하여 A, B 두 그룹 간 총선 인식 응답 차이를 검증해보고자 한다.

웹서베이 실험 배정 결과 : 지지 정당 분포의 불균형

[표9]의 양 기관이 발표한 정당지지율과 주관적 이념 성향 분포를 보면 이념 성향 분포에서의 차이는 무시할 만하나, 정당지지율 분포에서는 NBS와 한국갤럽 조사 결과 차이(NBS 조사 비율-갤럽조사 비율)가 두드러진다. 더불어민주당 지지율은 7월 1주에 NBS 조사와 갤럽조사 격차가 -4%p, 8월 3주 조사에서 무려 -8%p나 차이 났다. 무당파 비율의 경우 반대로 NBS가 7월 1주에 +2%p, 8월 3주에는 +8%p로 특히 8월 3주 시점에는 NBS보다 갤럽 조사가 민주당 지지율이 강세를 보여, 총선 인식 문항에서 NBS보다 정권 견제론-야당 승리 기대 입장이 강세를 보이는 현상에는 실험변수인 워딩 효과 외에 양 조사의 응답자 성향 분포 영향이 작동해왔음을 확인할 수 있는 결과다.

한편 웹 조사 실험에 표집 된 응답자를 보면 7월 1주, 8월 3주 두 차례 실험 조사에 포함된 응답자의 경우 동일 시점의 NBS, 갤럽 조사 대비 더불어민주당 지지율이 자가 과대 대표되고, 국민의힘 지지층과 무당파는 과소 대표되어 있음을 알 수 있다. 다만 주관적 이념 성향으로 보면 큰 차이는 없었다. 이는 웹 패널조사 방법(survey modes)의 일반적인 현상으로서 웹 자발적 참여 패널은 한국리서치 패널뿐 아니라 국내외의 모두 진보 성향 패널 응답자들이 과대 대표된다(Mercer et al. 2018; 조성겸·오승호 2021). 따라서 현재 웹 패널 조사 방법은 정확한 여론분포 추정과 변화를 추적하는 선거여론조사로는 활용하는 데 한계가 분명하지만, 본 연구처럼 실험 집단 무작위 배정과 실험 조작에 유용하기 때문에 서베이 실험에 많이 활용되고 있다.

[표9] 웹 실험 시점의 NBS-한국갤럽 조사의 정당 지지율과 주관적 이념 정체성 비율 차이

(단위 : %, %p)

| | NBS | | 한국갤럽 | | NBS-한국갤럽 | | 한국리서치 웹 실험 조사 | |
|------|-------|-------|-------|-------|----------|-------|---------------|-------|
| | 7월 1주 | 8월 3주 | 7월 1주 | 8월 3주 | 7월 1주 | 8월 3주 | 7월 1주 | 8월 3주 |
| 국민의힘 | 34 | 34 | 33 | 35 | +1 | -1 | 28 | 30 |
| 더민주 | 28 | 23 | 32 | 31 | -4 | -8 | 36 | 37 |
| 기타 | 6 | 9 | 5 | 5 | +1 | +5 | 6 | 8 |
| 무당파 | 32 | 34 | 30 | 27 | +2 | +8 | 30 | 26 |
| 진보 | 25 | 25 | 28 | 27 | -3 | -2 | 29 | 25 |
| 중도 | 33 | 36 | 32 | 34 | +1 | +2 | 35 | 36 |
| 보수 | 30 | 29 | 30 | 32 | 0 | -3 | 33 | 34 |
| 유보 | 12 | 10 | 10 | 10 | +2 | 0 | 3 | 5 |

[표10]의 각 조사 시점별로 A/B 실험 집단에 대한 무작위 배정의 결과를 살펴보자. 배정된 집단별로 (1) 정치 성향 변수(주관적 이념 정체성, 지지 정당) 및 인구학적 변수(성별/연령대/거주지역/학력/월평균 가구소득)들의 구성비(%)를 정리하고, 각각에 분할표에 대한 카이제곱 검정을 통해 각 시점 및 전체 총계에서 A/B 두 집단 간 구성비의 차이 여부를 검정한 결과 지지 정당 분포, 성별에서만 차이가 유의했다(7월 1주, 총계에

서 $p < 0.01$). 나머지 변수들의 경우 무작위 배정에서 기대한 실험 집단 간 공변인들의 동질화 효과가 뚜렷하다.³²⁾

[표10] 웹 실험 실험 집단 무작위 배정 결과 (카이제곱 검정)

(단위 : %)

| 조사 시점 | | 23년 7월 1주 (N=1,000) | | 23년 8월 3주 (N=1,000) | | 통합 총계 (N=2,000) | |
|----------|----------|------------------------|-----------------------|------------------------|-----------------------|-------------------------|------------------------|
| 실험 집단 | | A그룹 NBS 워딩 N=498 | B그룹 갤럽 워딩 N=502 | A그룹 NBS 워딩 N=488 | B그룹 갤럽 워딩 N=512 | A그룹 NBS 워딩 N=1000 | B그룹 갤럽 워딩 N=1000 |
| 사례 수 | | | | | | | |
| 이념 성향 | 진보(0-4) | 29 | 29 | 24 | 25 | 27 | 27 |
| | 중도(5) | 32 | 37 | 36 | 37 | 34 | 37 |
| | 보수(6-10) | 36 | 31 | 35 | 34 | 35 | 33 |
| | 모름 | 3 | 3 | 5 | 4 | 4 | 4 |
| 지지 정당 | 더불어민주당 | 31** | 40** | 35 | 39 | 33** | 40** |
| | 국민의힘 | 29** | 26** | 29 | 30 | 29** | 28** |
| | 기타정당 | 8** | 4** | 8 | 7 | 8** | 6** |
| | 무당파 | 31** | 29** | 28 | 24 | 30** | 27** |
| 성별 | 남자 | 51 | 48 | 52 | 47 | 52+ | 48+ |
| | 여자 | 49 | 52 | 48 | 53 | 48+ | 52+ |
| 연령 | 18-29세 | 18 | 15 | 16 | 16 | 17 | 16 |
| | 30-39세 | 13 | 17 | 13 | 17 | 13 | 17 |
| | 40-49세 | 19 | 17 | 19 | 17 | 19 | 17 |
| | 50-59세 | 19 | 20 | 21 | 18 | 20 | 19 |
| | 60세이상 | 32 | 30 | 31 | 31 | 32 | 31 |
| 거주지역 | 서울 | 17 | 20 | 19 | 19 | 18 | 19 |
| | 인천/경기 | 30 | 33 | 31 | 32 | 31 | 33 |
| | 대전/세종/충청 | 12 | 9 | 11 | 10 | 11 | 10 |
| | 광주/전라 | 9 | 10 | 10 | 10 | 9 | 10 |
| | 대구/경북 | 9 | 10 | 9 | 10 | 9 | 10 |
| | 부산/울산/경남 | 17 | 13 | 17 | 14 | 17 | 13 |
| | 강원/제주 | 5 | 4 | 3 | 5 | 4 | 5 |
| 학력 | 고졸이하 | 53 | 52 | 47 | 46 | 50 | 49 |
| | 대재이상 | 47 | 48 | 53 | 54 | 50 | 51 |
| 월평균 가구소득 | 200만원미만 | 13 | 13 | 14 | 15 | 14 | 14 |
| | 300만원 미만 | 20 | 13 | 14 | 15 | 17 | 14 |
| | 400만원 미만 | 15 | 17 | 17 | 15 | 16 | 16 |
| | 500만원 미만 | 14 | 16 | 12 | 12 | 13 | 14 |
| | 600만원 미만 | 12 | 13 | 14 | 13 | 13 | 13 |
| | 700만원 미만 | 7 | 7 | 9 | 10 | 8 | 8 |
| | 700만원 이상 | 19 | 21 | 21 | 19 | 20 | 20 |

자료: 한국리서치 웹 실험 조사 데이터(7~8월 통합)

주1: 카이제곱검정 + (90% 신뢰수준), * (95% 신뢰수준), ** (99% 신뢰수준), *** (99.9% 신뢰수준)에서 통계적으로 유의한 차이

결국 양 집단의 총선 인식 문항에 대한 응답 차이가 발생할 때 이 차이가 순수 워딩 효과가 아닌 양 실험 집단의 지지 정당 구성비 차이가 혼재된 결과일 수 있음을 의미한다.³³⁾ 흥미로운 점은 무작위 배정임에도 A그룹과 B그룹의 정당 지지 구성비가 실제 NBS와 갤럽 조사와 동일한 패턴으로 차이가 나타난다는 점이다. 국민의힘 지지율에서는 양 기관 차이가 없고, NBS 조사에서 민주당 지지가 상대적으로 낮고 무당파/기타정당 비율이 높은 반면, 갤럽조사는 민주당 지지가 높고, 무당파/기타정당 비율이 낮았던 패턴과 동일하게 A그룹의 더불어민주당 지지율은 B그룹보다 낮고, 무당파/제3당 지지율은 높은 경향성이 일치한다.³⁴⁾ 무작위 배정의

32) 7~8월 통합자료에서 성별 차이는 90% 신뢰수준에서 유의했지만 남녀 비율 보면 A 집단 52:48, B 집단은 48:52로 전체 유권자의 성비 구성비로 볼 때 유의미한 편향으로 보기 힘들다.

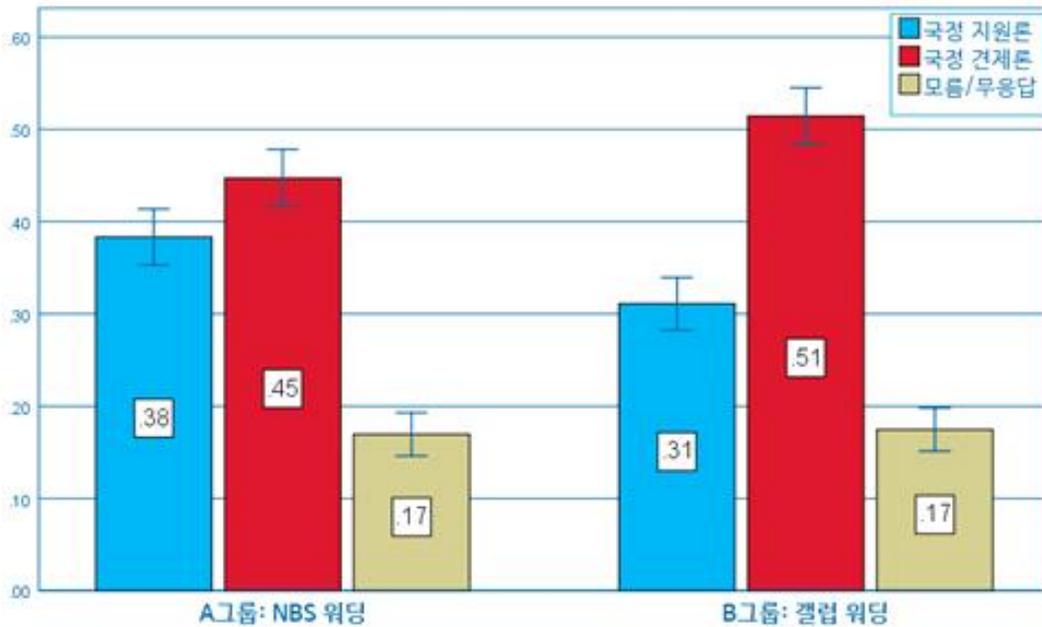
33) 결국 내생성(endogeneity) 문제를 해결하기 위한 실험에서 발생하는 내생성 문제인 셈이다(Morton and Williams).

결과이기 때문에 우연에 의한 결과라는 점에서 흥미롭다. 결과적으로 무작위 배정에도 불구하고 A/B 실험 집단에서 핵심 공변인(지지 정당)의 균형이 이루어지지 못한 상황이라 두 집단 응답 결과를 실험 처치(위딩 효과)의 결과로 해석할 수 없으며 지지 정당 분포의 불균형에 따른 교란 효과(confounding effect), 즉 '내생성(endogeneity)'의 문제가 발생할 수 있음을 의미한다. 따라서 양 집단의 차이에서 위딩 효과를 살펴보기 위해 양 집단 간 총선 구도 결과의 차이에 대한 관찰값의 비교와 함께, 정치 성향, 인구학적 공변인들의 영향력을 통제한 다항로지스틱 회귀분석 결과를 함께 분석하여야 한다.

지지 정당 통제 후에도 A/B 그룹 간 위딩효과 유의할까?

7~8월 통합자료의 A그룹(NBS 위딩), B그룹(갤럽 위딩)으로 측정한 총선 인식 구도 결과를 보면 NBS 위딩으로는 국정 안정론과 견제론의 차이가 -7%p(안정론 38%, 견제론 45%)인 반면, 갤럽 위딩에선 -20%p(안정론 31%, 견제론 51%)가 격차가 크다. 실제 NBS 조사에서 안정론이, 갤럽조사에서 견제론이 상대적으로 많은 경향과 일치한다. 양 집단에서 안정론과 견제론의 비율 격차(대응 표본 T 검정 결과)는 통계적으로 유의했다. 그러나 실제 NBS 조사에서는 10월 이전까지 조사에서 갤럽조사와 달리 안정론과 견제론의 비율 차이가 유의하지 않은 데 반해 본 웹 조사 실험에서 NBS 위딩을 사용한 그룹에서도 안정론과 견제론 차이가 유의했다. 이는 웹 패널조사의 정치 성향의 편향이 반영된 결과이다. 웹 패널조사는 동일한 패널에서 무작위로 배정된 집단들에서 위딩 간 차이 여부를 확인하는 자료로 사용해야 한다. 웹 패널조사를 활용하여 전체 여론의 분포를 추정하려면 별도의 가중치를 적용하거나, 확률 기반의 웹 조사 결과를 활용해야 할 것이다.

[그림20] 7~8월 통합 총선 인식 결과 비율 차이(지원론-견제론) 검증 : A그룹**, B그룹**



자료 한국리서치 웹 실험 조사 데이터(7~8월 통합)

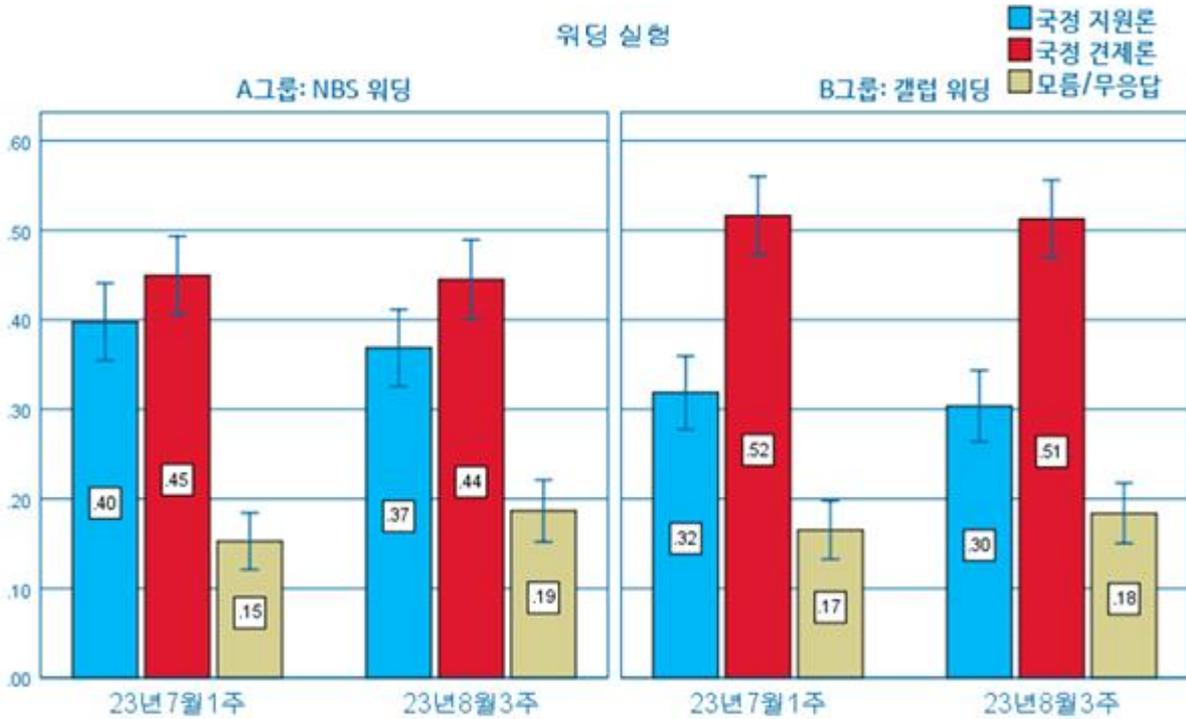
주: 대응 표본 T-test: 카이제곱검정 * (95% 신뢰수준), ** (99% 신뢰수준), *** (99.9% 신뢰수준)에서 통계적으로 유의한 차이

조사 시점별로 보면 A그룹의 경우 7월 조사에서는 안정론과 견제론의 차이가 신뢰구간을 통계적으로 유의하지 않았지만(p>0.05), 8월 조사에서는 그 격차가 커져 통계적으로 유의한 차이(p<0.05*)로 나타났다. 반면, 갤럽의 위딩을 사용하는 B그룹에서는 7월과 8월 모두 견제론이 크게 우세한 결과가 나왔고, 각각 대응 표본 T 검정 결과 99% 신뢰수준에서 통계적으로 유의한 차이였다(p<0.01**). 종합해보면 조사 시점에 따라

34) 이는 통계적으로 유의한 차이를 보인 7월 1차 실험뿐 아니라 8월 2차 실험에서도 통계적으로 유의한 차이가 아니었을 뿐 패턴은 동일하여 7~8월 통합자료에서도 실험 집단 간 지지 정당 분포의 불균형이 유지된 요인이다.

NBS 워딩을 사용한 A그룹에서 상대적으로 안정론 비율이 높게 나와 격차가 감소하는 경향이 나타났지만, 갤럽조사 워딩을 사용한 B그룹에서는 압도적으로 견제론 비율이 우세한 현상이 반복된 셈이다. 두 집단의 총선 인식 결과를 보면 조사 워딩에 따라 상이한 응답 분포가 도출될 수 있음을 시사해주는 결과다.

[그림21] 7월과 8월 개별 실험: 총선 인식 결과 비율 차이(지원론-견제론) 검증 : A그룹, B그룹**



자료 한국리서치 웹 실험 조사 데이터(7~8월 통합)

주: 대응 표본 T-test: 카이제곱검정 * (95% 신뢰수준), ** (99% 신뢰수준), *** (99.9% 신뢰수준)에서 통계적으로 유의한 차이

정치성향을 통제해도 워딩 효과가 작동하고 있다

그러나 이렇게 일관되게 나타나는 A그룹과 B그룹의 총선 인식 분포 차이를 실제 '워딩 효과'로 볼 수 있는지는 추가 검증이 필요하다. 무엇보다 A, B 양 그룹 간 총선 인식 결과의 차이에 영향을 미칠 다른 제3의 요인들을 통제한 조건에서도 워딩 효과가 유의미한지 검증해야 한다. 종속변수의 범주가 2개일 경우는 보통 로지스틱 회귀분석을 통해 검증하나 본 조사처럼 종속변수인 총선 인식 문항이 세 개의 범주 척도(1. 국정 지원론, 2. 국정 견제론, 9. 응답 유보)로 구성된 비 연속형 범주인 경우 기준 범주 대비 각각의 항목과 짝을 지어 이분형 로지스틱 회귀분석의 쌍을 분석하는 '다항로지스틱 회귀분석'(multiple logistic regression test)을 통해 설명 변수의 영향력을 검증한다. 조사 데이터는 7~8월 통합 데이터이다.

본 모델에는 실험변수를 워딩 효과 문항 (1=A그룹, 0=B그룹)과 함께 통제해야 할 공변인으로서 (1) 인구학적 변수: 성별(1= 남, 0=여자), 연령 더미(기준 변수 60대 이상=0), 거주지역(기준 변수 강원/제주=0), 월 가구소득(기준 변수=700만원 이상), 학력(1=고졸 이하, 0=대졸 이상) (2) 정치 성향 변수: 이념 성향(기준 변수 응답 유보층=0, 각 범주=1), 지지 정당(기준 변수 무당파=0, 각 범주=1) (3) 실험통제 변수: 조사시점(1=7월 조사, 0=8월 조사)로 코딩하여 다항로지스틱 회귀분석을 수행했다. 분석 결과표가 [표11]이다.

결과를 보면 본 분석 모델을 종속변수의 변이를 설명하는 모델로서 유의(카이제곱=1598.12, df=52 (p =.000***)하며, 종속변수의 변화량의 55.2% 정도를 설명(Cox and Snell's R² = 0.552) 하는 것으로 나타났다. 주요 분석 결과로는 첫째, 분석의 목적인 워딩 효과의 경우 총선 인식에 영향을 줄 수 있는 다른 인구학적 변인과 정치적 변인을 고려하더라도 지원론 대비 견제론의 선호 여부(모델 Y1), 지원론 대비 응답 유보의 선호 여

부(모델 Y2)에 영향을 미치는 변인으로 확인되었다. 모델 Y1의 회귀계수 B =-0.574, p<0.001, 모델 Y2에서도 회귀계수 B =-0.497, p<0.001로 유의한 변수임이 확인되었다. 회귀계수의 부호가 (-)라는 것은 B그룹(겔럽) 워딩 대비 A그룹(NBS) 워딩에 노출되었을 때, 국정 견제론이나 응답 유보 대신 국정 지원론을 강화하는 효과가 있음을 말해준다.

다른 공변인들의 영향을 살펴보면 국정 안정론 대비 견제론 선호에는 연령대(60대 대비 30대/40대/50대에 속할수록 견제론을 강화, 20대는 60대와 유사), 지역 변수(강원 제주 대비 수도권, TK, PK에서 국정 안정론 선호, B<0), 정향 변수로는 역시 이념 성향(진보층은 견제론 선호, 보수층은 안정론 선호)과 지지 정당(무당파 대비 민주당 지지/기타정당 지지자는 견제론 선호, 무당파 대비 국민의힘 지지자가 안정론을 선호)이 모두 총선 인식 구도에 미치는 영향력이 확인되었다. 이는 조사기관이 같은 개별 건의 조사라도 세대/지역, 이념 성향, 정당 지지의 분포가 불균형할 경우(한쪽에 특정 집단이 과대 대표되거나 과소 대표될 경우) 워딩 효과와 무관하게 총선 인식 응답 분포의 차이를 유발할 수 있음을 의미한다.

한편 응답 유보층의 경우 여자, 60대 이하 젊은 세대일수록, 이념 성향이 뚜렷하지 않고, 무당파 성향이 강할수록 총선 인식에 대한 응답을 유보하는 경향이 강함을 실증적으로 보여주었다.

[표11] 다항로지스틱 회귀분석(multiple logistic regression)³⁵⁾

| 총선 구도a N=2,000 | (기준) 국정 지원론(Y0) | 국정 견제론(Y1) | | | 응답 유보(Y2) | | |
|-------------------|----------------------------------|------------|----------|--|-----------|----------|--------|
| | | B | 표준오차(SE) | Exp(B) | B | 표준오차(SE) | Exp(B) |
| 절편 | | 1.359* | (0.617) | | 1.481* | (0.650) | |
| 성별(여자) | 남자 | -0.033 | (0.151) | 0.968 | -0.429* | (0.169) | 0.651 |
| 연령(60대+) | 20대 | 0.329 | (0.251) | 1.389 | 1.502*** | (0.270) | 4.492 |
| | 30대 | 0.429+ | (0.245) | 1.535 | 1.125*** | (0.281) | 3.081 |
| | 40대 | 0.665** | (0.239) | 1.945 | 1.359*** | (0.275) | 3.894 |
| | 50대 | 0.556* | (0.217) | 1.744 | 0.619* | (0.264) | 1.857 |
| 거주지역 (강원/제주) | 서울 | -0.952* | (0.404) | 0.386 | -0.700 | (0.461) | 0.497 |
| | 인천/경기 | -0.642+ | (0.388) | 0.526 | -0.302 | (0.440) | 0.739 |
| | 대전/세종/충청 | -0.470 | (0.436) | 0.625 | -0.039 | (0.490) | 0.962 |
| | 광주/전라 | -0.426 | (0.446) | 0.653 | -0.021 | (0.507) | 0.979 |
| | 대구/경북 | -0.970* | (0.431) | 0.379 | -0.577 | (0.486) | 0.562 |
| | 부산/울산/경남 | -0.740+ | (0.412) | 0.477 | -0.339 | (0.464) | 0.712 |
| 월 가구소득 (700만+) | 200만 미만 | -0.027 | (0.278) | 0.973 | 0.225 | (0.298) | 1.252 |
| | 300만 미만 | -0.223 | (0.262) | 0.800 | -0.015 | (0.281) | 0.985 |
| | 400만 미만 | -0.241 | (0.260) | 0.786 | -0.155 | (0.289) | 0.856 |
| | 500만 미만 | -0.336 | (0.264) | 0.715 | -0.590+ | (0.303) | 0.554 |
| | 600만 미만 | -0.060 | (0.267) | 0.942 | -0.529+ | (0.310) | 0.589 |
| | 700만 미만 | -0.001 | (0.319) | 0.999 | -0.477 | (0.365) | 0.621 |
| 학력(대재+) | 고졸 이하 | -0.170 | (0.176) | 0.844 | 0.102 | (0.196) | 1.108 |
| 이념 성향 (유보층) | 진보층 | 1.230** | (0.437) | 3.422 | -0.046 | (0.432) | 0.955 |
| | 중도층 | 0.084 | (0.404) | 1.087 | -0.442 | (0.376) | 0.643 |
| | 보수층 | -0.634 | (0.411) | 0.531 | -1.121** | (0.388) | 0.326 |
| 지지 정당 (무당파) | 더불어민주당 | 1.757*** | (0.202) | 5.798 | -0.655** | (0.246) | 0.519 |
| | 국민의힘 | -2.452*** | (0.212) | 0.086 | -2.963** | (0.235) | 0.052 |
| | 기타정당 | 0.989*** | (0.276) | 2.688 | -0.874* | (0.362) | 0.417 |
| 조사 시점(8월) | 7월 조사 | -0.231 | (0.151) | 0.794 | -0.491** | (0.169) | 0.612 |
| 워딩실험(B그룹) | A 그룹(NBS) | -0.574*** | (0.152) | 0.563 | -0.497** | (0.169) | 0.609 |
| 모델 적합도 | 카이제곱=1598.12, df=52 (p =.000***) | | | Cox and Snell's R ² = 0.552 | | | |

자료 한국리서치 웹 실험 조사 데이터(7~8월 통합)

주: 유의도: * (95% 신뢰수준), ** (99% 신뢰수준), *** (99.9% 신뢰수준)에서 통계적으로 유의한 차이

35) 계수 B > 0, Exp(B) > 1이면, 국정 지원론 대비 국정 견제(Y1)/국정 지원론 대비 응답 유보(Y2) 선택 확률을 강화시키고, 계수 B < 0, Exp(B) < 1이면 감소함을 의미

워딩 효과를 낳은 워딩 요인은 무엇인가?

남는 문제는 ‘왜 이런 워딩 효과가 발생하는가?’이다. 본 연구에서는 양 기관의 워딩 자체를 유발할 요인에 대한 몇 가지 잠정적인 가설을 제기한 후 필자들은 당파성이 약한 중도/무당파층, 중/저 관여층의 다차원적인 인식과 태도 갈등의 효과일 수 있음을 주장하고자 한다.

우선 양 기관의 질문 및 응답 워딩이 워딩 효과를 만들 만한 몇 가지 상이한 요소를 가설적인 차원에서 살펴보자. 첫째, 질문에서 두 가지 주장에 대한 ‘공감 여부’와 ‘동의 여부’ 워딩이 응답자의 응답에 영향을 줄 가능성을 살펴볼 필요가 있다. ‘공감’과 ‘동의’의 개념을 보면 원칙적으로 전자는 ‘정서적’ 태도를 지칭하고 후자의 동의는 상대적으로 ‘인지적’ 태도를 지칭함으로써 상이한 반응을 이끌 가능성이 있다. 사실 선거 때마다 조사기관의 결과 차이를 설명하면서 ‘적합도’, ‘선호도’, ‘정치적 지지’, ‘인지적 평가’ 중 어느 쪽을 활성화하는가에 따라 응답이 달라질 수 있다는 진단들이 끊이지 않은 점을 생각하면 ‘공감’과 ‘동의’의 워딩 차이가 결과의 차이를 유발할 가능성을 배제할 수는 없을 듯하다.³⁶⁾

둘째, 질문 응답 워딩을 보면 양 기관 모두 “투표 선택의 의미와 목적”에 해당하는 “지원 대 견제”와 실제 투표 선택의 “방향”과 관련한 “친여당 대 친야당”의 두 개의 차원이 혼재된 질문을 통합하여 하나의 질문으로 만든 질문 유형(double-barreled question)이다. 조사 방법론에서 피해야 할 질문 방식의 대표적인 사례로 (1) 두 차원이 일치하지 않는 응답자(정권 안정론을 지지하지만, 여당에 힘을 실어주고 싶지 않은 유권자나 정권심판론에 동의하지만, 야당 후보가 많이 당선되는 것도 꺼리는 유권자)들은 자신의 선택을 배제당하는 문제가 있고 (2) 따라서 집계된 결과에 대한 해석도 어렵다. 차원별로 묻는 방식을 고려해야 한다.

셋째, 투표의 의미에서 NBS는 “국정운영을 (앞으로) 더 잘하도록 vs 정부·여당을 견제할 수 있도록”으로 구분하여 갤럽의 “정부를 지원하기 위해 vs 견제하기 위해”와 의미상 차이가 나지는 않는 듯하지만, 주관적인 의견으로는 문구 대응의 균형이나 서술구조의 안정감을 보면 갤럽의 문구가 더 안정적으로 느껴진다. NBS는 “국정운영”이라는 표현이 정부나 여당보다는 ‘대통령’을 연상시키는 효과가 있을 듯하며 “더 잘하도록”이라는 표현이 상대적으로 미래지향적 기대를 강화하는 워딩 효과를 가져올 수도 있을 듯하다. 문구 대응의 균형으로 보면, 갤럽의 문구가 나아 보이나, NBS의 문구는 현실 캠페인에 더 부합할 수 있다는 설명도 가능하다. 현실 캠페인에서의 프레임 전략 시사점으로 보면 NBS의 문구도 나름의 의미가 있어 보인다.

넷째, 질문 문구 후반에서 NBS는 “정부와 여당 vs 야당” 구도를 명시한 반면, 갤럽의 워딩은 “여당 후보 승리 대 야당 후보 승리” 구도를 명시한 미묘한 차이가 있다. 현재 정부 심판/견제론의 분위기가 확산되어 이에 대한 동의나 공감이 커지고 있는 상황에서 야당에 대해서도 불신이 있는 유권자들을 가정해보면 NBS의 진술(“힘을 실어주어야 한다”)은 갤럽조사의 진술(“다수가 당선돼야 한다”)에 비해 응답자 본인의 선호와 의사가 더 강한 느낌을 주고 동시에 다른 사람들의 행동을 직접적으로 설득하는 워딩으로 들린다. 따라서 견제론의 선호가 강하지 않거나 야당 지지에는 꺼려하는 응답자들이 선뜻 선택하기에는 강한 주장처럼 보인다. 반면 갤럽 조사의 워딩은 상대적으로 선호나 의지가 아닌 본인 개인 차원의 판단과 주장의 성격이라 NBS에 비해 선택하는데 부담이 상대적으로 크지 않을 수 있다.

위에서 언급한 워딩 효과를 유발하는 워딩 요인에 대한 진단은 필자들의 주관적 가설 차원이며, 실제 고려하는 후속 연구의 과제로 남겨둬야 할 것 같다. 종합하면 워딩 효과를 통해 정부와 여당의 입장에서는 대통령의 국정을 미래지향적으로 전환하는 것에 대해서는 상대적으로 넓은 공감대를 얻을 수 있다는 시사점을 얻을 수 있을 것 같다. 이렇게 보면 현재 강화되고 있는 강경한 이념 노선의 강조나 국민의힘보다 정부 부처의 책임자들이 야당의 공세에 전면에 나서는 듯한 이미지가 “국정의 미래지향적 전향”이라는 공감대에 부합하는지에 대해 진지한 검토가 필요할 듯하다. 반대로 더불어민주당의 경우 정부·여당에 대한 적지 않은 견제 심리가 총선에서 야당이 다수의석을 얻는 것이 낫겠다는 데까지는 동의하지만, 선뜻 “야당 후보를 밀어줘야 한다”라는 집단적 행동 의사로 나아가지는 못하는 점에 주목해야 할 듯하다.

36) 다만 기존의 ‘적합도’, ‘선호도’ 등의 워딩 차이의 경우 실증적으로 검증된 적이 없기 때문에 하나의 가설로서만 이해되어야 한다는 점을 분명히 해두고자 한다. 한겨레신문. “선호도·호감도·적합도…미묘하게 다른 질문, 어떤 차이 있을까” (2021/08/19)

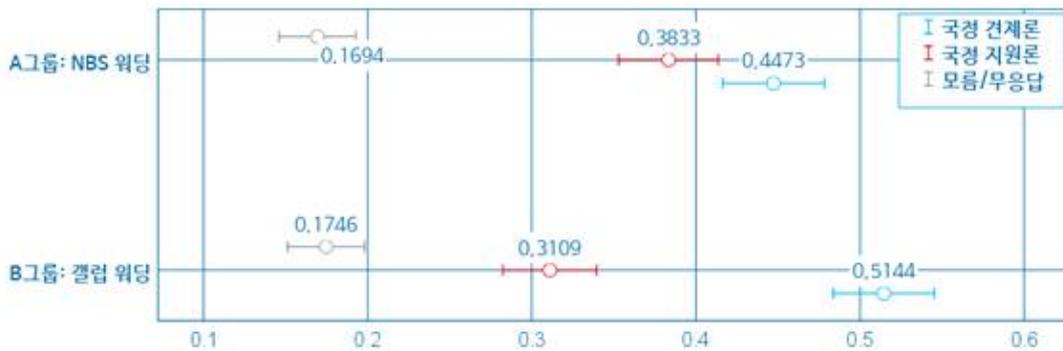
모든 차이를 워딩 효과로 환원해서는 안되며, 그 차이도 향후 줄어들 듯

현재 NBS/갤럽 간 총선 인식조사 결과 차이에 양 기관의 워딩 효과가 작동하고 있다는 것은 매우 흥미롭기도 하고, 주목할 만한 현상이다. 그러나 양 기관의 차이를 워딩 효과로는 환원하지는 말아야 한다. 한국 사회에서 여론조사가 소비될 때 일시적, 특정 시점과 특정 환경에서 나타나는 현상을 무리하게 일반화하거나 공식화하면서 추가적인 심층 검증이나 새로운 변화 요인에 대한 탐색을 억제하는 경향이 있기 때문이다.

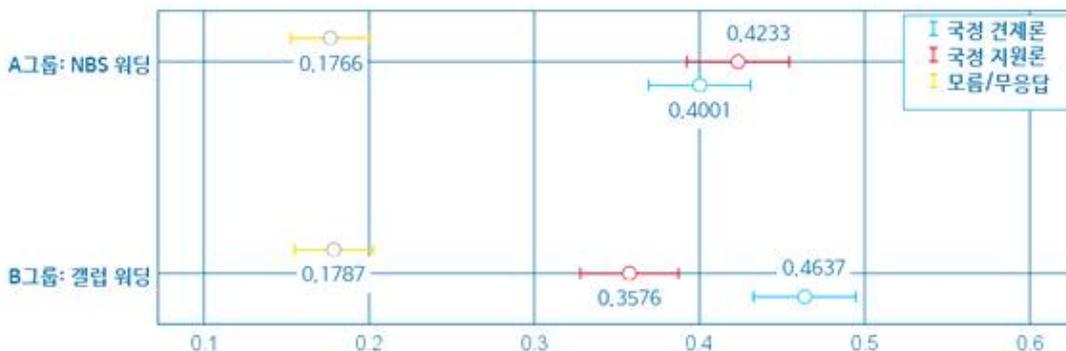
당장 본 웹 조사과정에서 확인된 NBS 워딩 조사 집단과 갤럽 워딩 조사 집단 간 차이에는 워딩 효과도 작동했지만, NBS/갤럽 간의 응답자 정치 성향 차이가 존재했고, 실험 연구에서도 우연히 개입된 A그룹, B그룹의 지지 정당 분포 차이가 워딩 효과와 결합한 결과이기도 하다. 그 효과를 직관적으로 확인해보기 위해 실험 조사가 있었던 7월 1주, 8월 3주 시점의 NBS와 갤럽조사의 정당지지율을 평균한 수치를 기준으로 가중치를 부여하여 본 실험 조사의 응답자 분포를 보정하면 [그림22]처럼 보정 전에 비해 견제론과 지원론 간의 격차는 절반 이하 수준으로 줄어들 것이라는 것을 알 수 있다. 조사 결과에서 나타나는 편향이나 차이의 문제는 여러 요인이 복합적으로 작용한 결과이며, 조사 시점이나 환경에 따라 상이한 결과를 만들어낸다.

앞에서 살펴본 것처럼 NBS와 갤럽의 정치성향 분포의 차이가 주로 갤럽이 2023년 7월 조사 이전까지 유무선 RDD방법을 사용했던 시기에 집중되었지만, 이후 갤럽의 응답률이 상승하고, 응답률의 영향도 같은 방향으로 수렴해가고 있기 때문에 앞으로 양 기관의 정치성향 분포의 차이는 줄어들 것으로 예상할 수 있다. 만약 그렇다면 총선구도 인식에서도 양 기관의 워딩 효과 차이가 상당 부분 감소할 것으로 조심스럽게 전망해본다.

[그림22] 정치 성향 보정 전후 양 기관의 총선 인식 결과 차이의 변동
(1) 정당 지지 분포 보정 전



(2) NBS/갤럽의 정당지지율 평균 기준 가중치 보정 후



자료 한국리서치 웹 실험 조사 데이터(7~8월 통합)

주: 본 실험 조사 시점의 NBS와 갤럽의 정당지지율 평균을 구하여 립가중 방식으로 본 웹 조사 실험 집단들이 정당 지지율을 보정함.

누가 워딩 효과에 반응하는가도 중요하다

방법론적으로만 접근하면 어느 워딩이 나은가의 문제로 논의가 귀결되지만 양 기관의 총선 구도 측정 문항은 다음과 같은 점에서 근본적인 한계를 갖고 있다는 점을 놓쳐서는 안 된다. 총선 인식 구도는 대체로 대통령 임기 중에 치러지는 총선이나 지방선거가 대통령에 대한 “중간평가” 성격을 갖는다는 점에 착안하여 대통령과 정부 여당에 대한 ‘회고적 평가(retrospective evaluation)’에 기초한 총선 태도 측정 문항이라는 점에서 공통적이다. 질문 형식은 하나의 질문(일차원 질문)에서 가능한 양극단의 응답 중 하나를 택하는 일차원-양극화 모델에 기반한 측정 방식이다. 그러나 이러한 질문 방식은 여러 가지 점에서 문제가 있다. 물론 회고적 평가가 유권자의 투표행태에 미치는 중요한 변수임은 맞지만, 사실 이 단일 차원으로 단순하게 결정되지 않는다. 물론 정부·여당은 ‘정권 안정론’을 야당은 ‘정권 심판/견제론’을 내세워 경쟁하기도 하지만, 대부분의 총선 승부가 지역구 투표에서 결정된다는 점을 고려하면 정부와 여당은 ‘정권 안정론’ 대신 ‘지역발전론’을 내세워 ‘정권 심판론/견제론’의 예봉을 피하는 전략을 취함으로써 ‘지역 발전론’과 ‘정권 교체론’ 구도를 모색하기도 한다(강원택 2010; 정한울 2011, 2013, 2020; 장승진 2016, 2020; 허석재·정한울 2019).

또한 한국 사회에서 일차원-이분법 모델에서 탈피하여 다차원적인 상충성/양가성 태도에 기반하여 현 정부나 선거 경쟁 정당들을 평가하는 유권자층이 늘어난 점도 고려해야 한다. 2010년 이후에는 정부·여당에 대한 단일 차원에서 지지 후보를 선택하는 것이 아니라 여당은 여당대로, 야당은 야당대로 각각 평가하여 태도를 결정하는 신 중도층/스마트 유권자층이 등장했다. 즉 ‘정권 심판론’ 찬반으로 결정하지 않고 ‘야당’에 대해서도 심판 여부를 독립적으로 판단하면서 소위 ‘야당 심판론’의 축이 별도로 작동하는 유권자층이 늘어났다. 2012년 총선, 2016년 총선에서 선거 예측이 쉽지 않았던 것은 소위 진보-보수를 동시에 심판하겠다는 동시 심판론자들이 증가하면서 기존의 단일 차원 질문으로 유권자의 투표 방향을 제대로 예측하기 어려워졌기 때문이다. 이렇듯 회고적 평가의 관점에서 여러 차원이 공존할 수 있으며, 과거에 대한 회고적 평가 못지않게 미래의 전망적 기대도 별도로 작동한다는 점을 고려하면 한 문항으로 선거 구도를 파악하겠다는 것은 무리다(정한울 2011, 2013; 장승진 2016, 2020; 허석재·정한울 2019).

선거 구도 관련 워딩 효과는 복합적이고 다차원적인 태도를 하나의 차원으로 축약 시키는 과정에서 발생하게 된다는 것이 필자들의 생각이다. 2010년대 이후 유권자들은 정부와 여당에 대한 찬반만으로 투표하기보다는 그에 맞서는 야당에 대해서도 독립적인 평가를 통해 투표를 선택하는 ‘양면적/상충적 태도 갈등 층’의 규모가 커졌다. 이에 따라, 정부·여당에 대한 평가(정부·여당 심판론)와 함께 그에 대응하는 제1야당에 대한 평가(야당 심판론)도 동시에 고려해야 한다는 주장이 확산되어 왔다. 즉 정부/여당 대 야당을 ‘일차원 이분법적인 양자택일의 선택’이 아니라 정부·여당에 대한 찬반과 대안인 야당에 대한 찬반을 동시에 고려하는 ‘이차원 상충적 태도’를 측정해야 기존의 이분법 틀에서 포착되지 않는 소위 정부·여당과 야당 모두를 동시에 심판하고 싶어 하는 양면적/상충적 태도의 소위 ‘동시 심판론자’들을 분석해야 한다는 것이다. 이 경우 총선 인식 조사는 ‘정부·여당 심판/견제론’에 대한 동의 여부, ‘야당에 대한 심판/견제론’에 대한 동의 여부를 각각 측정하여 적어도 두 문항 이상의 복수 문항을 분석하는 것이 바람직하다.

이차원 상충적 태도 측정 문항은 정부·여당과 야당에 대한 양면적 태도 갈등이 늘어난 시점(예를 들면 2010~12년 안철수 현상의 확산 시기나 2022년 대선에서 양 후보에 대한 비호감이 공존하는 시기)에 더욱 유용하다. 이런 시기에는 워딩 효과나 질문 순서 효과와 같은 미묘한 차이가 현실화 될 가능성이 크다. 반대로 2017년 촛불, 탄핵 국면처럼 유권자들이 전체적으로 어떤 일방에 대한 비호감이 집중하고, 다른 한쪽을 대안으로 생각하는 경향이 확고한 시점에는 양면적 유권자층이 감소하고, 워딩과 질문 순서와 같은 미묘한 차이에 따라 응답이 달라질 가능성은 작아진다는 것이 본 보고서의 가정이다. 이는 조사기관 간의 차이, 조사 시점과 환경 등에 따라서 워딩 효과에 영향을 주는 응답자들의 태도나 편견의 강도 또한 달라질 수 있음을 고려해야 한다는 것이다(Schriesheim et al. 2011).³⁷⁾

37) 그럼에도 한 문항으로 측정하는 전통적인 문항을 사용하는 이유는 지금까지 전통적으로 한 문항으로 총선 구도를 파악해온 관성과 함께, 전화조사의 경우 대면면접이나 웹조사에 비해 1회 조사에 포함할 수 있는 문항 수가 제한적으로 ‘총선 구도 문항’ 파악에 여러 복수의 문항을 배치하는 데 현실적인 제약이 따르기 때문이다.

다차원 태도갈등층(여야 동시 심판론자)과 무당파층에서 워딩 효과가 두드러진다

이상의 논의를 종합하면 워딩 효과와 같은 미묘한 차이에 따라 응답이 달라지는 현상은 당파적 유권자층보다는 여와 야, 진보와 보수 양 진영에 대한 실망이 공유하는 양면적/상충적 태도의 유권자층이나, 태도가 유동적인 중도/무당파 성향의 유권자층에서 더욱 집중적으로 나타날 것을 가정해볼 수 있다. 즉 워딩 효과가 모든 유권자층에서 일괄적으로 나타나는 현상이라기보다는 진보와 보수, 여와 야 중 양자택일하지 못하고 양자 사이에서 갈등하는 양면적/상충적 태도 층이 많을수록 워딩이 강조하는 프레임의 영향이 클 수 있다는 가정이다. 실제로 정권 심판과 야당 심판 사이에 고민하고 있거나(둘 다 공감) 유동적인 태도(지지 정당이 없는 중도 무당파 성향의 유권자)를 가진 층에서 워딩 효과의 영향이 클 것이라는 본 보고서의 가설을 탐색해보자. 우선 태도 갈등 층을 분석하기 위해 2012년 총선부터 선거 구도 파악 시 진보 vs. 보수, 정부·여당 안정론 vs. 견제론/심판론 중 택일하는 방식이 아니라 정부·여당에 대한 심판론의 찬반을 묻고, 별도로 야당 심판론에 대한 찬반을 각각 물어 교차하여 2*2 네 개 유형으로 분류하여 회고적 평가 태도를 측정해왔다.³⁸⁾

문1. “내년 총선에서 윤석열 정부와 여당인 국민의힘을 심판해야 한다”는 주장에 대해 어떻게 생각하십니까?

1. 매우 공감한다
2. 대체로 공감한다
3. 별로 공감하지 않는다
4. 전혀 공감하지 않는다
9. 모름/무응답(읽지 말 것)

문2. “내년 총선에서 야당인 더불어민주당을 심판해야 한다”는 주장에 대해 어떻게 생각하십니까?

1. 매우 공감한다
2. 대체로 공감한다
3. 별로 공감하지 않는다
4. 전혀 공감하지 않는다
9. 모름/무응답(읽지 말 것)

자료: 한국일보·한국리서치(2023.5)

이러한 2차원 분석 틀은 아래와 같이 두 문항을 직접 교차하면 [표12]와 같은 2×2 네 개의 인식 유형화가 가능하다. 이 중 정부도 심판하고 싶은데, 야당도 동시에 심판하고 싶어 하는 ‘동시 심판론자’들이야말로 대표적인 태도 갈등이자 부동층의 핵심 집단으로 볼 수 있다.

- (1) 일방적 정부·여당 심판론 : 정부·여당 심판론 공감* 야당 심판론 비공감(전통적 진보/민주당 핵심 지지 성향)
- (2) 일방적 야당 심판론 : 정부·여당 심판론 비공감* 야당 심판론 공감(전통적 보수/국민의힘 핵심 지지 성향)
- (3) 여야 동시 심판론 : 정부·여당 심판론 공감* 야당 심판론 공감(양 진영의 이탈층/중도 무당파/중점관여층)
- (4) 심판론 냉소/무관심층 : 정부·여당 심판론 비공감* 야당 심판론 비공감(정치 냉소, 무관심 비 관여층)

올 5월에 한국일보 기획으로 진행한 조사 결과에 따르면 ‘일방적 정부·여당 심판론’과 ‘일방적 야당 심판론자’들이 당파적 성향이 강한 양 당의 상대적으로 안정적인 지지기반의 세력 분포를 보여준다. 일방적 정부·여당 심판론이 30%, 일방적 야당 심판론자들이 24%로 약간 정권심판론이 약간 우세하지만, 대체로 경합하는 국면이며, 양쪽 모두를 심판하고 싶어 하는 동시 심판론자도 24%로 캐스팅보트(Casting voter)로서 영향력을

38) 이러한 이차원 여야 심판론에 기초한 선거 기획보도를 진행해왔고, 중앙일보, 한국일보 등 다양한 언론을 통해 소개된 바 있다. 중앙일보 “MB 심판 선거 63%, 야당 심판이다. 59%”(2012/04/09); 한국일보 “총선 ‘야당 심판론’이 ‘정권심판론’ 앞질러”(2016/02/25); “정권심판론 52%. 야당 심판론 앞섰다, 야당 심판론 우세했던 한 달 전과 정반대.”(2016/04/01), “더민주, 야 분열 책임 전가 자충수. “야당 동시 심판론 공감” 응답 중 호남, 20대 지지층 비율 높아”(2016/04/08); SBS “민주 압승 파란 물결 몰아쳤다. 민심, 보수 야당 심판.”(2018/06/14); 주간조선 “4대 변수로 본 2020 총선 관전법.”(2019/05/13); 한국일보 “정부·여당 심판론 대 야당 심판론 백중세” (2023/05/10). ‘정권심판론’과 ‘야당심판론’을 교차한 유형화를 통해 유권자의 투표행태를 분석한 논문과 보고서로는 정한울(2013; 2020), 장승진(2016; 2020)을 참조할 것.

발휘할 수 있는 크기이다. 어느 쪽에도 공감하지 않은 냉소/무관심층 23%로 이들은 대체로 정치에 관한 관심이 거의 없으며, 실제 투표 참여도 떨어지는 비 관여층에 가까운 집단이다. 그러나 동시 심판론자들은 선거 시기에 집중적으로 판단을 위한 정보 취득에 나서고, 실제 투표 선택의 변동도 가장 많이 일어나는 대표적인 부동층이자, 탈당파적/탈이념적 성향이 강한 층이며, 이들이 어느 쪽으로 연합하느냐에 따라 판세가 결정되는 핵심 캐스팅보트(Casting voter)가 되기도 한다. 어느 쪽으로 선호가 쏠리지 않고, 정부·여당도 심판하고 싶고, 야당도 심판하고 싶은 동시 심판론자들이야말로 중도/무당파의 핵심 구성 집단 중 하나이며, 워딩 효과의 영향을 가장 크게 받을 수 있다.

[표12] 여야 심판론에 근거한 이차원 회고적 선거 구도 문항으로 본 2024년 총선 인식 유형 분포

| 2차원 분석 모델 | | 야당 심판론 (보수) | |
|----------------------|----------------|--|--|
| | | 찬성(공감) 동시 심판론 (상충적/양면적 태도 갈등) 24% | 반대(비공감/유보) 일방적 정권심판론 (전통적 진보) 30% |
| 정부·여당 심판론 (진보) | 찬성 (공감) | | |
| | 반대 (비공감/유보) | 일방적 야당 심판론 (전통적 보수) 24% | 냉소/무관심층 23% |

출처: 정한울(2013; 2020), 자료: 한국일보·한국리서치(2023.5)

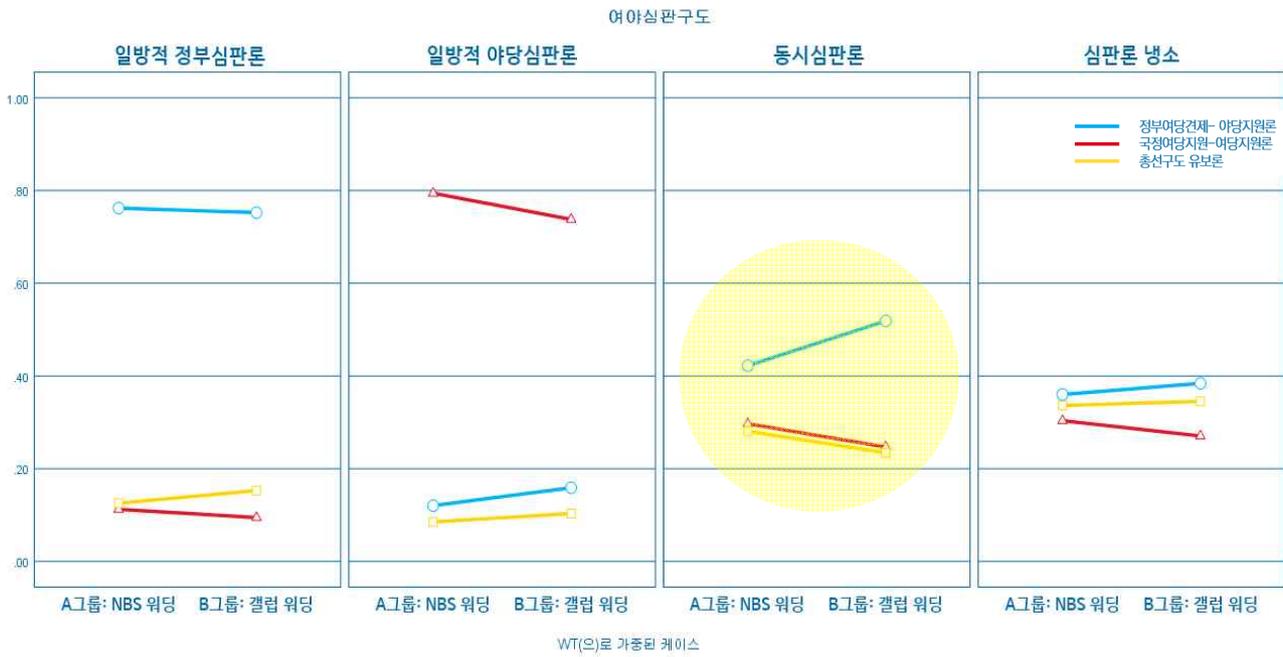
[그림23]-(1)은 이번 실험 조사 시 여야 심판론의 유형별 워딩 효과의 크기 차이가 있는지 살펴보기 위해 이들 네 집단별로 워딩 실험에 얼마나 다른 응답을 하는지 확인한 그래프이다. 8월 2차 실험 조사에 위의 정부·여당 심판론과 야당 심판론에 대해 질문을 하여 네 집단으로 분류한 후 A그룹의 NBS 워딩이 주어진 경우와 B그룹의 깎벌 워딩이 주어졌을 때 얼마나 응답 차이가 나는지 살펴보았다. 본 보고서의 가정대로 당파적 태도가 확고한 일방적 정부 심판론자나 일방적 야당 심판론자, 그리고 정치에 관심과 참여도가 낮은 냉소 층에서는 A그룹과 B그룹에서의 응답 변화가 크지 않았지만, 태도 갈등 층인 동시 심판론자들에게서 강한 워딩 효과가 확인된다.

마찬가지로 각 진영의 대표적 정치 지도자에 대한 태도에서 나타나는 태도갈등 여부나 탈당파적 성향(상충적/양면적 태도층)에서 워딩효과가 두드러진다. 8월 실험조사에서는 기존의 워딩실험 외에 윤석열 대통령에 대한 호감도, 이재명 더불어민주당 대표에 대한 호감도를 각각 측정하고 이를 교차하여 네 가지 응답유형을 분류하여 A, B 실험 워딩별 차이의 크기를 확인하였다. 일방적 태도층인 ‘일방적 윤석열 호감층(윤 호감, 이 비호감)’이나 ‘일방적 이재명 호감층(윤 비호감, 이 호감)’에서보다 ‘동시 비호감층(윤 비호감, 이 비호감)’ 층과 ‘동시 호감층(윤 호감, 이 호감)’ 층에서 A, B 워딩간 차이가 크게 나타났다(부록-그림1을 참조할 것).³⁹⁾

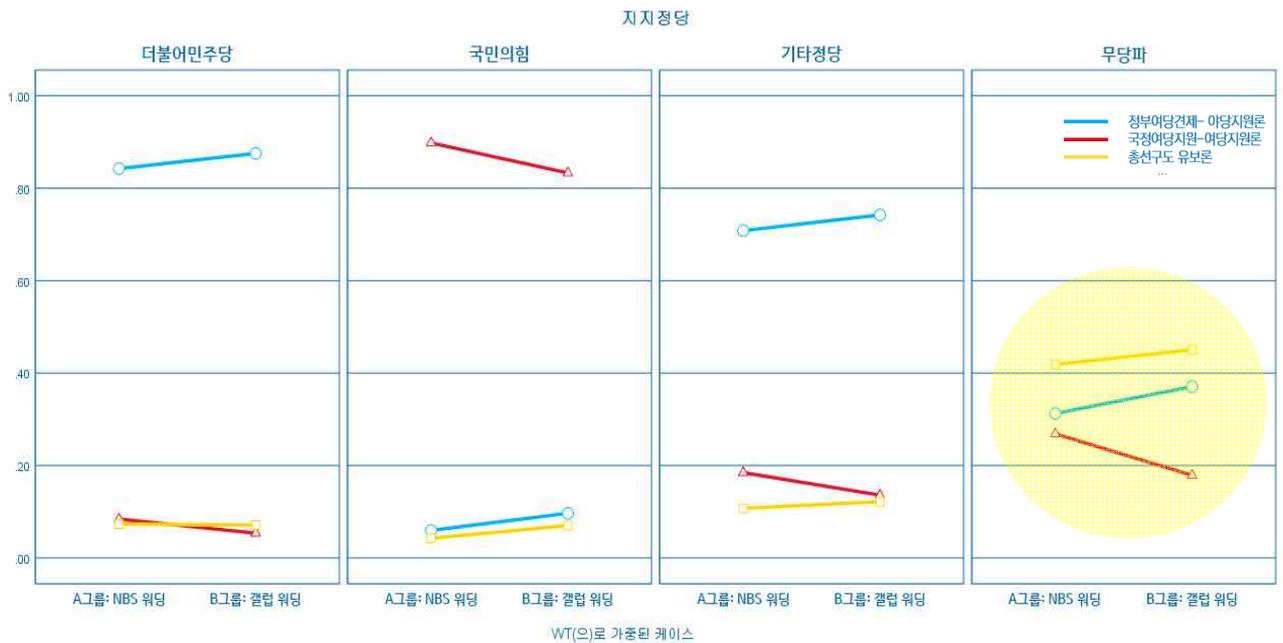
또한 [그림23]-(2)에서도 확인되듯이 정당 태도를 보면 더불어민주당이나 국민의힘, 기타정당 등 지지하는 정당이 있는 당파적 유권자층에 비해 지지하는 정당이 없는 무당파층에서 총선 인식 질문 워딩에 따라 응답 분포의 차이가 뚜렷함을 알 수 있다. 결국 워딩 효과가 모든 응답자에게 일률적으로 동일한 효과를 유발하기 보다는 특히 태도 갈등이 있고, 변동 가능성이 있는 부동층에서 더 집중되는 현상임을 알 수 있다. 그렇다면 문제의 해법은 단일한 워딩으로 조정하기에 앞서 다차원적이고 복합적인 태도를 가진 유권자층의 생각을 무리하게 일차원적인 문항으로 측정하는 관행에서 벗어나는 데에서 시작해야 할 듯하다.

39) 크기로 보면 ‘동시 비호감층’보다 ‘동시 호감층’에서 워딩 간 차이가 더 컸는데 이는 ‘동시 호감층’의 응답자 규모가 너무 적어서 나타난 결과로 보인다. ‘일방적 윤 호감-이 비호감’ 유형이 26%(255명), ‘일방적 윤 비호감-이 호감’ 유형(275명)이 28%로 팽팽한 반면, ‘윤 비호감-이 비호감 동시 비호감’ 층(458명)이 46%로 과반에 육박한 반면, ‘윤 호감-이 호감의 동시 호감층’은 1%(12명)에 불과했다. 윤석열 대통령, 이재명 대표 공히 호감 비율이 30%에 미치지 못하고, 비호감 비율은 둘 다 70%를 상회하는 강한 비토층을 갖고 있다는 점에서 공통적이다.

[그림23] 워딩 효과가 집중적으로 나타나는 집단 : 태도 갈등/탈진영/중저 관여층
 (1) 여야 심판론 유형별 워딩 효과의 차이



(2) 지지 정당별 워딩 효과



5. 요약 및 전망

본 보고서는 선거여론조사 방법으로서 기존의 전화면접조사와 ARS 조사 간 조사 품질과 정확성 논쟁 과정에 대한 기존 연구를 검토하고, 최근 이슈가 되고 있는 동일한 전화면접조사 간 결과 차이에 대해 실증적인 실태 파악과 시론적 차원에서 문제 진단을 정리하고자 했다. 우선, 최근 생산된 선거여론조사 데이터들에 대한 집계자료를 검토한 기존 연구를 통해 전화면접조사와 ARS 조사 품질의 차이와 실제 차이가 나타나는 요인들을 정리하였다.

첫째, 전화면접조사와 ARS 조사의 비교는 크게 (1) 정치적 편향 (2) 신뢰성과 대표성(품질 지표) (3) 정확성을 중심으로 논쟁이 이루어져 왔다. 조사 품질과 관련해서는 여심위가 규제하는 (1) 응답률 (2) 가중배율 지표에서 전화면접조사의 품질이 우월하며 이러한 조사 품질이 주로 표본 구성에서 무당파 비율 등 중/저 관여층의 대표성에서 차이가 나는 것으로 알려져 있다. 무당파나 중/저 관여층의 표집이 제대로 이루어지지 않으며 ARS 조사에서 당파적 유권자층, 극단적 태도 층이 과대 표집되는 문제로 연결되었고, 그 결과 “전화면접은 민주당/이재명 후보에 유리하고, ARS는 국민의힘에 유리하다”라는 편향성에 대한 선부른 규정들이 상식처럼 자리 잡게 되었다. 그러나 불과 1~2년 사이 윤석열 정부 시기에는 반대로 “전화면접이 국민의힘에 유리하고 ARS 조사는 민주당에 유리하다”는 반대 공식이 유행하는 상황도 벌어지고 있다. 한편 실제 선거 결과와 조사 결과와의 차이(편향척도 A)를 통해 측정된 정확성에 있어서는 상이한 연구 결과들이 소개되고 있지만, 현재 정확성의 기준을 선거 결과와의 편차로 측정할 경우 (1) D-6 시점의 구조적인 오차 요인이 개입 할 수도 있고, (2) 선거 국면의 변동성이 큰 경우 분석 기간 설정에 따라 평가가 달라지며 (3) 경주마 보도의 틀을 벗어나지 못한 여론조사 활용의 문제를 낳을 수 있다는 점에 주목했다. 기존 연구들은 대체로 이러한 문제를 극복할 새로운 정확성 지표의 개발, 메타분석을 활용한 조사연구 등의 확장이 필요하다는 점을 강조해왔다.

그러나 최근에는 전화조사와 ARS 조사 간의 차이 외에 전화면접조사를 활용한 대표적인 선거여론조사인 NBS와 갤럽의 조사 결과에서, 나아가 NBS 조사에 참여하는 기관들의 발표 결과 사이에서도 상당한 혼선을 낳는 상황이 잦아지면서 논의가 새롭게 전개되고 있다. 특히 선거행태를 좌우하는 태도 결정요인인 정당 지지율 분포나 총선 구도 인식 문항에서 체계적인 응답 차이가 지속적으로 확인되면서 혼란이 가중되고 있다.

이에 관해 본 연구에서는 윤석열 정부 시기(2022년 5월-2023년 10월까지) NBS 조사와 갤럽조사 방법 관련 정보에 대한 집계자료와 실제 조사 결과 데이터를 수집하여 양 기관 조사 결과의 차이가 집중적으로 나타나는 지점이 무엇이고, 차이를 만들어내는 요인에 대한 잠정적인 진단 결과를 본 보고서를 통해 소개하고자 했다. 데이터 분석 결과 양 기관의 결과 차이는 국정 평가, 국민의힘 지지율에서는 큰 차이가 없으나 민주당 지지율/무당파+제3정당 비율, 그리고 차기 총선 구도 문항에 집중적으로 나타나고 있다. 즉 NBS 조사가 갤럽조사 대비 민주당 비율은 낮고, 무당파/제3정당 지지율이 높으며, 총선 인식 문항에서도 국정 안전론에 대한 지지가 상대적으로 높고, 갤럽조사는 민주당 지지율과 국정 견제론이 상대적으로 높게 나오는 것이 확인되었다.

그 원인으로 본 보고서에서는 (1) 중/저 관여층의 표집에 영향을 주는 응답률 효과 (2) 총선 구도 문항의 경우 양 조사의 정치 성향 분포 요인과 함께 ‘위딩 효과’의 영향을 검증하고자 했다. 실제로 양 기관의 응답률 평균을 보면 갤럽조사 대비 NBS 조사가 높아 ‘높은 응답률=높은 무당파층=낮은 민주당 지지율’을 낳는다는 또 다른 통념을 낳고 있다. 그리고 이러한 응답률의 차이는 주로 양 기관의 조사 방법의 차이로 설명해오곤 했다. NBS는 100% 무선 가상번호를 활용한 층화확률 표집을, 갤럽은 유무선 RDD 방법으로 조사를 진행해왔기 때문이다. 이는 갤럽이 상대적으로 낮은 응답률을 기록했던 요인으로 지적되었다. 그러나 분석 결과 양 조사기관을 통제하여 살펴보면 응답률과 표본 성향 분포하고는 뚜렷한 상관관계가 보이지 않았으나 추가분석을 통해 응답률이 조사 시점, 조사기관 효과와 상호작용하면서 응답률의 효과를 상쇄시킨 결과로 해석된다. 즉 응답률은 조사기관에 따라, 조사 시점에 따라 상반된 방향으로 영향력을 미치고 있었음이 확인되었고, 응답률의 효과를 살펴보기 위해서는 성급한 일반화보다 복합적인 요인 간 상호작용 효과까지 고려한 심층적인 추가 검증이 필요하다는 시사점을 얻을 수 있었다. 다만 갤럽 조사가 23년 7월 이후 NBS와 동일한 가상번호

층화확률 표집 방법으로 전환하면서 양 기관의 조사방법 상의 큰 차이는 사라졌고, 그동안 통계적으로 유의한 차이를 보였던 양 기관의 응답률 격차는 좁혀지는 추세이다. 또한 향후 추세를 지켜봐야 하지만, 조사 방법 조정 이후 NBS와 갤럽 조사 사이의 더불어민주당 지지율 격차가 감소하는 추세를 보인다는 점도 새롭게 발견할 수 있었다.

총선 인식 구도 문항도 최근 그 격차가 좁혀지는 추세이기는 하지만, NBS 조사는 여전히 국정 안정론과 정부 견제론이 경합 구도를 보여준다. 반면, 갤럽조사는 일관되게 정부·여당 견제론의 우위를 보여 여전히 차이가 줄어들 조짐을 보이지 않는다. 실제로 총선 인식 문항 결과의 차이가 발생한 시점에는 양 기관의 정치 성향 분포(정당지지율, 이념 성향 분포)에서 상당한 차이가 발견되어 이 역시 양 기관 조사표본의 정치 성향 분포 차이 탓으로 이해되어온 측면이 있다. 본 보고서에서는 갤럽의 조사 방법 전환 이후 무당파 비율의 차이는 유의미할 정도로 나타나지 않았고, 특히 최근에는 반대로 NBS에서의 더불어민주당 지지율이 상승하여 양 조사 간 격차가 줄어드는 추세이다. 그런데도 총선 인식에서 상당한 차이가 발견되면서 본 보고서에서는 일각에서 제기되고 있는 '위딩 효과'의 영향력을 웹 실험 조사를 통해 검증하였다. 7월과 8월에 각각 조사 응답자를 A, B 두 그룹으로 무작위 배분한 후 A그룹에는 NBS 조사 위딩을, B그룹에는 갤럽 조사 위딩을 제시하여 조사한 결과 실제 NBS, 갤럽 조사처럼 A그룹에서 안정론과 견제론의 격차가 줄어들고, B그룹에서는 견제론 우위 현상이 뚜렷하게 나타났다. 무작위 배정의 결과로 A, B 그룹의 공변인 분포가 대부분 균질적으로 배분되었지만, 우연의 결과로 정당 지지율에서는 A그룹에 민주당 지지층 비율이 상대적으로 과소 대표되어 정치 성향 분포 차이 효과가 위딩 효과와 혼재되어 있음이 확인되었다. 웹 실험 결과 인구학적 요인(성/세대/지역/학력/소득) 및 정치 성향 분포(이념 성향/지지 정당) 변수들을 포함하여 분석한 결과 정치적 성향, 세대-지역 등의 인구학적 요인도 총선 인식에 영향을 강하게 미치고 있으나, 이들 요인을 통제한 조건에서도 양 실험 집단 간의 총선 구도 인식 결과 차이가 나타나 위딩 효과가 작동하고 있음이 확인되었다. 그러나 양 실험 집단 간 정치 성향 분포 요인이 만드는 차이도 뚜렷하여 실험응답자들의 정치 성향 분포를 가중치를 사용하여 NBS/갤럽의 평균 분포에 맞추면 A/B 두 실험 집단 간 총선 인식 결과 차이는 절반 이하로 감소함을 확인할 수 있었다. 그동안 주목하지 않았던 위딩 효과가 작동한다는 발견은 중요한 시사점을 주지만 그렇다고 과도한 확대해석도 경계할 필요가 있음을 강조하고자 한다. 다 나아가 이러한 위딩 효과가 발생하는 요인에 대해 보고서에서는 위딩 효과를 유발할 요인들에 대한 몇 가지 가설(질문 위딩의 차이-공감 대 동의, 응답 선택지 위딩 차이-NBS는 대통령의 국정운영 대 야당의 견제, 갤럽은 정부 지원/견제 목적 위한 여당 대 야당 구도라는 점, NBS 조사의 안정론 위딩은 갤럽조사 대비 대통령 국정 전망적 차원의 지원을 부각할 가능성)들을 제기하였다.

본 보고서에는 위딩 효과를 유발하는 요인으로 위딩 자체 못지않게 질문의 형태에 대해서도 논의하였다. NBS 질문과 갤럽 질문 모두 총선의 의미와 목적을 정부·여당에 대한 태도(지지 vs 견제)와 함께 '여야에 대한 태도'라는 두 개의 독립적인 차원이 혼재된 위딩으로 묻고 있어. 정부·여당에 태도와 야당에 대한 태도가 상충하는 유권자들의 입장은 배제되는 문항이라는 한계를 가지고 있다고 지적한다. 따라서 정부·여당에 대해 심판하고 싶지만, 야당에 대해서도 심판하고 싶은 양면적 태도/갈등을 가진 유권자들일수록 질문에 대한 응답 선택 시 내면의 태도 갈등을 유발하여, 위딩에 따라 응답이 영향받을 가능성이 큰 것에 주목하였다. 이러한 가설이 타당하다면 당파적 태도가 강한 일방적 심판자나 정당 태도가 분명한 유권자보다 여야를 동시에 심판해야 한다고 생각하는 태도 갈등 층이나 당파적 태도가 약한 무당파층이 위딩 차이에 따라 더 민감하게 반응할 것으로 보았다. 실제 웹 실험 결과를 분석한 결과 정당 태도가 분명한 당파적 유권자층이나 어느 한쪽을 심판해야 한다는 일방적 심판론자에 비해 무당파층이나 여야 동시 심판론자들이 A/B 그룹 간 위딩 차이에 따라 상이한 응답을 하는 경향이 강했다. 위딩 효과를 순수 위딩의 차이에서 비롯되는 현상이 아닌, 다차원적인 태도를 일차원적 양자택일의 선택 문제로 압축하는 과정에서 다차원적이고, 양면적 태도를 가진 유권자층의 내적 갈등을 유발하면서 위딩 효과를 강화할 것이라는 가설이 설득력을 가질 수 있음을 시사해주는 결과라 할 수 있다.

이상의 논의를 종합하면서 본 보고서는 조심스럽지만 향후 양 기관의 정치적 성향 응답 분포의 차이나 총

선 인식 구도에서의 결과 차이가 줄어들 가능성에 주목하고자 한다. 우선 양 기관의 조사 방법론이 조정되면서 응답률이나 정치성향 분포에 영향을 줄 방법론적 차이가 줄어들었고, 아직 일반화하기는 이르지만 정당 지지율 차이가 감소하는 추세로 돌아섰다. 물론 총선 문항의 경우 독립적인 워딩 효과임이 확인되었기에 정당 지지율에 비해 일정한 차이가 유지될 가능성이 있다. 그러나 본 연구에서 주장한 것처럼 워딩 효과가 당파성이 약하거나 여야 모두에 불만을 갖고 있는 상층적/양가적 유권자 층에서 집중되는 것이라면 최근 윤석열 정부의 '홍범도 동상 이전' 등의 이념노선 강화, '이재명 대표 구속 적부심 기각', '강서구청장 보궐선거 패배' 등으로 여야 동시심판론이 감소하고 '정권심판론/견제론'쪽으로 무게 중심이 이동할 경우 워딩 효과의 영향력도 감소할 가능성이 존재한다.

본 보고서는 기존의 전화면접조사 대 ARS 조사 간의 방법론 논쟁을 넘어 심화되고 있는 전화면접조사 간 차이를 만들어내는 요인에 대해 여러 가설들을 정리하면서 가능한 선에서 실증적으로 검증해보고자 시도한 시론적 차원의 연구 결과이다. 문항 순서 효과나 기타 요인들에 대해서는 본 보고서에서 다루지 못한 한계도 있으며, 본 연구에서 응답률과 조사기관, 조사 시점 간 상호작용의 영향은 발견하였지만, 왜 이러한 현상이 발생하는지에 대한 해석은 매우 미진한 것이 사실이다. 그러나 그동안 여론조사 방법에 대한 쟁점에 대해 몇 가지 그럴듯한 가설 수준의 논의를 실증적으로 검증하기 위한 노력 대신 성급한 일반화나 공식화를 통해 미봉적으로 대응했던 경향에서 탈피하고자 했던 시론적 연구로서 의미를 갖기를 기대한다. 조사 방법 논쟁의 2라운드가 보다 생산적이고 심층적인 후속 연구로 진화하길 기대한다.

[부록-표1] 동일 시점 양 조사 간 국정 평가 비율 차이(% , N=60)

| 독립표본 T 검증 H _a μ _{NBS} ≠ μ _{한국갤럽} | 조사기관 Effect | 신뢰구간 | | |
|--|-------------|---------|------|------|
| | | 평균 M | LL | UL |
| 긍정 평가 N = 30 p= 0.296 | NBS | 35.8 | 33.4 | 38.1 |
| | 한국갤럽 | 34.0 | 31.6 | 36.4 |
| | NBS - 한국갤럽 | 1.8 | -1.6 | 5.1 |
| 부정 평가 N = 30 p=0.220 | NBS | 53.6 | 50.2 | 56.9 |
| | 한국갤럽 | 56.5 | 53.2 | 59.9 |
| | NBS - 한국갤럽 | -3.0 | -7.7 | 1.8 |
| 모름/무응답 N = 30 p=0.165 | NBS | 10.7 | 9.4 | 11.9 |
| | 한국갤럽 | 9.5 | 8.2 | 10.7 |
| | NBS - 한국갤럽 | 1.2 | -0.6 | 3.0 |

[부록-표2] 동일시 점 양 조사 간 정당 지지율 차이(% , N=60)

| 독립표본 T 검증 H _a μ _{NBS} ≠ μ _{한국갤럽} | 조사기관 Effect | 신뢰구간 | | |
|--|-------------|---------|-------|------|
| | | 평균 M | LL | UL |
| 국민의힘 지지율 N = 30 p= 0.947 | NBS | 35.3 | 33.9 | 36.7 |
| | 한국갤럽 | 35.2 | 33.8 | 36.6 |
| | NBS - 한국갤럽 | +0.07 | -1.92 | 0.99 |
| 민주당 지지율 N = 30 p < 0.001** | NBS | 28.9 | 28.0 | 29.9 |
| | 한국갤럽 | 32.4 | 31.5 | 33.4 |
| | NBS - 한국갤럽 | -3.53 | -4.88 | 0.67 |
| 제3정당 지지율 N = 30 p < 0.001** | NBS | 6.6 | 6.2 | 7.0 |
| | 한국갤럽 | 5.2 | 4.8 | 5.6 |
| | NBS - 한국갤럽 | +1.40 | 0.27 | 0.87 |
| 무당파 지지율 N = 30 p=0.05+ | NBS | 29.0 | 27.6 | 30.4 |
| | 한국갤럽 | 27.1 | 25.7 | 28.5 |
| | NBS - 한국갤럽 | +1.93 | 0.98 | 3.90 |

주) + (90% 신뢰수준), * (95% 신뢰수준), ** (99% 신뢰수준) *** (99.9% 신뢰수준)에서 통계적으로 유의한 차이

[부록-표3] 동일 시점 양 조사의 응답률 차이(N=60)

| 독립표본 T 검증 p < 0.001*** | 사례 수 | 평균 | 중앙값 | 표준편차 | 표준오차 |
|------------------------|------|------|------|------|------|
| 전체 | 60 | 14.1 | 14.3 | 3.7 | 4.8 |
| NBS | 30 | 17.2 | 17.3 | 2.1 | 3.9 |
| 한국갤럽 | 30 | 11.1 | 10.6 | 1.9 | 3.5 |

[부록-표4] 2021년-22년 한국리서치 전화면접 선거조사의 문항 수별 응답률 평균 차이(ANOVA Test)

| ANOVA Test: p < 0.001*** | 사례 수 | 응답률 (%) | | 95% 신뢰구간 (%) | |
|--------------------------|------|---------|------|--------------|--------|
| | | 평균 | 표준오차 | 하한(LL) | 상한(UL) |
| 전체 | 42 | 24.2 | 0.7 | 22.8 | 25.7 |
| 15~19개 | 22 | 27.5 | 0.6 | 26.1 | 28.8 |
| 20~24개 | 11 | 22.2 | 0.9 | 20.3 | 24.1 |
| 25개 이상 | 9 | 18.8 | 1.0 | 16.4 | 21.1 |

[부록-표5] NBS·갤럽조사 결과로 본 민주당·국민의힘·제3정당·무당파 비율에 대한 영향요인: OLS 회귀분석

| 예측 모델 N=103 | 국민의힘 | | 민주당 | | 제3정당 | | 무당파 | |
|-------------------------|-----------------|------|-----------------|------|-----------------|------|-----------------|------|
| | 추정값(B) | 표준오차 | 추정값(B) | 표준오차 | 추정값(B) | 표준오차 | 추정값(B) | 표준오차 |
| (Intercept) | 0.346**(0.005) | | 0.303**(0.003) | | 0.061**(0.001) | | 0.290**(0.004) | |
| 조사 기관 (1=NBS, 0=갤럽) | -0.008(0.016) | | -0.045**(0.012) | | 0.021**(0.005) | | 0.037**(0.013) | |
| 조사 방법 (1=가상, 0=RDD) | 0.017(0.021) | | 0.013(0.015) | | -0.008(0.006) | | -0.010(0.016) | |
| 조사 시점 1. (~23년 6월) | -0.021**(0.008) | | -0.010+(0.005) | | 0.004+(0.002) | | 0.031**(0.006) | |
| (0=22년12월)2. (~23년 10월) | -0.041**(0.014) | | -0.032**(0.010) | | 0.011**(0.004) | | 0.063**(0.011) | |
| 응답률 | -0.161(0.233) | | -0.042(0.165) | | -0.027(0.066) | | -0.051(0.184) | |
| 모델 F(df1, df2) | F(5, 97)=3.43** | | F(5, 97)=14.5** | | F(5, 97)=10.2** | | F(5, 97)=18.4** | |
| Adj R ² | 0.106 | | 0.399 | | 0.312 | | 0.460 | |

주) + (90% 신뢰수준), * (95% 신뢰수준), ** (99% 신뢰수준) *** (99.9% 신뢰수준)에서 통계적으로 유의한 차이

[부록-표6] 웹 실험 A/B 그룹 응답자의 인구학적 구성비 (7~8월 조사 통합)

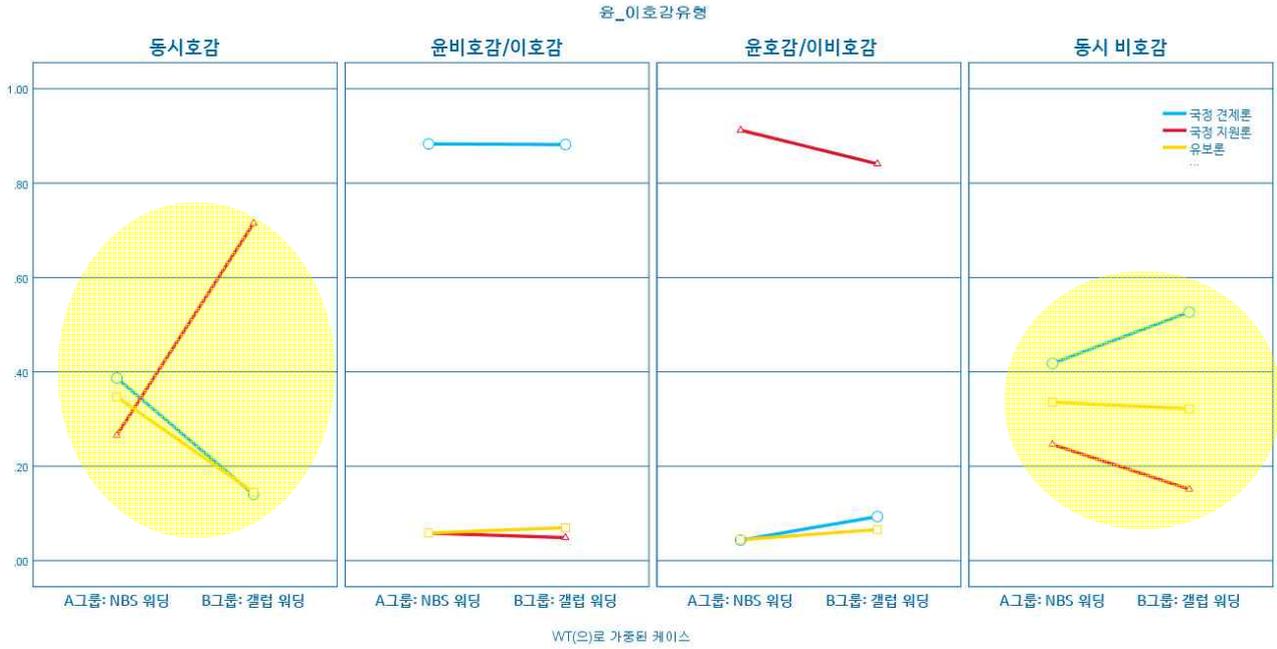
| | | A그룹: NBS 워딩 | | | | B그룹: 갤럽 워딩 | | | | 총계(2,000명) | | | |
|---------------------------------|----------|-------------|----|----|----|------------|----|----|----|------------|----|----|----|
| 성 별 | 남자 | 55 | 53 | 41 | 52 | 53 | 47 | 41 | 48 | 54 | 50 | 41 | 50 |
| | 여자 | 45 | 47 | 59 | 48 | 47 | 53 | 59 | 52 | 46 | 50 | 59 | 50 |
| 연 령 | 18-29세 | 16 | 14 | 28 | 17 | 11 | 13 | 32 | 16 | 14 | 13 | 30 | 16 |
| | 30-39세 | 10 | 14 | 17 | 13 | 15 | 17 | 21 | 17 | 12 | 16 | 19 | 15 |
| | 40-49세 | 11 | 25 | 21 | 19 | 11 | 20 | 20 | 17 | 11 | 22 | 20 | 18 |
| | 50-59세 | 19 | 22 | 16 | 20 | 13 | 24 | 14 | 19 | 16 | 23 | 15 | 19 |
| | 60세이상 | 44 | 26 | 17 | 32 | 49 | 26 | 13 | 31 | 46 | 26 | 15 | 31 |
| 거 주 지 역 | 서울 | 19 | 17 | 17 | 18 | 24 | 18 | 15 | 19 | 21 | 18 | 16 | 19 |
| | 인천/경기 | 31 | 30 | 30 | 31 | 28 | 34 | 37 | 33 | 30 | 32 | 34 | 32 |
| | 대전/세종/충청 | 11 | 13 | 11 | 11 | 10 | 9 | 11 | 10 | 11 | 10 | 11 | 11 |
| | 광주/전라 | 6 | 12 | 10 | 9 | 4 | 14 | 8 | 10 | 5 | 13 | 9 | 10 |
| | 대구/경북 | 12 | 7 | 10 | 9 | 15 | 8 | 9 | 10 | 13 | 7 | 10 | 10 |
| | 부산/울산/경남 | 18 | 15 | 20 | 17 | 15 | 12 | 13 | 13 | 16 | 14 | 17 | 15 |
| | 강원/제주 | 3 | 6 | 2 | 4 | 3 | 5 | 7 | 5 | 3 | 5 | 5 | 4 |
| 학 력 | 고졸이하 | 57 | 45 | 48 | 50 | 51 | 48 | 46 | 49 | 55 | 47 | 47 | 50 |
| | 대재이상 | 43 | 55 | 52 | 50 | 49 | 52 | 54 | 51 | 45 | 53 | 53 | 50 |
| 월 평 균 가 구 소 득 | 200만원미만 | 14 | 11 | 20 | 14 | 14 | 13 | 17 | 14 | 14 | 12 | 18 | 14 |
| | -300만원미만 | 18 | 15 | 18 | 17 | 13 | 12 | 21 | 14 | 16 | 14 | 20 | 16 |
| | -400만원미만 | 17 | 17 | 13 | 16 | 19 | 15 | 17 | 16 | 18 | 16 | 15 | 16 |
| | -500만원미만 | 16 | 10 | 12 | 13 | 12 | 17 | 11 | 14 | 14 | 14 | 11 | 14 |
| | -600만원미만 | 13 | 13 | 10 | 13 | 12 | 14 | 11 | 13 | 13 | 14 | 10 | 13 |
| | -700만원미만 | 5 | 11 | 6 | 8 | 8 | 9 | 8 | 8 | 7 | 10 | 7 | 8 |
| | 700만원이상 | 17 | 21 | 22 | 20 | 21 | 21 | 17 | 20 | 19 | 21 | 19 | 20 |

[부록-표7] 웹 실험 A/B 그룹 응답자의 공변인(정치성향 차이) (7-8월 조사 통합)

| | | A그룹: NBS 워딩 | | | | B그룹: 갤럽 워딩 | | | | 총계(2,000명) | | | |
|------------------|----------|-------------|----|----|----|------------|----|----|----|------------|----|----|----|
| 이 념 | 진보(0-4) | 7 | 47 | 17 | 27 | 6 | 44 | 12 | 27 | 6 | 46 | 15 | 27 |
| | 중도(5) | 26 | 35 | 51 | 34 | 26 | 37 | 56 | 37 | 26 | 36 | 53 | 36 |
| 성 향 | 보수(6-10) | 64 | 15 | 22 | 35 | 66 | 16 | 21 | 33 | 65 | 16 | 22 | 34 |
| | 모름 | 3 | 2 | 10 | 4 | 2 | 2 | 11 | 4 | 3 | 2 | 10 | 4 |
| 지 지 정 당 | 더불어민주당 | 7 | 63 | 14 | 33 | 7 | 67 | 17 | 40 | 7 | 65 | 15 | 36 |
| | 정의당 | 2 | 7 | 3 | 4 | 1 | 6 | 3 | 4 | 2 | 7 | 3 | 4 |
| | 진보당 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| | 국민의힘 | 66 | 5 | 10 | 29 | 78 | 5 | 9 | 28 | 71 | 5 | 9 | 29 |
| | 기본소득당 | 0 | 3 | 0 | 1 | 0 | 2 | 1 | 1 | 0 | 3 | 0 | 1 |
| 국 정 평 가 | 시대전환 | 1 | 1 | 2 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| | 다른 정당 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | 무당파 | 23 | 20 | 70 | 30 | 13 | 20 | 71 | 27 | 18 | 20 | 71 | 28 |
| 국 정 평 가 | 잘하고 있다 | 72 | 6 | 10 | 32 | 79 | 5 | 9 | 29 | 75 | 5 | 9 | 30 |
| | 못하고 있다 | 24 | 92 | 72 | 63 | 18 | 93 | 73 | 66 | 21 | 92 | 73 | 64 |
| | 모르겠다 | 4 | 2 | 18 | 6 | 4 | 2 | 19 | 6 | 4 | 2 | 18 | 6 |

주) 한국갤럽의 8월 3주 결과는 2주/4주 결과 평균

[부록-그림1] 윤석열 대통령-이재명 대표 호감도 유형별 총선인식 워딩 효과



[참고문헌]

- 강원택. 2010. “2002년 지방선거의 정치적 의미: 중간평가 혹은 대선 전초전. 『한국 선거정치의 변화와 지속: 이념, 이슈, 캠페인과 투표참여』 파주: 나남. 271-294.
- 김영원. 2023. “추석 민심 여론조사 차이가 나는 이유는?” 김영원 교수 페이스북 https://m.facebook.com/story.php?story_fbid=2495914243915378&id=100004904791122&mibextid=ZbWKwL (검색일. 2023.10.10.)
- 고길곤·김대중. 2018. “선거 여론조사의 비관측 오차 영향요인에 관한 연구: 20대 총선 여론조사 결과를 중심으로.” 『조사연구』 19.1. 27-60.
- 구본상. 2018. “제7회 전국동시지방선거 여론조사 평가와 제안.” 『제7회 전국동시지방선거 선거여론조사 백서』 121-128.
- _____. 2017. “ARS 조사방식의 젊은 연령대 여성 표집의 실패: 정치적 의견이 강한 유권자들의 비율에서의 성별 차이를 중심으로.” 『조사연구』 18.1. 31-60.
- 구본상·박원호. 2022. “선거여론조사의 쟁점과 제안: 제20대 대선과 제8회 지선.” 『2022년 중앙선거관리위원회 연구용역 결과보고서』(한국정치학회). 423-505.
- 박인호·임종호·박민규. 2019. 「선거여론조사 신뢰성 제고 조사기법 및 조사환경 개선방안 연구」 2019년 중앙선거여론조사심의위원회 연구용역 과제 보고서(한국통계학회). 1-80.
- 박종희. 2021. “의뢰기관과 정치여론조사.” 한국조사연구학회 춘계학술회의. 서울. 5월.
- _____. 2013. “제18대 대선 여론조사에서 나타난 조사기관 편향” 『조사연구』 14.1. 1-30.
- 심형인. 2017. “설문조사에서의 질문문항순서 및 선택지순서 효과” 『한국행정학보』 51.1. 145-169.
- 이소영·변영학·장우영·조희정. 2017. 「유·무선전화 비율 등 바람직한 여론조사 방법에 관한 연구」 2017년 중앙선거여론조사심의위원회 연구용역 결과보고서(대한정치학회). 1-139.
- 이상신·민태은·윤광일·구본상·Antonio Fiori·Marco Milani. 2023(근간). 「KINU 통일의식조사[1/3년차] 최종보고서」 통일연구원.
- 장덕현. 2021. “정치관심도와 여론조사 표본대표성.” 한국조사연구학회 춘계학술회의. 서울. 5월.
- 장덕현·홍영택·조성겸. 2014. “가중방법으로 선거여론조사의 편향을 얼마나 더 줄일 수 있다.” 『조사연구』 15.2. 105-121.
- 장승진. 2020. “유권자들은 총선에서 누구를 언제 심판하는가?” 『한국정치학회보』 50. 4. 151-169.
- _____. 2016. “제20대 총선의 투표선택: 회고적 투표와 세 가지 심판론.” 『한국정치학회보』 50. 4. 151-169.
- 조성겸·오승호. 2021. 「웹기반 선거여론조사의 쟁점과 신뢰성 제고 방안 연구」 2021년도 중앙선거여론조사심의위원회 연구용역 결과 보고서(한국조사연구학회).
- 조진만·장승진·구본상·강우창. 2021. 「선거여론조사기준 개정방안 연구」 2021년도 중앙선거여론조사 연구용역보고서(한국 정당학회). 1-113.
- 정한울. 2020. “여야 심판론의 관점에서 본 21대 총선과 보수 혁신의 딜레마” 『EAI 워킹페이퍼』 제2020-05호. 1-21.
- _____. 2019. “선거여론조사의 신뢰도 위기와 여론으로 본 대안.” 『중앙선거여론조사심의위원회 선거여론조사 객관성·신뢰성 제고를 위한 세미나』.
- _____. 2016. “외주민주주의 시대의 여론조사: 여론조사가 투표선택에 미친 영향” 『한국정당학회보』 15.1. 63-105.
- _____. 2013. “18대 대선에서 중도는 어떻게 선택했는가” 이창곤·한귀영 편 『18 그리고 19: 18대 대선으로 본 진보개혁의 성찰과 길』 90-136.
- _____. 2011. “50퍼센트 지지율 대통령이 왜 심판 받았을까” 이내영·임성학 공편 『변화하는 한국유권자4: 패널 조사를 통해 본 2010 지방선거』 95-124.
- 중앙선거여론조사심의위원회. 2022a. 「2022년 양대선거 선거여론조사 백서」.
- _____. 2022b. 「선거여론조사 가이드북」.
- _____. 2020. 「제21대 국회의원선거 선거여론조사 백서」.
- _____. 2018. 「제7회 전국동시지방선거 선거여론조사 백서」.
- _____. 2017. 「제19대 대통령선거 선거여론조사 백서」.
- 중앙선거여론조사공정심의위원회. 2016. 「제20대 국회의원선거 선거여론조사 심의백서」.
- 최종호·김태균·백다예·왕정하·한강욱. 2022. “조사방법 차이가 정치적 태도 조사결과에 미치는 영향: 중앙여심위 선거여론

- 조사 통합자료 분석.” 『한국정당학회보』 21.3. 5-40.
- 한국갤럽. 2023. “표본추출 프레임 변경 안내” 『한국갤럽 데일리오피니언』 제549호 (2023년 7월1주), 한국갤럽 홈페이지 <https://www.gallup.co.kr/gallupdb/reportContent.asp?seqNo=1403#Z2> (검색일: 2023/10/8).
- _____. 2022. “조사담(調査談): 무선전화 RDD, 가상번호, 알뜰폰: 이용자 특성과 성향” (2022/01/18), 한국갤럽 홈페이지 <https://www.gallup.co.kr/gallupdb/columnContents.asp?seqNo=133> (검색일: 2023/10/8).
- _____. 2021. “조사담(調査談): 정치 관심도와 여론조사” (2021/03/18), 한국갤럽 홈페이지 <https://www.gallup.co.kr/gallupdb/columnContents.asp?seqNo=127> (검색일: 2023/10/8).
- 허석재·정한울. 2019. “태도 상충성과 분할투표: 2016년 20대 국회의원 총선거 분할투표 분석” 『의정논총』 14. 1. 137-162.
- AAPOR Education/Resources. “Response Rates—An Overview.” <https://www-archive.aapor.org/Education-Resources/For-Researchers/Poll-Survey-FAQ/Response-Rates-An-Overview.aspx> (검색일 2023/10/20) .
- Alvarez, R. Michael and John Brehm. 2002. *Hard Choices, Easy Answers: Values, Information, and American Public Opinion*. Princeton: Princeton Univ. Press.
- Callegaro, Mario, Katja Lozar Manfreda, Vasja Vehovar, . 2015. *Web Survey Methodology: Research Methods for Social Scientists*. SAGE Publications. Kindle Edition.
- Deane, Claudia, Courtney and Scott Keeter. 2019. “A Field Guide to Polling: Election 2020 Edition.” Pew Research Center. <https://www.pewresearch.org/methods/2019/11/19/a-field-guide-to-polling-election-2020-edition/> (검색일 2023/10/20).
- Erikson, Robert A. and Kent L. Tedin, 2005. *American Public Opinion: It's Origins, Contents and Impact*. New York: Pearson Longman.
- Goidel, Robert K., Todd G. Shields, and Mark Peffley. 1997. “Priming Theory and RAS Models: Toward an Integrated Perspective of Media Influence.” *American Politics Quarterly*. 25. 3. 287-318.
- Groves, Robert M. and Emilia Peytcheva. 2008. “The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis.”
- Groves, Robert M., Floyd J. Fowler Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2004. *Survey Methodology*. 김석호·이정환·황명진·고성호 역. 『서베이 방법론』 (서울: 터닝포인트, 2015).
- Hughes, Adam and Bradley Jones. 2019. “‘Good jobs’ vs. ‘jobs’: Survey Experiments Can Measure the Effects of Question Wording—and More” Pew Research Center. <https://www.pewresearch.org/short-reads/2019/01/29/good-jobs-vs-jobs-survey-experiments-can-measure-the-effects-of-question-wording-and-more/> (검색일 2023/10/19).
- Lavine, Howard G., Christopher D. Johnston, and Marco R. Steenbergen. 2012. *The Ambivalent Partisan: How Critical Loyalty Promotes Democracy*. NY: Oxford University Press.
- Lee, Seulki and Hyunsub Kum. 2022. “설문조사에 있어서 문항 순서효과에 대한 연구.” 『조사연구』 22, 2. 99-130.
- Mercer, Andrew, Courtney Kennedy, and Hannah Klein. 2018. “For Weighting Online Opt-In Samples, What Matters Most?” Pew Research Center. <https://www.pewresearch.org/methods/2018/01/26/for-weighting-online-opt-in-samples-what-matters-most/>(검색일 2023/10/19).
- Morton, Rebecca and Kenneth Williams. 2010. *The Methodology of Experimental Political Science and the Study of Causality: From Nature to the Lab*. Cambridge: Cambridge University Press.
- Peytchev, Andy and Emilia Peytcheva. 2017. “Reduction of Measurement Error to Survey Length: Evaluation of the Split Questionnaire Design Approach.” *Survey Research Methods*. 11. 4. 361-368.
- R Core Team (2022). R: A Language and environment for statistical computing. (Version 4.1) [Computer software]. Retrieved from <https://cran.r-project.org>. (R packages retrieved from CRAN snapshot 2023-04-07).
- Schriesheim, Chester A., Joshua B. Wu, and Cecily D. Cooper. 2011. “A Two-study Investigation of Item

Wording Effects on Leader-Follower Convergence in Descriptions of the Leader-Member Exchange(LMX) Relationship.” *The Leadership Quarterly*. 22. 881-892.

Singleton, Jr. Royce A. and Bruce C. Straits, 1999. *Approches to Social Research: Third Edition*. New York: Oxford University Press.

The Jamovi project (2023). jamovi. (Version 2.4) [Computer Software]. Retrieved from <https://www.jamovi.org>.

Zaller, John R. 1992. *The Nature and Origins of Mass Opinion*. Cambridge: Cambridge University Press.

한국리서치 주간리포트

여론속의

여輿論論

Hankook  Research

