연구논문 「환경정책」 제34권 제3호 2026. 9: 139-179
DOI http://dx.doi.org/10.15301/jepa.2026.34.3.139
pISSN 1598-835X eISSN 2714-0601

CBD 제7차 국가보고서의 진척도-증거 간극: 134개국 비교와 한국의 정책과제*

The Progress-Evidence Gap in CBD Seventh National Reports: A Comparison of 134 Countries and Policy Priorities for Korea

이후승**
Who-Seung Lee

요약: 쿤밍-몬트리올 글로벌 생물다양성 프레임워크(Kunming-Montreal Global Biodiversity Framework, 이하 KMGBF) 채택 이후 제출된 CBD 제7차 국가보고서는 각국이 국가 목표 이행을 보고한 최초의 공식 자료이나, 목표 설계·모니터링 역량·판정기준이 달라 자체평가 진척도를 직접 비교하기 어렵다. 본 연구는 134개국 보고서에서 국가 목표 3,120건을 추출하여 국가생물다양성전략(NBSAP) 정렬, 자체평가 진척도, 이를 뒷받침하는 증거수준(Evidence level)을 비교하였다. 응답국의 98.1%가 NBSAP를 개정했거나 개정 중이었으나, 평가 대상 국가 목표의 40.8%만 '궤도 또는 달성'으로 보고되었고, 진척도 근거의 69.6%는 활동 수준 이하에, 생태적 상태변화 이상은 12.5%에 그쳤다. 진척도와 증거수준은 국가 단위에서 음의 상관을 보였으나, 개별 목표 단위에서는 상관관계가 낮았다. 증거수준은 실제 성과가 아니라 모니터링 역량·보고 관행을 함께 반영하는 보고 기반 프록시이므로, 이 관계는 자체평가 등급의 단순 서열화가 이행성과를 왜곡할 수 있음을 시사한다. 한국은 전체 및 고소득국 평균을 상회하는 상세한 보고서를 제출했으나, 국가 목표의 약 81%가 생태적 상태변화 이상의 증거에 이르지 못해 결과 중심 평가로의 전환 과제가 뚜렷하다.

핵심주제어: 쿤밍-몬트리올 글로벌 생물다양성 프레임워크, 제7차 국가보고서, 증거수준, 이행평가, 진척도-증거 간극

Abstract: The CBD Seventh National Reports (7NR) are the first official documents in which Parties report progress on national targets following the Kunming-Montreal Global Biodiversity Framework (KMGBF), yet differences in target design, monitoring capacity, and progress criteria make self-assessed progress difficult to compare directly. This study computationally extracted 3,120 national targets from the reports of 134 countries and compared National Biodiversity Strategy and Action Plan (NBSAP) alignment, self-assessed progress, and the evidence level supporting it. Although 98.1% of reporting countries had revised or were revising their NBSAP, only 40.8% of scored targets were rated 'on track' or 'achieved'; 69.6% of the underlying evidence remained at or below the activity level, and only 12.5% reached ecological state change or above. Progress and evidence level were negatively correlated at the country level (Spearman ρ=-0.196, p=0.043) but were weak and not robust at the individual-target level. Because the evidence level is a reporting-based proxy reflecting monitoring capacity and reporting practices rather than

* 본 논문은 "생태계 계정 구축 및 생물다양성 주류화 정책 개발(RE2026-09)" 과제의 지원을 받아 수행됨.

** 한국환경연구원 자연환경연구실 실장

actual performance, this relationship suggests that simple ranking by self-assessed ratings can distort implementation performance. Korea submitted a detailed report exceeding both the overall and the high-income country means, yet about 81% of its targets did not reach evidence of ecological state change or above, underscoring the need to shift toward outcome-based evaluation.
Key Words: Kunming-Montreal Global Biodiversity Framework, CBD Seventh National Report, Evidence Level, Implementation Review, Progress-Evidence Gap

I. 서론

생물다양성은 인간의 생존과 경제활동을 떠받치는 기반이지만(Costanza et al., 1997, pp.253-260; TEEB, 2010), 인간 활동으로 유례없는 속도로 감소하고 있으며(IPBES, 2019), 2020년까지의 Aichi 목표는 어느 것도 완전히 달성되지 못하였다(Butchart et al., 2010, pp.1164-1168; Secretariat of the CBD (이하 SCBD), 2020). 이러한 반성 위에서 2022년 제15차 생물다양성협약 당사국총회(COP15)에서 채택된 쿤밍-몬트리올 글로벌 생물다양성 프레임워크(Kunming-Montreal Global Biodiversity Framework, 이하 KMGBF)는 2050년 장기목표와 2030년까지 달성해야 할 23개 행동목표를 제시하고(CBD, 2022a, Decision 15/4), 이행상황을 점검하기 위한 모니터링 프레임워크(CBD, 2022b, Decision 15/5)와 계획·모니터링·보고·검토(PMRR) 메커니즘(CBD, 2022c, Decision 15/6)을 함께 채택하였다. KMGBF는 각 당사국에 국가생물다양성전략 및 행동계획(National Biodiversity Strategy and Action Plan, 이하 NBSAP)의 개정, 국가목표 수립, 지표 기반 모니터링, 그리고 주기적 이행평가를 요구한다.

생물다양성협약(Convention on Biological Diversity, 이하 CBD) 제7차 국가보고서는 KMGBF 채택 이후 당사국이 국가목표의 이행상황을 종합적으로 보고하는 최초의 공식 자료로서, 2030년 글로벌 검토와 목표 달성 가능성을 평가하는 핵심 근거가 된다. 그러나 국가보고서는 각국의 자체평가에 기반하므로, 목표 설정방식, 이행기간, 모니터링 역량, 지표 가용성, 진척도 판정기준이 서로 다르다. 어떤 국가는 생태계 면적·종 개체군·위협요인의 변화를 정량적

으로 보고하는 반면, 다른 국가는 법령 제정·계획 수립·사업 시행·교육인원 등 정책활동과 산출을 주된 성과로 제시한다. 따라서 동일한 진척도 등급이라도 이를 뒷받침하는 증거의 수준과 정책적 의미는 크게 다를 수 있다.

한국도 『제5차 국가생물다양성전략(2024-2028)』을 KMGBF와 연계하여 수립하고, 부처별 이행실적과 국가목표별 진척도를 제7차 국가보고서에 제시하였다. 향후 한국의 생물다양성 정책은 국제목표와의 형식적 정렬을 넘어, 국가목표별 기준선·성과지표·모니터링 자료·정책환류 체계를 강화해야 한다. 이를 위해서는 한국의 이행평가 체계를 다른 국가와 비교하여 상대적 강점과 한계를 진단할 필요가 있다.

본 연구는 제7차 국가보고서를 제출한 다수 당사국을 비교분석하여 KMGBF 국가 이행평가의 구조와 특성을 파악하고, 이를 바탕으로 한국의 생물다양성 정책 및 국가보고 체계 개선 방안을 제시하는 것을 목적으로 한다. 이를 위해 본 연구는 다음의 질문에 답하고자 한다. 첫째, 각국은 KMGBF의 23개 글로벌 목표를 국가생물다양성전략과 국가목표에 어느 정도 반영하고 있는가? 둘째, 국가 목표별 자체평가 진척도는 어떻게 분포하며, 목표·지역·소득수준에 따라 어떠한 차이를 보이는가? 셋째, 진척도 판단은 정책활동·행정적 산출·정책적 결과·생태적 상태변화 가운데 어느 수준의 증거에 기반하는가? 넷째, 국가가 보고한 진척도와 이를 뒷받침하는 증거의 강도는 어느 정도 정합적인가? 다섯째, 한국의 국가목표 설정, 진척도 평가, 증거수준은 다른 국가와 비교할 때 어떠한 상대적 강점과 한계를 보이는가? 이상의 질문은 각각 전략 정렬, 진척도 분포, 증거기반, 진척도-증거 정합성, 그리고 한국의 상대적 위치라는 다섯 국면에 대응한다.

II. 이론적·제도적 배경

1. KMGBF의 이행 및 모니터링 체계

KMGBF는 2050년 4대 목표(A-D)와 2030년 23개 행동목표로 구성되며,

당사국은 이를 자국 여건에 맞추어 국가목표로 전환한다. 모니터링 프레임워크(Decision 15/5)는 헤드라인 지표, 구성지표, 보완지표, 이진(binary)지표로 구성된 위계적 지표체계를 제시하고, 당사국은 국가목표별로 이들 지표를 선택·연계하여 진척도를 보고한다. 글로벌 검토는 국가보고서를 종합하여 2026년과 2029년에 이루어진다.

2. CBD 제7차 국가보고서의 기능과 한계

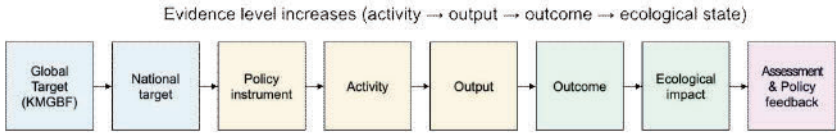
제7차 국가보고서는 CBD 온라인 보고도구(Online Reporting Tool, 이하 ORT)의 공통 양식으로 작성된다. 보고서는 (Ⅰ)보고서 작성과정, (Ⅱ)NBSAP 개정 현황, (Ⅲ)국가 목표별 이행조치·진척도·지표, (Ⅳ)글로벌 목표(A-D 및 23개 행동목표)에 대한 국가 기여 평가, (Ⅴ)결론으로 구성된다. 국가 목표별 진척도는 '목표 달성', '달성 궤도', '진전은 있으나 속도가 불충분함', '유의한 변화 없음' 등 표준 범주 중에서 국가가 선택한다. 그러나 이 자체평가는 국가의 목표 야심수준, 모니터링 역량, 평가문화에 좌우되므로, 보고된 진척도를 곧 실제 생물다양성 성과로 간주하거나 국가 간 단순히 서열화하면 이행수준을 왜곡할 수 있다.

3. 환경정책 성과평가와 증거수준

환경정책의 성과관리와 모니터링·평가(M&E)에서 성과는 일반적으로 투입–활동–산출–결과–영향의 연쇄(결과사슬, results chain)로 이해된다. 예산·인력은 투입, 법령 제정·사업 시행·조사 수행은 활동, 보호지역 지정면적·복원면적·교육인원은 산출, 압력요인의 감소는 결과, 종 개체군 회복·멸종위험 감소·생태계 건전성 향상은 생태적 영향에 해당한다. 결과기반관리(results-based management)의 핵심 통찰은, 정책 초기에는 활동·산출이 중요한 선행조건이지만 이를 최종 성과와 동일시하면 정책효과가 과대평가된다는 것이다. 따라서 성과보고의 타당성은 '어떤 등급을 부여했는가' 뿐 아니라 '그 등급이 결과사슬의 어느 단계 증거에 근거하는가'에 달려 있다. 국제

평가·보고체계 가운데 이러한 증거의 위계와 불확실성을 명시적으로 다룬 대표적 사례로 IPCC의 신뢰도(confidence)·가능성(likelihood) 언어를 들 수 있으며(Mastrandrea et al., 2010; Schneider and Moss, 1999), 그 적용과 한계도 검토되어 왔다(Janzwood, 2020, pp.1655- 1675; Mach, Mastrandrea, Freeman and Field, 2017, pp.1-14). 반면 CBD 제7차 국가보고서의 진척도 자체평가에는 이에 상응하는, 등급과 그 증거기반을 함께 보고하도록 하는 보정된 체계가 부재하다. 본 연구는 결과사슬 관점을 국가보고서 분석에 적용하여, 국가가 제시한 진척도가 어느 수준의 증거에 기반하는지를 체계적으로 평가한다〈그림 1〉.

〈그림 1〉 연구의 개념적 분석틀: 글로벌 목표에서 정책환류에 이르는 결과사슬의 증거수준



4. 선행연구와 본 연구의 차별성

본 연구와 관련된 선행연구는 세 문헌군으로 정리할 수 있다. 첫째 문헌군은 생물다양성 손실과 국제 이행평가에 관한 연구, 둘째 문헌군은 평가·보고체계가 증거와 불확실성을 다루는 방식에 관한 연구, 셋째 문헌군은 다수 국가의 정책문서를 체계적으로 비교분석하는 방법에 관한 연구이다.

첫째 문헌군을 보면, 생물다양성 손실과 그 이행평가에 관한 국제 문헌은 꾸준히 축적되어 왔다. 세계 생태계 서비스의 경제적 가치가 방대함이 일찍이 제시되었고(Costanza et al., 1997, pp.253-260; TEEB, 2010), IPBES (2019)는 인간활동에 의한 유례없는 생물다양성 감소를 경고하였다. 그러나 2020년까지의 Aichi 목표는 20개 중 어느 것도 완전히 달성되지 못하였으며(Butchart et al., 2010, pp.1164-1168; SCBD, 2020), 그 원인의 하나로 목표의 측정가능성 부족과 지표·자료의 공백이 지적되었다. KMGBF(CBD, 2022a, Decision 15/4)와 그 모니터링 프레임워크(Decision 15/5)는 이러한

반성 위에서 지표 기반의 이행점검을 강화하였다. 최근 기업·금융 부문의 생물다양성 영향 공개를 다루는 자연관련 재무정보공개(TNFD, 2023) 같은 제도적 흐름도 목표 14·15·19의 이행 맥락을 형성한다.

둘째 문헌군은 성과관리와 국제 평가·보고체계가 '증거와 불확실성을 어떻게 다루는가'를 논한다. 결과기반 관리와 M&E 문헌은 산출지표와 결과·영향지표를 구분하고 후자에 근거한 성과판정을 지향해 왔으며, 그 대표적 제도화 사례로 IPCC는 증거의 양·질과 일치도에 기반한 신뢰도 언어를 표준화하였다(Janzwood, 2020, pp.1655-1675; Mach et al., 2017, pp.1-14; Mastrandrea et al., 2010; Schneider and Moss, 1999). 국내에서도 IPCC 보고서의 불확실성·확실성 표현 변화가 분석되었다(김선회·윤순진, 2026, pp.1-38; 윤순진·김선회, 2024, pp.799-814). 이들 연구의 공통된 통찰은 증거의 수준과 그 명시적 소통이 보고의 신뢰성을 좌우한다는 점이다(Oppenheimer, O'Neill, Webster and Agrawala, 2007, pp.1505-1506; Oreskes, 2004). 그러나 CBD 제7차 국가보고서의 진척도 자체평가에는 이에 상응하는, 등급과 증거기반을 함께 요구하는 체계가 없어, 진척도와 그 증거기반의 관계 자체가 실증 분석의 대상이 된다.

셋째 문헌군을 보면, 다수 국가의 정책문서를 비교분석하는 접근은 환경정책 연구에서 확립된 방법이다. 국내에서도 주요국의 산림부문 적응정책을 공통 항목별로 5점 척도로 평가하고 한국을 대조한 연구가 이루어졌으며(홍민아·송민경, 2026, pp.67-82), 국가 범주화의 방법론적 쟁점(Victor, Gerlagh and Baiocchi, 2014, pp.34-36)과 다국가 비교분석(Hase, Mahl, Schäfer and Keller, 2021; Schmidt, Lvanova and Schäfer, 2013, pp.1233-1248)도 논의되어 왔다. 대규모 문서를 체계적으로 코딩하는 데에는 계량 텍스트 분석이 널리 활용되어 왔으며(Blei, Ng and Jordan, 2003, pp.993- 1022; Denny and Spirling, 2018, pp.168-189; Roberts, Stewart, Tingley, Lucas, Leder-Luis, Gadarian et al., 2014, pp.1064-1082), 본 연구는 이러한 방법적 논의를 참고하여 다수 국가의 보고서를 공통 항목으로 구조화하였다. 그럼에도 KMGBF 채택 이후 제출된 제7차 국가보고서를 대상

으로 진척도와 그 증거기반, 보고 거버넌스를 함께 비교한 연구는 아직 찾기 어렵다.

이에 본 연구는 세 가지 점에서 기존 연구와 구별된다. ①특정 국가나 목표에 한정하지 않고 제출된 국가보고서 전체를 비교 대상으로 삼는다. ②국가가 보고한 진척도를 곧바로 성과로 간주하지 않고, 그 진척도를 뒷받침하는 증거의 강도를 별도의 변수로 측정하여 양자의 관계를 규명한다. ③이러한 국제 비교의 맥락 속에서 한국의 이행평가 체계를 진단하여 국내 정책 개선 과제와 연결한다.

Ⅲ. 연구자료 및 방법

1. 분석자료와 분석단위

본 연구는 CBD 온라인 보고도구(ORT) 제7차 국가보고 포털(https://ort.cbd.int/national-reports/nr7)에 등록된 134개국의 제7차 국가보고서를 분석 대상으로 한다.[1] 각 보고서는 ORT의 공통 양식에 따라 작성되어 국가정보, NBSAP 개정 현황, 국가목표별 이행조치·진척도·지표, 그리고 글로벌 목표에 대한 국가 기여 평가를 포함한다. 본 연구는 각 보고서에서 이들 공통 항목을 추출하여 국가·국가목표·글로벌 목표의 세 수준으로 구조화하였으며, 최종 데이터셋은 국가목표 3,120건(국가당 중앙값 23개, 범위 1-104)을 포함한다.

2. 분석틀과 변수

본 연구의 분석틀은 앞서 제시한 결과사슬〈그림 1〉을 국가보고서 분석에 맞추어 조작화(operationalization)한 것으로, 국가 이행평가를 다섯 개 분석 영역으로 구분하고 각 영역의 핵심질문과 측정변수를 설정하였다〈표 1〉. 구

1) 2026년 4월 30일 기준.

체적으로 전략 정렬성은 KMGBF가 국가생물다양성전략과 국가목표에 반영된 정도를, 진척도는 국가가 부여한 자체평가 등급을, 증거수준은 그 등급을 뒷받침하는 근거의 강도를, 지표 연계는 측정 가능한 지표의 존재와 유형을, 이행 맥락은 지역·소득·OECD 가입 여부 등 구조적 조건을 포착한다. 이 가운데 진척도와 증거수준이 본 연구의 핵심 변수이며, 각 변수의 조작적 정의는 이어지는 절에서 상술한다.

〈표 1〉 국가 이행평가 분석틀과 주요변수

분석영역	핵심질문	주요 변수
전략 정렬성	KMGBF가 국가전략에 반영되었는가	NBSAP 개정 여부, 국가목표 수, KMGBF 목표 연계
진척도	목표 이행을 어떻게 자체평가하는가	5단계 표준화 진척도(0–4)
증거수준	진척도는 어떤 근거로 뒷받침되는가	활동·산출·결과·생태적상태·목표대비 평가(0–5)
지표연계	측정 가능한 지표가 있는가	헤드라인/이진/구성지표, GBF 지표코드 연계
이행맥락	지역·소득 등 구조적 조건	UN지역, World Bank 소득군, OECD 여부

1) 진척도 표준화

국가보고서의 진척도 범주를 결과지향 순서로 표준화하였다〈표 2〉. 점수는 국가의 자체평가를 수치화한 것으로서 객관적 생물다양성 성과 순위를 의미하지 않는다. 본 코퍼스(corpus)에서는 '목표에서 멀어짐'과 '달성 초과 궤도'는 관측되지 않았다.

〈표 2〉 국가목표 진척도 표준화 기준(실측 범주 기준)

점수	표준화 범주(ORT 원문)	비고
4	Achieved / On track to exceed target	달성 또는 초과 달성
3	On track to achieve target	현 추세 유지 시 달성 가능
2	Progress towards target but at an insufficient rate	진전 있으나 속도 불충분
1	No significant change	뚜렷한 진전 없음
0	Moving away from target	목표와 반대 방향
NA	Unknown / Not applicable	평가불가·미보고

2) 증거수준

진척도 서술에 포함된 근거를 결과사슬에 따라 0-5단계로 코딩하였다〈표 3〉. 코딩은 각 국가목표의 '진척도 요약' 서술과 첨부지표를 대상으로 자동 텍스트 규칙(정량단위, 생태적 상태 용어, 기준선·목표값·시계열 표현의 탐지)으로 1차 부여한 뒤, 한국 등 표본에 대해 연구자가 수기로 검증하였다. 여기서 강조할 것은, 증거수준은 국가가 진척도의 근거로 결과사슬의 '어느 단계'에 위치한 증거를 제시했는가를 포착하는 보고 기반 프록시(reporting-based proxy)라는 점이다. 즉 증거수준은 제시된 과학적 자료의 정확성·품질이나 실제 생태 성과의 크기를 측정하는 것이 아니며, 보고 서술의 상세성·완결성에 부분적으로 좌우된다. 따라서 높은 증거수준이 곧 우수한 생물다양성 성과를, 낮은 증거수준이 곧 부실한 이행을 의미하지는 않는다.

〈표 3〉 진척도 평가의 증거수준

수준	증거유형	판정기준
0	근거없음	진척도만 선택, 실질 서술 없음
1	정성적 활동	법·계획·협의체·사업의 존재 서술
2	정량적 산출	면적·예산·사업량·참여자 등 산출값
3	정책·사회적 결과	압력·행태·제도 운영 성과의 변화
4	생태적 상태변화	종·생태계·유전다양성의 변화
5	목표 대비 성과평가	기준선·목푯값·시계열을 종합

3) 증거수준 프록시의 타당도 검증

증거수준 자동 코딩의 타당도를 검증하기 위해, 자동 증거수준(1-5)별로 9개씩, 총 45개 국가목표를 층화표집하고,[2] 연구진이 코드북에 따라 각 목표의 진척도 서술을 0-5단계로 수기 코딩하였다. 코더는 각 국가목표의 '진척도 요약' 서술과 첨부지표를 코드북〈표 3〉의 단계별 판정기준에 따라 검토하여 0-5단계를 부여하였고, 자동 프록시 값을 참조하지 않은 상태에서 코딩한 뒤

2) 영어 서술 목표에 한정.

판정이 모호한 사례는 코드북 기준을 재적용하여 확정하였다. 초기의 단순 키워드 프록시는 'species', 'ecosystem', 'forest', 'restoration' 등 생태 관련 명사를 상태변화 여부와 무관하게 포착하여 증거수준을 체계적으로 과대추정하였다.[3] 이에 생태 상태변화 용어가 변화·추세·수치 표현과 근접(±180자)하여 등장할 때에만 생태적 증거로 인정하도록 규칙을 정교화하였다.

정교화한 프록시는 수기 코딩과 비교하여 정확일치 53.3%, ±1단계 이내 82.2%, 편향 -0.27단계, 이차 가중 κ=0.50, Spearman ρ=0.58 (p=0.002)의 일치를 보였다〈표 4〉. 4단계와 5단계의 세부 구분은 신뢰도가 상대적으로 낮아, 증거수준은 개별 목표의 정밀 척도가 아니라 활동·산출과 생태적 증거를 구분하는 집합적 지표로 해석한다. 정교화 과정에서 생태적 증거(4단계 이상)의 비율은 자동 코딩의 20.0%에서 12.5%로 하향 조정되었는데, 이는 초기 자동 코딩이 생태적 증거를 과대추정하였음을 의미한다. 따라서 국가 이행보고가 활동·산출에 편중되어 있다는 본 연구의 해석은 보수적으로 성립한다.

〈표 4〉 증거수준 프록시의 일치도 검증

검증	정확일치	±1단계 이내	이차가중 κ	편향 (자동−수기)
초기 키워드 프록시 vs 수기 (n=45)	55.6%	75.6%	0.41	+0.76
정교화 프록시 vs 수기 (n=45)	53.3%	82.2%	0.50	−0.27
정교화 프록시 vs. LLM 코더 (n=312)	61.5%	84.0%	0.36	+0.35

주: 편향이 양(+)이면 자동 프록시가 증거수준을 과대추정함을 뜻함. n=45는 연구자의 수기 검증, n=312는 코드북 기반 LLM 코더 재코딩 검증임.

나아가 완성된 코드북에 기반하여, 전체 국가목표의 10%에 해당하는 312개 목표를 세 벤더 계열의 5개 코드북 기반 대규모 언어모델(LLM) 코더가 자동값을 참조하지 않고 독립적으로 재코딩하였다.[4] 자동 프록시와 이들 LLM 코더의 일치는 정확일치 61.5%, ±1단계 이내 84.0%, 단순 Cohen's κ=0.33, 이차 가중 κ=0.36(95% 신뢰구간 0.24–0.48)으로 '보통(fair)' 수준이었다〈표

3) 수기 대비 편향 +0.76, 이차 가중 κ=0.41; 생태증거 판정의 정밀도 0.39.
4) 앞의 45개 표본은 직접 수기 코딩한 검증이다.

4〉. LLM 코더는 자동값보다 체계적으로 보수적인 수준을 부여하여(편향(즉, 자동값-LLM 다수결) +0.35단계) 활동·산출 편중이라는 해석을 재확인하였다. 이차 가중 κ 0.36은 내용분석의 일반적 해석 기준에서 '보통(fair)' 수준(0.21–0.40)에 해당하며, 이는 6단계(0–5)의 세밀한 순서형 분류에서 코더 간 완전일치가 어렵다는 점을 반영한다. 이 결과는 자동 프록시가 개별 목표의 세밀한 수준 판정에서 코드북 기반 재코딩과 완전히 일치하지 않음을 시사하며,[5] 이에 본 연구는 증거수준을 총량·방향(활동·산출 대 생태증거)의 해석에 한정하여 사용한다. 자동 프록시와 독립 코더 비교에서 '활동·산출 중심'이라는 전체 분포의 방향은 일관되게 확인되었다. 다만 세 벤더 계열의 5개 LLM 모델의[6] 다수결 합의값을 사용한 부표본에서는 목표 단위의 음의 상관이 재현되지 않았으므로(Ⅳ장 7절), 진척도–증거 관계의 핵심 해석은 전체 자료를 이용한 국가 단위 분석에 한정한다.

본 연구에서 활용한 대규모 언어모델(LLM) 코더의 구성과 근거자료는 다음과 같다. 코더로는 서로 다른 세 벤더 계열의 5개 모델(Anthropic Claude Opus·Sonnet·Haiku, OpenAI GPT-5.6, Google Gemini 3.1 Pro)을 사용하였다. 각 모델에는 코드북의 0–5단계 정의와 판정 예시를 담은 동일한 영어 코딩 지시문을 제시하고, 그 아래에 자동 프록시 값을 제거한 블라인드 상태의 312개 목표 서술(각 국가목표의 '진척도 요약' 원문과 첨부지표에서 추출)을 입력하여, 모델이 목표별로 0–5단계 코드만 산출하도록 하였다. 각 모델은 서로의 결과를 참조하지 않고 독립적으로 코딩하였으며, 5개 모델의 최빈값(다수결)을 기준값(gold)으로 삼아 규칙 기반 자동 프록시와의 일치도 및 코더 간 일치도를 산정하였다.

아울러 동일한 312개 표본을 코드북에 따라 복수의 독립 코더가 겹쳐 코딩하여 코드북의 재현성을 점검하였다. 5개 LLM 코더 간 Fleiss κ는 0.70('상당한 일치') 수준이었으며, 이 다수결 결과에 대해 규칙 기반 자동 프록시는 이차 가중 κ 0.37의 '보통' 일치를 보였다. 다만 이 절차는 연구자의 코더 신뢰

5) 저자 수기 코딩 검증과도 일관된다.
6) 세 개 벤더에 속한 5대 LLM 모델(Opus 4.8, Sonnet 5, Haiku 4.5, GPT-5.6, Gemini 3.1 Pro).

도의 대체물이 아니라 '코드북이 서로 다른 코더에 걸쳐 재현되는 정도'를 확인하는 것이며, 자동 프록시는 개별 목표의 정밀 판정이 아니라 총량·방향 파악의 지표로 활용한다. 코더 이질성(동일 모델 대 이종 모델, 벤더 간 차이)에 따른 일치도 변화의 세부와 앙상블 구성은 부록에 제시한다. 개별 목표 수준의 확정적 신뢰도 산정은 서로 독립적인 평가자 코더 2인 이상의 전수 재코딩(중복 10% 이상)을 필요로 하며, 이는 후속 과제로 남는다.

3. 분석절차와 한계

문서의 구조화와 코딩은 계량 텍스트 분석의 절차(Denny and Spirling, 2018, pp.168-189; Roberts et al., 2014, pp.1064-1082)를 준용하였다. 각 보고서에서 (1)NBSAP 개정상태, (2)"National" 표지로 구분되는 국가목표, (3)표준 진척도 라벨, (4)GBF 지표코드에 기반한 목표 연계, (5)증거수준을 추출하였다. 진척도 라벨은 제출 언어와 무관하게 영어 표준 문구로 기록되어 언어와 독립적으로 코딩하였다. 다만 일부 국가(중국어·스페인어 등의 ORT 인터페이스로 제출)에서는 표준 진척도 필드가 추출되지 않아, 진척도 범주 분석은 110개국 및 2,558개 목표를 대상으로 하였다.

각 국가는 유엔 지역그룹(아프리카, 아시아·태평양, 중남미, 서유럽 및 기타, 중·동유럽)과 세계은행 소득군(고소득, 상위중소득, 하위중소득, 저소득), OECD 가입 여부로 분류하였다. 소득군은 세계은행의 회계연도 소득분류를, 지역은 유엔 지역그룹의 구분을 따랐다. 국가 범주화가 지니는 해석적 한계(Victor et al., 2014, pp.34-36)를 고려하되, 분류의 투명성과 재현성을 위하여 분석대상 134개국의 지역·소득군·OECD 여부·NBSAP 개정상태·국가목표 수를 부록 〈표 A1〉에 제시한다.

정량분석은 다음과 같이 수행하였다. 진척도와 증거수준의 관계는 순위상관(Spearman ρ)으로 목표 단위와 국가 단위에서 각각 검토하였다. 두 변수의 결정요인은 세 모형으로 추정하였다. 모형 M1과 M2는 국가 단위 최소자승회귀(OLS)로, 각각 국가별 평균 증거수준과 평균 진척도를 종속변수로 하며, 설명변수로 소득수준(순서형, 저소득 1–고소득 4), 지역(더미변수, 기준

범주=서유럽 및 기타), OECD 가입 여부, 국가목표 수를 포함한다(M2에는 평균 증거수준을 추가). 모형 M3은 목표 단위 진척도를 종속변수로 하고 국가별 임의 절편을 두는 다층모형(multilevel model)으로, 소득수준·지역·증거수준을 설명변수로 하며 급내상관계수(ICC)를 함께 산출하였다.

국가별 국가목표 수의 편차가 크므로, M1·M2는 이분산에 강건한 표준오차(HC3), 국가목표 수를 가중치로 한 가중최소자승(WLS), 국가목표 5개 미만 국가를 제외한 부표본으로 각각 재추정하여 강건성을 확인하였다. 또한 진척도가 순서형 변수임을 고려하여, M3의 선형 다층모형과 별도로 순서형 로짓(cumulative-link) 설정으로도 진척도–증거수준 관계의 방향을 재확인하였다. 다만 이 순서형 로짓은 진척도 척도의 순서형 성격에 대한 보조 분석으로서 국가 임의 효과를 포함하지 않는 통합(pooled) 모형이므로, 국가 간 차이와 국가 내 차이가 혼합된 결과이며 국가 내 연관성을 추정하는 M3의 대체모형은 아니다(민감도 분석 결과는 Ⅳ장 7절 및 부록 참조). 아울러 진척도–증거수준의 음의 관계가 프록시 선택에 얼마나 민감한지를 검토하기 위해, 증거수준을 결과수준 이상(3단계 이상/미만)과 생태적 상태변화 이상(4단계 이상/미만)의 두 문턱으로 이분한 설정과 5개 LLM 코더 합의값을 사용한 부표본에서도 관계를 검토하였다. 모든 정량분석과 시각화는 Python 3.11(pandas, NumPy, SciPy, statsmodels, matplotlib)로 수행하였으며, 코더 간 일치도 히트맵 등 신뢰도 관련 보조 그림은 부록 〈그림 A1〉에 제시한다.

본 연구의 한계는 (1)국가 자체평가 자료라는 점, (2)자동 코딩의 오차(앞의 타당도 검증으로 부분적으로 보완), (3)진척도 표준필드가 추출되지 않는 비영어 인터페이스 보고서, (4)표본 선택편의이다. 특히 CBD 당사국 약 196개국 가운데 본 분석은 제출된 134개국에 한정되며, 미제출국은 자료·보고 역량이 상대적으로 제한된 국가에 편중될 수 있으므로 이러한 선택편의는 결과의 일반화 가능성을 제약한다.

Ⅳ. 134개국 제7차 국가보고서 분석결과

분석결과를 제시하기에 앞서, 본 연구에서 사용하는 주요 지표의 분석모집단(분모)과 정의를 〈표 5〉에 정리한다. 진척도 관련 통계는 라벨이 부여된 목표에서 '평가 불가·미상'을 제외한 채점 대상 2,407개 목표를 기준으로 하고, 증거수준 관련 통계는 전체 국가목표 3,120건을 기준으로 하며, 국가 간 비교가 필요한 경우 국가별 평균을 다시 평균한 값(국가평균)을 사용한다. 서로 다른 분모에서 산출된 수치가 혼동되지 않도록, 이하 본문에서는 각 통계의 기준을 함께 표기한다.

〈표 5〉 주요 결과 지표의 분석모집단과 정의

지표	분석모집단(n)	값	정의·비교
진척도 라벨 부여 목표	전체 라벨 목표	2,558개	진척도 범주가 기재된 목표
진척도 채점 대상	채점 목표	2,407개	라벨 목표에서 평가불가(90)·미상(61) 제외
"궤도 또는 달성" 비율	채점 목표(2,407)	40.8%	진척도 3-4단계 목표 비율
"불충분 이하" 비율	채점 목표(2,407)	59.2%	진척도 0-2단계 목표 비율
평균 진척도(목표단위)	채점 목표(2,407)	2.42	목표별 진척도(0-4)의 단순 평균
평균 진척도(국가평균)	국가(110)	2.48	국가별 평균 진척도를 다시 평균
활동수준 이하(0-1) 비율	전체 목표(3,120)	69.6%	증거수준 0-1단계 목표 비율
생태증거 이상(4-5) 비율	전체 목표(3,120)	12.5%	증거수준 4-5단계 목표 비율
생태증거 이상(4-5) 비율	국가평균	13.7%	국가별 4-5단계 비율의 평균
평균 증거수준(목표단위)	전체 목표(3,120)	1.50	목표별 증거수준(0-5)의 단순평균
평균 증거수준(국가평균)	국가(134)	1.52	국가별 평균 증거수준을 다시 평균

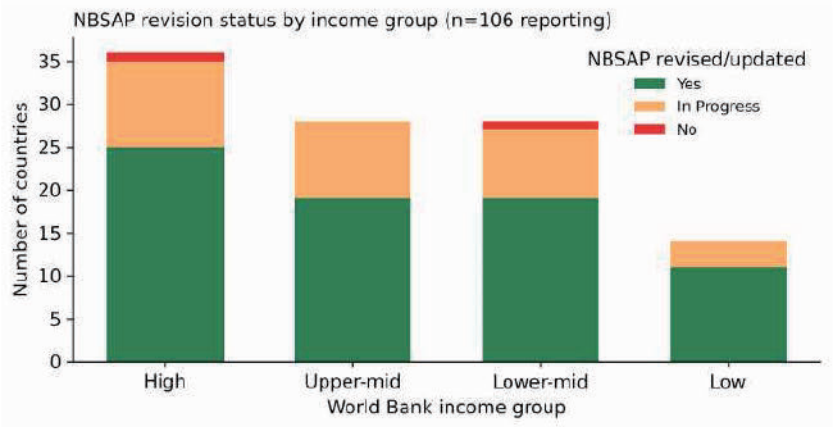
주: 진척도 통계는 채점 대상(2,407개), 증거수준 통계는 전체 국가목표(3,120개)를 기본 분모로 함. "국가평균"은 국가 간 규모차이를 보정하기 위해 국가별 값을 다시 평균한 것임.

1. NBSAP 개정 및 국가목표 정렬 현황

NBSAP 개정상태가 확인된 106개국 가운데 KMGBF에 맞추어 NBSAP를 개정 완료했다고 보고한 국가는 74개국(69.8%), 개정 진행 중은 30개국이었으며, 미개정은 2개국에 그쳤다. 즉 응답국의 98.1%가 전략 정렬에 착수하였다. 국가목표 수는 국가당 중앙값 23개로 KMGBF 23개 목표와 대체로 대응

하였으나, 벨기에·독일·스웨덴 등은 하위목표를 세분하여 40개 이상을 설정한 반면 일부 국가는 6개 이하로 설정하는 등 편차가 컸다. 소득군별로 개정 완료율에 뚜렷한 위계는 나타나지 않아, 전략의 형식적 정렬은 소득수준과 무관하게 광범위하게 진행되었다〈그림 2〉.

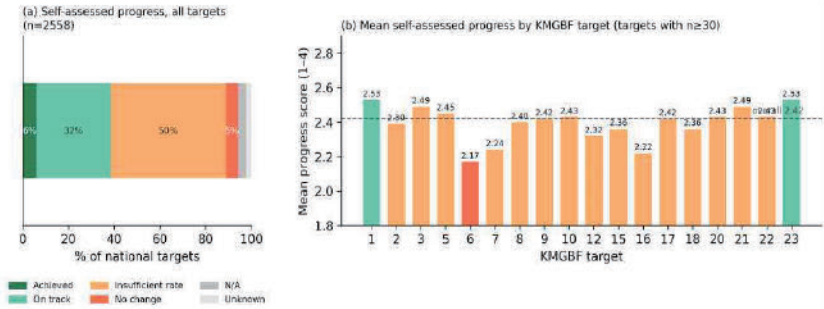
〈그림 2〉 소득군별 NBSAP 개정 상태



2. KMGBF 목표별 자체평가 진척도

진척도 라벨이 부여된 목표는 2,558개이며, 이 가운데 '평가 불가(Not applicable, 90개)'와 '미상(Unknown, 61개)'을 제외한 평가 대상 2,407개 목표를 기준으로 진척도 분포를 분석하였다. 가장 큰 비중은 '진전은 있으나 속도가 불충분함'(53.6%)이었고, '달성 궤도에 있음'은 34.3%, '목표 달성'은 6.5%, '유의한 변화 없음'은 5.7%였다. 결과적으로 '달성 궤도 또는 달성'으로 평가된 목표는 40.8%에 그치고, '불충분 이하'가 59.2%를 차지하였다(평균 진척도 2.42/4, 목표 단위). 목표별로는 외래종(목표 6, 평균 2.17), 오염(목표 7, 2.24), 지속가능한 소비(목표 16, 2.22)에서 진척도가 상대적으로 낮았다〈그림 3〉.

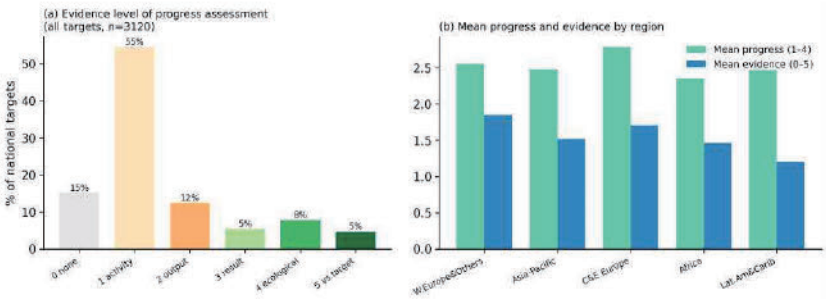
〈그림 3〉 (a)전체 국가목표의 자체평가 진척도 분포, (b)KMGBF 목표별 평균 진척도(n≥30 목표)



3. 진척도 평가의 증거수준

증거수준 분포는 전체 국가목표 3,120건을 기준으로 산정하였다. 정성적 활동 수준(1단계)이 54.6%로 가장 많았고, 근거가 사실상 제시되지 않은 경우(0단계)도 15.1%였다. 정량적 산출(2단계) 12.5%, 정책·사회적 결과(3단계) 5.4%였으며, 종·생태계의 상태변화를 근거로 제시한 경우(4단계)는 7.8%, 기준선·목표값·시계열을 종합한 경우(5단계)는 4.7%에 그쳤다〈그림 4〉.

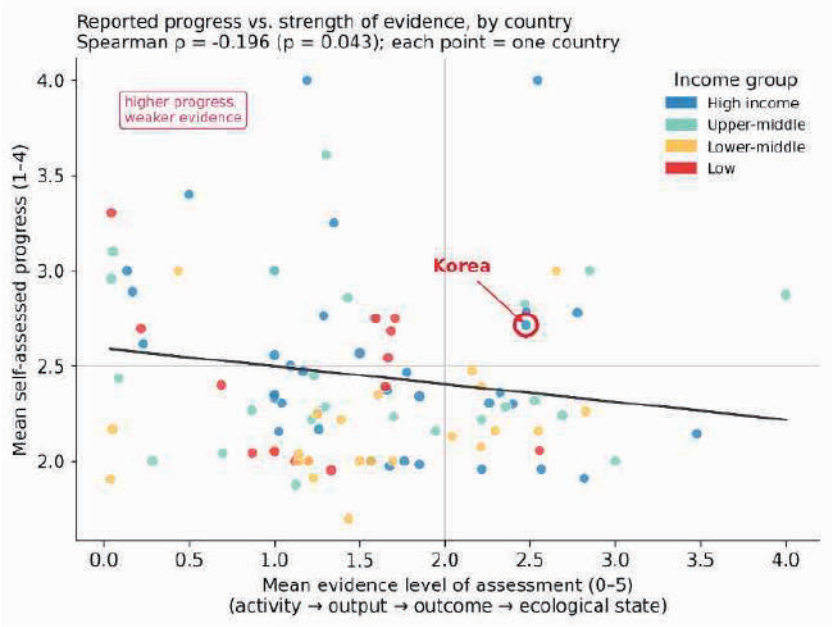
〈그림 4〉 (a)국가목표 진척도 평가의 증거수준 분포, (b)지역별 평균 진척도·증거수준



종합하면 진척도 서술의 69.6%가 활동 수준 이하(0–1단계)에 머물렀고, 정책·사회적 결과 이상(3단계 이상)은 17.9%, 생태적 상태변화 이상(4–5단계)의 증거를 제시한 목표는 12.5%(국가평균 기준 13.7%)에 불과하였다. 이는 국가 이행보고가 정책활동과 행정산출을 기록하는 데에는 진전이 있으

나, 정책이 생물다양성 상태에 미친 효과를 검증하는 단계까지는 발전하지 못했음을 보여준다. 지역별로는 서유럽·기타(WEOG) 지역의 평균 증거수준이 1.85로 가장 높고 중남미(LAC)가 1.21로 가장 낮았다〈그림 4〉.

〈그림 5〉 국가가 보고한 진척도와 증거수준의 관계(점=국가, 색=소득군)



4. 진척도와 증거수준의 정합성

국가가 보고한 진척도와 그 증거수준의 관계는 본 연구의 핵심 발견이다. 이 관계는 특히 국가 단위에서 뚜렷하여, 국가별 평균 진척도와 평균 증거수준은 음(-)의 상관을 보였다(Spearman ρ=-0.196, p=0.043)〈그림 5〉. 국가 단위 4분면 분석에서 '높은 진척도-약한 증거'에 해당하는 경우가 35%로, '높은 진척도-강한 증거'(5.8%)의 여섯 배에 달하였고, '낮은 진척도-강한 증거'도 11.3% 존재하였다. 다만 이 음의 상관은 주로 국가 간(between-country) 차이에서 나타나며, 개별 목표 단위의 상관은 상대적으로 약하다(목표 단위 ρ =-0.151)는 점을 유의할 필요가 있다. 소득수준은 진척도와 약한 양의 상관(ρ

=0.22, p=0.023)을 보였으나 증거수준과는 유의한 상관이 없어(ρ=0.085, p=0.383), 증거수준이 단순한 국가 발전단계의 함수는 아님을 시사한다. 앞서 밝혔듯 증거수준은 실제 생태 성과가 아니라 결과사슬 상 증거의 위치를 포착하는 보고 기반 프록시이므로, 이 음의 관계를 "측정을 많이 할수록 성과가 낮다"는 인과로 해석해서는 안 된다. 오히려 이 결과는 동일한 진척도 등급이라도 그것을 뒷받침하는 증거의 수준이 국가마다 크게 다르며, 따라서 자체 평가 등급을 그대로 국가 간 성과 순위로 환원하면 이행수준이 왜곡될 수 있음을 보여준다.

5. 심화분석: 30×30(KMGBF 목표3)

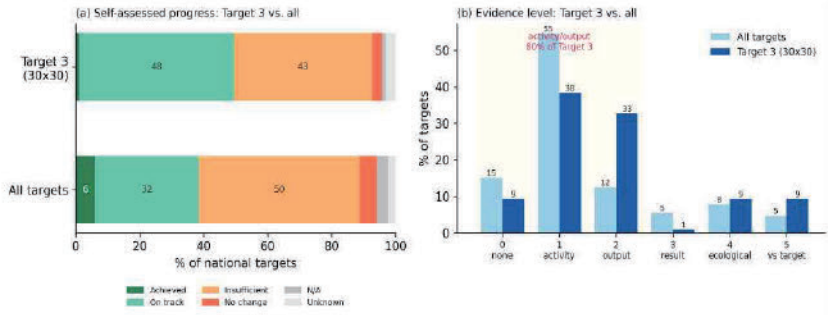
가장 구체적인 행동목표인 목표 3(2030년까지 육상·해양의 각 30%를 보호지역 및 OECM으로 보전, 이른바 "30×30")을 별도로 분석하였다. 134개국 중 78개국이 목표 3에 연계된 국가목표를 설정하였고, 이 가운데 44개국이 30% 수치를 명시하였다. 목표 3의 평균 진척도는 2.49로 전체 평균(2.42)을 다소 상회하여 48.4%가 '달성 궤도', 43.2%가 '속도 불충분'으로 보고되었다.

그러나 보호지역 면적 통계가 국제적으로 가장 잘 축적된 목표임에도, 진척도 근거의 약 80%가 활동·산출 수준(증거 2단계 이하)에 머물렀고 생태적 상태·대표성 증거(4단계 이상)는 18.6%에 그쳤다〈그림 6〉. 즉 다수 국가는 '얼마나 지정했는가(면적)'는 보고하지만 '보호지역이 실제로 생물다양성을 대표하고 효과적으로 보전하는가'는 거의 보고하지 않는다. 한국의 목표 3은 자동 프록시에서 4단계로 분류되었으나, 수기 검토에서 명확히 확인되는 핵심 근거는 보호지역·OECM 지정면적 증가(육상 2023년 17.3%→17.8%, 30% 목표와 상당한 격차)로 결과사슬상 산출에 해당하였다. 이는 개별 목표의 세부 단계 판정에는 오류 가능성이 있음을 보여주며, 본 연구가 증거수준을 개별 사례의 확정적 등급이 아니라 집합적 분포와 국가 단위 방향성으로 제한해 해석한 이유이기도 하다.[7] 다만 본 연구가 활용한 자료는 보고서상 서술

7) Ⅲ장의 타당도 검증과 민감도 분석에 이미 반영됨.

과 지정면적에 한정되므로, 현재의 커버리지 격차와 보고 정보만으로는 '달성 궤도'라는 한국의 자체 판정이 타당한지를 외부에서 검증하기 어렵다. '궤도'라는 자체평가가 큰 커버리지 격차 및 제한된 결과 증거와 공존하는 이 사례는 진척도 판정과 그 증거기반을 분리해 검토할 필요성을 보여준다.

〈그림 6〉 30×30(목표 3) vs 전체: (a)자체평가 진척도, (b)증거수준 분포



6. 진척도·증거수준의 결정요인

진척도와 증거수준이 어떤 구조적 요인과 연관되는지 확인하기 위해 국가 단위최소자승회귀(OLS)와 목표단위 다층모형(국가 임의절편)을 추정하였다 〈표 6〉. 첫째, 국가별 평균 증거수준은 소득·지역·OECD·목표수로 거의 설명되지 않았다(모형 M1, R^2=0.05). 이는 보고의 증거수준이 소득·지역·OECD 여부 등 본 연구가 포함한 구조적 변수만으로는 충분히 설명되지 않으며, 측정·보고 관행을 포함한 미관측 국가 특이적 요인이 작용할 가능성을 시사한다.

〈표 6〉 진척도·증거수준의 결정요인

변수	M1 증거수준(국가)	M2 진척도(국가)	M3 진척도(목표·다층)
소득수준(1-4)	+0.013	+0.102	+0.091
지역: 중·동유럽	−0.147	+0.168	+0.224
지역: 아시아·태평양	−0.311	−0.003	+0.070
지역: 중남미	−0.625†	−0.104	+0.004

지역: 아프리카	−0.322	−0.045	+0.036
OECD	+0.082	+0.011	–
국가목표 수	−0.006	−0.001	–
증거수준	–	−0.149* (SE 0.058)	−0.012
상수	1.903***	2.443***	2.172***

주: 기준지역=서유럽·기타(WEOG). 계수는 비표준화 회귀계수. † p<0.10, * p<0.05, *** p<0.001. 핵심 계수인 M2의 증거수준 계수는 −0.149(SE=0.058, p=0.011)임. 모형별 n의 차이(M1=133, M2=107)는 진척도·NBSAP 상태·소득·지역·OECD 변수의 결측에 따른 완전 사례(complete-case) 분석 결과임. M2의 증거수준 계수는 HC3 강건표준오차, 국가목표 수 가중 WLS, 저(低) 목표국 제외 부표본에서 방향과 유의성이 유지됨.

둘째, 국가별 평균 진척도 모형(M2)에서 평균 증거수준의 계수는 유의하게 음(−0.149, p=0.011)이었다. 구조적 요인을 통제하더라도 증거수준이 높은 국가일수록 진척도를 낮게 보고하는 경향이 남는다는 뜻이며, 이 음의 계수는 이분산에 강건한 표준오차(HC3, p=0.031), 국가목표 수 가중회귀(WLS, b=−0.178, p=0.006), 국가목표 5개 미만 국가를 제외한 부표본(b=−0.145, p=0.014)에서 모두 방향과 유의성이 유지되었다.

셋째, 목표단위 다층모형(M3)의 급내상관(ICC)은 0.45로, 진척도 분산의 약 45%가 국가 수준에 귀속되었다. 다만 국가 임의절편을 통제한 뒤 목표 단위에서 증거수준의 계수는 작고 통계적으로 유의하지 않았는데(선형 다층모형 b=−0.012, p=0.26), 이는 진척도-증거의 음의 관계가 개별 목표 내부보다 국가 간 수준의 차이와 더 밀접하게 관련됨을 시사한다.

종합하면 국가별 특이적 보고·평가 관행이 진척도 자체평가에 크게 작용하며, 자체평가 진척도를 성과지표로 국가 간 직접 비교하는 것의 한계가 재확인된다. 다만 국가별로 설정·보고한 KMGBF 목표의 구성이 상이하므로, 국가 평균의 차이에는 평가문화뿐 아니라 측정 난이도가 다른 목표 포트폴리오의 차이도 일부 반영될 수 있다.

7. 민감도 및 강건성 분석

핵심 발견인 진척도–증거수준의 음의 관계가 특정 척도화·모형 설정에 대해 얼마나 민감한지를 평가하기 위해 여러 민감도 분석을 수행하였다〈표 7〉.

〈표 7〉 진척도-증거수준 관계의 민감도·강건성 분석

설정	지표	추정치	유의성
목표단위 순위상관(연속 0–5)	Spearman ρ	−0.151	$p<0.001$
결과수준 이상(≥3) 목표 평균진척도	고증거 대 저증거	2.29 대 2.44	$p<0.001$
생태증거 이상(≥4) 목표 평균진척도	고증거 대 저증거	2.34 대 2.43	$p=0.0936$ (n.s.)
국가단위 순위상관(평균)	Spearman ρ	−0.196	$p=0.0432$
국가단위: 진척도-결과수준 증거비율	Spearman ρ	−0.191	$p=0.0489$
국가회귀 M2 증거계수(OLS)	b	−0.149	$p=0.011$
국가회귀 HC3 강건표준오차	b	−0.149	$p=0.031$
국가회귀 목표수 가중(WLS)	b	−0.178	$p=0.0062$
국가회귀 저목표국 제외(≥5)	b	−0.145	$p=0.0143$
목표단위 순서형 로짓 증거계수	b	−0.186	$p<0.001$
5개 LLM 코더 합의 증거 부표본(n=292)	Spearman ρ	0.105	$p=0.0727$ (n.s.)

주: "결과수준 이상(≥3)"은 코드북(표3)의 3단계(정책·사회적 결과) 이상, "생태증거 이상(≥4)"은 4–5단계(생태적 상태변화·목표대비 성과)를 뜻함. 국가 단위 관계는 여러 설정에서 방향이 유지된 반면, 목표 단위 관계는 생태증거(≥4) 이분화와 5 모델 합의 부표본에서 유의하지 않았음. 따라서 본 연구의 핵심 해석은 국가 단위 집계관계에 한정함. 순서형 로짓은 국가 임의효과를 포함하지 않는 통합모형임. n.s.=통계적으로 유의하지 않음.

여기서 증거수준의 이분화 문턱은 코드북〈표 3〉에 따라 두 가지로 구분한다. 즉 3단계는 정책·사회적 결과, 4–5단계는 생태적 상태변화·목표대비 성과에 해당하므로, '결과수준 이상 증거(3단계 이상)'와 '생태적 상태변화 이상 증거(4단계 이상)'를 각각 적용하였다. 첫째, 증거수준을 연속척도가 아니라 이분변수로 바꾸어도 관계의 방향은 유지되었다. 결과수준 이상 증거(3단계 이상)를 제시한 목표의 평균 진척도(2.29)는 그렇지 않은 목표(2.44)보다 낮았고(Mann–Whitney $p<0.001$), 문턱을 생태적 상태변화 이상(4단계 이상)으로 높였을 때에도 방향은 동일하였으나 그 차이는 통계적으로 유의하지 않았다(2.34 대 2.43, $p=0.0936$). 국가 단위에서 평균 진척도와 '결과수준 이상 증거 목표 비율'의 순위상관도 음($\rho=-0.191$, $p=0.0489$)이었다. 둘째, 앞서의 국가단위회귀(M2)에서 증거수준 계수의 음의 방향은 HC3·WLS·부표본 설정 전반에서 견고하였다. 셋째, 진척도의 순서형 성격을 반영한 통합(pooled) 순서형 로짓(cumulative-link) 설정에서도 증거수준의 계수는 음이었다($b=-0.186$, $p<0.001$). 다만 이 순서형 로짓은 국가 임의효과를 포함하지

않아 국가 간·국가 내 변동이 혼합된 결과이므로, 국가 내 연관을 추정한 선형 다층모형(M3, b=-0.012, non-significance)과 직접 비교하기보다 척도 민감도의 보조 근거로 해석한다.

한편 관계의 강도는 분석 수준에 따라 달랐다. 규칙 기반 프록시가 아니라 서로 다른 계열의 5개 대규모 언어모델이 합의(다수결)한 증거수준을 사용해 무작위 블라인드 부표본(n=292)에서 목표 단위 상관을 재계산하면, 진척도–증거의 상관은 사실상 '0'에 가깝거나 미약한 양(ρ=0.105, p=0.0727)으로 나타났다. 이는 전체 자동 프록시에서 관찰된 목표 단위 상관의 강건성이 제한적임을 보여주며, 따라서 본 연구의 핵심 해석은 목표 단위 관계가 아니라 전체 자료의 국가 단위 집계 관계에 두어야 함을 시사한다. 요컨대 진척도–증거의 음의 관계는 국가 단위에서 여러 설정에 걸쳐 일관되게 지지되지만 개별 목표 내부에서는 강건하지 않으며, 그 효과 크기 또한 작으므로 과대해석하지 않도록 주의할 필요가 있다.

요약하면, 진척도–증거수준의 음의 관계는 이분화·강건 표준오차·가중회귀·순서형 설정 등 여러 통계적 명세에서 국가 단위로는 일관되게 유지되나, 개별 목표 단위에서는 약하고 유의하지 않다. 따라서 본 연구의 핵심 주장은 "자체평가 등급을 국가 간에 그대로 비교하는 것이 위험하다"는 국가 단위 명제에 한정되며, 개별 목표 수준의 인과적 일반화로 확장되지 않는다.

V. 한국의 국가 이행평가체계 진단

1. 한국 제7차 국가보고서의 구조

한국의 제도적 강점과 개선과제는 앞의 134개국 정량분석에 더하여, 한국 제7차 국가보고서와 『제5차 국가생물다양성전략(2024-2028)』의 이행·평가 구조에 대한 문서검토를 병행하여 도출하였다. 한국의 제7차 국가보고서는 『제5차 국가생물다양성전략』의 국가목표별 이행실적과 관계부처자료를 기반으로 작성되었으며, 전문가·시민사회 참여와 국가생물다양성위원회 검

토를 거쳐 확정되었다. 국가목표는 KMGBF 23개 목표와 1:1에 가깝게 매핑되어 있고,[8] 부처별 시행계획과의 연결이 비교적 분명하다.

2. 한국의 상대적 위치

분석 결과 한국은 국가목표 21개에0대해 평균 진척도 2.71로 전체 평균(2.48) 및 고소득국 평균(2.60)을 상회하였다. 증거수준에서도 한국의 평균은 2.48로 전체 평균(1.52)·고소득국 평균(1.60)·OECD 평균(1.73)을 모두 웃돌았으며, 증거수준이 산출된 고소득국 47개국 가운데 상위 약 21% 수준(47개국 중 10위)으로 비교적 상세한 보고서를 제출한 것으로 나타났다. 이는 한국을 단순한 '행정산출 중심의 후발 보고국'으로만 보기 어렵다는 점을 시사한다. 그러나 동시에, 한국 국가목표 중 생태적 상태변화(4-5단계)에 근거한 목표는 19%에 그쳐 나머지 약 81%는 생태적 상태변화 이상의 증거에 이르지 못하였다. 즉 한국의 상대적 강점은 '보고의 상세성'이지 '생태적 결과에 기반한 판정'은 아니며, 결과 중심 평가로의 전환 과제가 분명하다〈표 8, 그림 7〉.

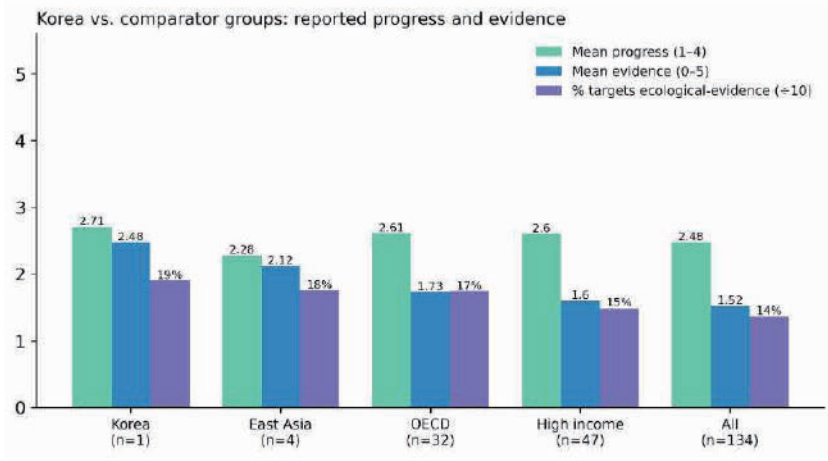
〈표 8〉 한국과 주요 비교집단의 이행평가 지표

비교집단	국가수	평균진척도 (0-4)	평균 증거수준(0-5)	생태적 증거 목표비율(%)	평균 국가목표수
한국	1	2.71	2.48	19	21
동아시아(4개국)	4	2.28	2.12	17.6	27
OECD	32	2.61	1.73	17.4	28.9
고소득국	47	2.60	1.60	14.8	26.3
전체	134	2.48	1.52	13.7	23.5

주: 각 집단의 값은 국가별 평균을 다시 평균한 국가평균이며(한국은 단일국이므로 목표단위 평균과 동일), 〈표 5〉의 "국가평균" 기준과 일치함. 동아시아는 일본·한국·중국·몽골. 증거수준은 독립 코더 검증(이차가중 κ 0.36)을 거친 자동 코딩 프록시이며, 4-5단계 세부 구분보다 "생태증거 여부"의 대비가 더 신뢰할 수 있음.

8) 본 분석에서 21개 국가목표를 확인함.

〈그림 7〉 한국과 비교집단의 진척도·증거수준



〈표 8〉의 지표는 각각 자체평가 등급(평균 진척도 0–4)과 그 근거의 강도(평균 증거수준 0–5)를 나타내며, 한국은 두 지표 모두에서 고소득국·OECD 평균을 상회한다. 다만 이 강점이 '생태적 결과에 기반한 판정'으로 이어지지 못한 지점은 개별 목표에서 뚜렷이 드러난다. 예컨대 30×30에 해당하는 목표 3의 경우 한국은 '달성 궤도'로 판정하였고 자동 프록시에서 4단계로 분류되었으나, 보고서에서 확인되는 핵심 정량 근거는 보호지역 지정면적의 증가(본래 산출 수준)였고, 지정된 보호지역이 생태적 대표성·연결성·관리효과성 측면에서 실제로 생물다양성을 얼마나 보전하는지에 대한 상태변화 증거는 제한적이었다. 이는 '면적 지정'이라는 산출 중심 판정과 '생태적 상태변화' 기반 판정 사이의 간극을 단적으로 보여주는 동시에, 자동 프록시의 단계 판정이 실질 근거보다 높게 나타날 수 있음을 예시한다.

3. 한국 체계의 강점과 개선과제

한국 체계의 강점은 (1)국가전략과 관계부처 시행계획을 연결하는 범정부적 관리, (2)전문가·시민사회가 참여하는 평가절차와 국가위원회 검토, (3)시민과학·국가 생물다양성 정보체계의 확대, (4)국제적으로 비교적 상세하고

증거가 풍부한 보고서 작성 역량에 있다. 그러나 개선과제도 분명하다. 첫째, 증거 서술은 풍부하나 최종 진척도 판정이 여전히 활동·산출지표에 의존하는 목표가 존재한다. 둘째, 국가목표별 기준선·정량 목표값의 일관성이 부족하다. 셋째, 부처별 조사방법·공간단위가 상이하다. 넷째, 평가의 불확실성·신뢰도가 명시되지 않는다. 다섯째, 평가 결과가 차년도 예산·사업 조정에 어떻게 반영되는지 확인하기 어렵다.

Ⅵ. 논의

1. 보고된 이행과 실제 성과의 간극

본 연구가 드러낸 진척도–증거 간극은, 국가가 보고한 진척도와 그 진척도를 뒷받침하는 증거수준 사이의 괴리를 실증적으로 보여준다. 이는 국제환경협약의 자체보고가 지니는 근본적 긴장과 맞닿아 있다. 자체보고에 기반한 이행점검은 당사국의 주권과 행정부담을 존중한다는 장점이 있으나, 보고 주체가 곧 평가 주체이기도 하다는 점에서 이론적으로 성과의 과대표시(over-claiming) 유인을 내포한다. 다만 본 연구는 실제 생태 성과를 외부 자료로 측정하지 않았으므로 과대표시 여부를 직접 검증한 것이 아니며, 오히려 그러한 가능성을 외부에서 식별하기 어렵게 만드는 보고구조의 이질성을 계량화한 것이다. 이 이질성은 추상적 우려가 아니라 제7차 국가보고서 전반에 걸쳐 관측되는 체계적 양상이다. 즉 진척도 등급 자체는 비교적 풍부하게 보고되지만,[9] 그 등급을 뒷받침하는 증거는 대부분 실질적 근거가 없거나 정성적 정책활동 수준에 머물러[10] 생태적 상태변화 이상으로 검증된 경우는 목표 단위 12.5%에 그쳤다.

9) 평가 대상의 53.6%가 "속도 불충분", 40.8%가 "궤도 또는 달성"임.
10) 전체 목표의 69.6%가 증거 0-1단계임.

2. 진척도-증거 음의 관계에 대한 세 가지 해석

이러한 간극은 무엇보다 KMGBF 이행보고에 보정된 증거·불확실성 언어가 제도화되어 있지 않다는 점과 관련된다. IPCC는 증거의 양·질과 일치도에 기반한 신뢰도 언어를 표준화하고 이를 지속적으로 정교화해 왔으며(Mach, Freeman, Mastrandrea and Field, 2016, 2017, pp.1-14; Mastrandrea et al., 2010; Slade, Pathak, Connors, Tignor, Okem and Leprince-Ringuet, 2024), 이 프레임워크의 강점과 한계는 방대한 후속 연구의 대상이 되어 왔다(Harris, 2025; Helgeson, Bradley and Hill, 2018, pp.517-525; Janzwood, 2020, pp.1655-1675; Kause, de Bruin, Persson, Thorén, Olsson, Wallin et al., 2022, pp.1-18). 반면 CBD의 진척도 자체평가는 국가가 등급을 선택하도록 하되, 그 등급이 어떤 수준의 증거에 근거하는지를 표준화된 형식으로 함께 제시하도록 요구하지 않는다. 그 결과 동일한 '달성 궤도' 등급이 한 국가에서는 종·생태계의 상태 개선을, 다른 국가에서는 사업 추진 사실만을 의미할 수 있다. 본 연구의 증거수준 분석은 바로 이 '보정되지 않은 등급'이 실제 보고관행에서 산출하는 이질성을 계량화한 것이다.

진척도와 증거수준이 국가 단위에서 음의 상관을 보이고〈그림 5〉 진척도 분산의 약 45%가 국가 간 차이에서 비롯된다는 다층모형 결과(ICC=0.45)〈표 6〉는 특히 신중한 해석을 요한다. 앞서 강조했듯 증거수준은 실제 생태 성과가 아니라 모니터링 역량, 보고의 완결성, 목표의 성격, 보고 전략이 함께 반영된 보고 기반 지표이다. 따라서 이 음의 관계를 '측정을 많이 한 국가일수록 성과가 나쁘다'는 인과로 읽어서는 안 되며, 최소한 세 가지 비배타적 설명이 가능하다. 첫째, 모니터링 역량이 높은 국가일수록 생물다양성의 감소추세와 불확실성을 더 정밀하게 포착하여 보수적으로 자체평가하는 경향이 있을 수 있다. 이는 증거의 정밀한 검토가 오히려 한계와 불확실성을 부각한다는 과학적 합의·신뢰성 문헌의 통찰과 맥을 같이한다(Oppenheimer, O'Neill, Webster and Agrawala, 2007, pp.1505-1506; Oreskes, 2004). 둘째, 국가별 평가문화(optimism bias)의 차이로, 측정 근거가 약한 국가일수록 낙관적 등급을 부여하기 쉽다. 셋째, 목표 야심수준·목표 성격의 차이로, 도전적이거

나 결과 측정이 까다로운 목표를 세운 국가가 낮은 진척도를 보고할 수 있다. 이 음의 관계는 이분화·가중회귀·순서형 로짓 등 여러 설정에서 방향이 유지되었으나, 주로 국가 간 수준에서 성립하고 개별 목표 내부에서는 약하다는 점에서 그 크기를 과대해석하지 않아야 한다. 소득·지역이 진척도·증거 수준을 유의하게 설명하지 못한 회귀 결과〈표 6〉는 이 현상이 단순한 발전단계의 함수가 아니라 측정·평가 역량과 보고문화라는 제도적 요인에 더 크게 좌우됨을 시사한다. 어느 설명을 취하든 정책적 함의는 일관된다. 즉 자체평가 등급을 그대로 국가 간 성과 순위로 환원하면 국가별 측정·보고 관행의 차이를 성과 차이로 오인할 가능성이 있다.

3. 목표군별 증거기반의 편중

이 문제는 목표의 성격에 따라 다르게 나타난다. 보호지역 통계가 국제적으로 가장 잘 축적된 30×30(목표 3)조차 진척도 근거의 약 80%가 지정 면적 등 활동·산출에 머물렀고 대표성·연결성·관리효과성 같은 생태적 질(質)에 관한 증거는 드물었다. 이는 '면적 목표의 달성'이 곧 '생물다양성 성과'를 담보하지 않는다는 보전생물학의 오랜 경고를 국가보고 차원에서 재확인한다. 반대로 기업의 생물다양성 영향, 민간재원, 지속가능한 소비, 유해보조금(목표 14·15·16·18·19)에서 증거수준이 낮은 것은, 자연관련 재무정보공개(TNFD, 2023)나 생태계서비스 가치평가(Costanza et al., 1997, pp.253-260; TEEB, 2010)와 같은 측정·공개 인프라가 아직 국가 이행보고에 충분히 통합되지 못했음을 반영한다. 따라서 증거기반의 취약성은 균질한 것이 아니라 목표군에 따라 구조적으로 편중되어 있으며, 이는 KMGBF 모니터링 프레임워크(CBD, 2022b, Decision 15/5)의 지표 개발이 우선순위를 두어야 할 영역을 지목한다.

4. 글로벌 검토에 대한 함의

이상의 결과는 2026년과 2029년의 글로벌 검토에 직접적 함의를 갖는다.

사무국과 당사국총회가 국가보고서를 종합할 때 진척도 등급의 단순 집계에 의존하면, 증거가 약하지만 낙관적으로 보고한 목표와 증거가 강하지만 보수적으로 보고한 목표가 동일하게 취급되어 전 지구적 이행수준이 실제보다 낙관적으로 요약될 위험이 있다. 본 연구는 글로벌 검토가 진척도 등급과 함께 이를 뒷받침하는 증거수준·지표 완결성·불확실성을 가중하여 종합하는 '증거가중(evidence-weighted) 접근'으로 이행할 필요가 있음을 시사한다.

5. 해석상의 유보와 한계

끝으로 본 연구의 해석에는 몇 가지 유보가 필요하다. 분석은 국가 자체보고 자료에 의존하므로 실제 생태 성과와의 괴리를 직접 검증하지는 못하며, 제출·수집된 국가에 한정되어 미보고국(보고 역량이 상대적으로 낮을 가능성이 있는 국가)을 포괄하지 못하는 선택편의를 지닌다. 또한 단일 보고주기의 단면 분석이므로 시계열적 변화는 후속 연구의 과제이다. 증거수준은 자동 코딩 프록시로 1차 부여한 뒤 다중 코더로 검증하였으나(Ⅲ장), 개별 목표의 세밀한 수준 판정에서는 여전히 연구자 코더의 확정이 필요하다. 특히 프록시와 독립 코더의 일치가 이차 가중 κ 0.36('보통') 수준에 머무른 만큼, 본 연구의 결론은 개별 목표의 정확한 등급이 아니라 "활동·산출 대 생태증거"라는 집합적 방향에 한정해 해석해야 한다. 다만 이러한 유보에도 불구하고, 자동 프록시가 생태적 증거를 오히려 과대추정하는 경향을 고려하면 '활동·산출 중심'이라는 본 연구의 핵심 결론은 보수적으로 성립한다.

Ⅶ. 한국 생물다양성 이행평가체계의 개선방안

본 연구의 실증분석이 한국에 던지는 핵심 메시지는 하나로 요약된다. 한국의 상대적 강점은 '보고의 상세성'이지 '생태적 결과에 기반한 판정'이 아니라는 것이다. 한국은 국제적으로 비교적 상세하고 증거가 풍부한 보고서를 제출하였으나, 국가목표의 약 81%가 여전히 생태적 상태변화에 근거하지 않았

다. 따라서 한국의 과제는 이행 노력의 부족이 아니라 '행정적 실적관리에서 생태적 결과 중심 평가로의 전환'에 있다. 아래에서는 이 진단에 직접 대응하는 네 가지 제언을 제시하되, 각 제언을 (1) 본 연구의 실증발견, (2) 그로부터 진단되는 제도적 문제, (3) 이에 대응하는 제도개혁의 순서로 구성하였다.

1. 결과·영향지표에 근거한 검증 가능한 목표체계로의 재설계

본 연구에서 진척도 근거의 다수가 활동·산출 수준에 머물렀고, 한국 역시 국가목표의 약 81%가 생태적 상태변화에 근거하지 않았다. 이는 한국의 국가목표에 기준연도·기준값·2030년 목표값이 일관되게 제시되지 않고, 최종 진척도 판정이 지정면적·예산·사업 수 등 산출지표에 의존하는 데에서 비롯된다. 따라서 각 국가목표에 기준값·정량 목표값·자료생산 책임기관을 명시하도록 의무화하고, 활동·산출지표는 과정 점검에만 활용하되 최종 진척도 판정은 원칙적으로 종 개체군·생태계 면적과 건전성·서식지 질 등 결과·영향지표에 근거하도록 평가체계를 재설계해야 한다. 이는 제8차 국가보고서의 진척도 판정을 외부에서 검증 가능하게 만드는 전제 조건이다.

2. 분산된 모니터링의 통합 조정과 데이터 표준의 일원화

본 연구는 증거수준이 소득·지역보다 국가별 특이적 측정·보고 역량에 좌우됨을 보였다. 한국에서는 부처별로 분산된 조사체계가 측정단위·공간단위·주기의 이질성을 낳아 국가 차원의 통합·비교를 제약해 왔다. 산림·해양·농업·국가유산 등 실제 조사 및 모니터링은 각 부처·기관이 수행할 수밖에 없으므로, 여기서 제안하는 '일원화'는 모든 모니터링 업무를 한 부처로 이관한다는 뜻이 아니라 데이터 표준과 평가 조정기능을 일원화한다는 의미이다. 2025년 10월 환경·기후·에너지 기능을 통합한 기후에너지환경부의 출범은 이러한 조정을 강화할 제도적 계기가 될 수 있다. 통합 행정체제를 활용하여 관계부처 자료의 메타데이터·공간단위·조사주기를 표준화하고, 국가 차원의 생물다양성 정보·평가 조정기능(예: 국가 생물다양성 정보체계)

을 통해 자료의 수집·품질관리·연계를 조정하며, 재생에너지 인프라 확충과 서식지 보전 사이의 상충과 같이 기후·에너지 정책과 생물다양성 이행이 교차하는 지점을 통합적으로 평가·조정하는 체계를 마련할 필요가 있다. 장기 생태 시계열의 확보와 시민과학 자료의 품질관리도 함께 요구된다.

3. 증거수준·신뢰도의 병기와 시장기반 성과의 검증체계 구축

본 연구에서 동일한 진척도 등급이 국가 및 목표마다 상이한 수준의 증거에 근거함이 드러났다. 한국의 보고서 역시 서술은 풍부하나 진척도 판정의 근거와 불확실성이 명시되지 않는 경우가 많다. 이에 각 국가목표에 증거수준과 신뢰도 등급을 병기하고 지표별 자료 출처와 산정방법을 공개하며, 외부 전문가 및 시민사회의 독립적 검토를 제도화할 필요가 있다. 나아가 생물다양성 순증(biodiversity net gain)이나 자연크레딧(nature credits)과 같이 향후 국내에도 도입·확산될 수 있는 시장기반 보전수단은 그 성과가 결국 '생물다양성 상태의 실질적 개선'으로 환산되어야 하므로 본 연구의 증거수준 논의와 직접 맞닿는다. 이러한 수단은 성과의 측정·보고·검증(MRV)이 부실할 경우 오히려 과대표시를 증폭시킬 수 있으므로, 이들 수단의 성과를 결과·영향지표와 독립 검증에 연동하는 국가 차원의 검증체계를 도입 이전 단계에서 미리 설계해 둘 필요가 있다. 이는 본 연구가 제안한 증거수준 개념을 국내 보고·시장 제도에 내재화하는 방안이기도 하다.

4. 평가–예산 환류와 국가–지방 이행의 연계

본 연구의 결과는 자체평가가 후속 정책과 연결되는 환류체계의 필요성을 뒷받침한다. 한국에서는 평가결과가 차년도 예산·사업 조정에 어떻게 반영되는지 확인하기 어렵고, 확산되는 지방생물다양성전략의 자료도 국가보고로 체계적으로 환류되지 못하고 있다. 따라서 목표 미달 부처의 보완계획 제출을 의무화하고 중간평가 결과를 국가목표 수정과 예산 배분에 연동하며, 국가목표와 지역목표를 매핑하는 공통지표와 지방정부 자료의 표준 전송체

계를 구축하여 지역 이행성과가 국가 단위로 집계·환류되도록 해야 한다. 제8차 국가보고서를 위한 상시 데이터베이스의 구축이 이 순환을 뒷받침한다.

〈표 9〉 한국 생물다양성 이행평가체계 개선과제

제언	실증적 진단	제도개혁 방향	기대효과
① 목표·지표 재설계	활동·산출 중심 판정, 기준값 부재	기준값·2030목표값·책임기관 의무화, 결과·영향지표 기반 최종판정	진척도 판정의 검증 가능성
② 모니터링 조정·데이터 표준화	부처별 조사체계 분산	기후에너지환경부 통합체제 활용, 데이터 표준·평가 조정기능 일원화	국가자료의 비교가능성
③ 증거·신뢰도 병기	불확실성·근거 미제시	증거수준·신뢰도 등급 병기, 시장수단 MRV·독립검증 연동	평가의 투명성·신뢰성
④ 평가-환류·국가 지방연계	평가-예산 연결 불명확, 지방자료 미연계	미달과제 보완계획·예산연동, 지방자료 표준 전송체계	성과관리의 실효성

이상의 네 제언은 〈표 9〉에 요약되며 상호 보완적이다. 특히 (1)의 목표·지표 재설계가 선행되어야 (3)의 증거·신뢰도 병기와 (4)의 환류가 실질적으로 작동할 수 있다.

Ⅷ. 결론

본 연구는 KMGBF 채택 이후 제출된 CBD 제7차 국가보고서를 추출·코딩하여 국가 이행평가의 구조를 실증적으로 규명하고, 그 비교 속에서 한국의 위치와 과제를 진단하였다. 분석 결과 응답국의 98.1%가 국가생물다양성전략을 개정하거나 개정 중이어서 전략의 형식적 정렬은 광범위하게 진전되었으나, 국가목표의 59.2%가 '속도 불충분' 이하로 평가되었고 진척도 근거의 69.6%가 정책활동 수준에 머물렀다. 무엇보다 국가가 보고한 진척도와 이를 뒷받침하는 증거의 강도는 국가 단위에서 음의 상관을 보였으며, '높은 진척도-약한 증거' 목표(35%)가 '높은 진척도-강한 증거'(5.8%)의 여섯 배에 달하였다.

본 연구의 기여는 세 층위에서 정리된다. 개념적으로는 국가가 보고한 '진

척도'와 그것을 뒷받침하는 '증거수준'을 분리하여, 이행평가의 비교가능성 문제를 성과의 격차가 아니라 측정·평가 역량의 격차로 재정의하였다. 방법론적으로는 다수 국가보고서를 공통 양식에 따라 재현 가능하게 구조화하는 절차와, 증거수준 코딩을 서로 다른 계열의 대규모 언어모델로 교차 검증하는 신뢰도 설계를 제시하였다. 경험적으로는 진척도–증거 간극과 국가 간 분산(ICC=0.45)을 계량화하고, 30×30을 포함한 목표별 증거편중을 확인하였다. 이는 국제환경협약의 자체보고가 지닌 타당성 문제를 생물다양성 분야에서 체계적으로 문서화한 시도의 하나이다.

정책적으로 이 결과는 국가 간 이행성과를 자체평가 등급으로 단순 서열화하는 것의 위험을 경고한다. 모니터링 역량이 높은 국가일수록 감소추세와 불확실성을 더 정밀하게 포착하여 보수적으로 평가할 수 있으므로, 낮은 진척도가 곧 낮은 정책역량을 의미하지는 않는다. 따라서 2026년과 2029년의 글로벌 검토는 진척도 등급과 함께 그 증거수준·지표 완결성·불확실성을 가중하여 종합할 필요가 있다. 한국의 경우, 범정부적 이행관리와 전문가·시민사회 참여, 국가위원회 검토라는 제도적 기반은 비교적 견고하나, 풍부한 서술을 국가목표별 기준값·목표값·장기 생태지표·불확실성에 근거한 결과중심 평가로 전환하고 그 결과를 예산·정책으로 환류하는 것이 제8차 국가보고서를 향한 핵심 과제이다.

본 연구는 국가 자체보고 자료라는 근원적 제약, 미보고국을 포괄하지 못하는 선택편의, 단일 보고주기의 단면성, 그리고 자동 코딩의 한계를 지닌다. 후속 연구에서는 독립적 사람 코더에 의한 전수 재코딩으로 신뢰도를 확정하고, 국가보고 자료를 원격탐사·현장 모니터링 등 독립적 생태 자료와 대조하여 자체평가의 타당성을 직접 검증하며, 제8차 보고주기와의 비교를 통해 이행평가의 시계열적 변화를 추적할 필요가 있다. 궁극적으로 국가의 정책행동을 비교 가능한 생태적 성과의 증거로 전환하는 측정·평가 역량의 강화야말로, 2030년 목표의 달성 가능성을 실질적으로 높이는 동시에 한국 생물다양성 정책이 국제적 신뢰성을 확보하는 길이 될 것이다.

▮참고문헌▮

고하정·김도은·권혁수·김정인, 2026, "뉴스 미디어를 통한 환경담론의 의제화 과정: 토픽 간 관계와 담론 전환점 탐지," 『환경정책』, 34(1), pp.263-287.

관계부처 합동, 2023, 『제5차 국가생물다양성전략(2024-2028)』, 세종.

김선희·윤순진, 2026, "IPCC 평가보고서에 나타난 과학적 확실성의 강화: 기후변화의 실재성과 인간의 책임성에 관한 담론 변화," 『환경정책』, 34(1), pp.1-38.

대한민국 정부, 2026, 『생물다양성협약 제7차 국가보고서』, 세종.

윤순진·김선희, 2024, "IPCC 종합 보고서의 정책결정자를 위한 요약 원본과 한국어 번역본의 과학적 불확실성 및 저자 입장 표현 비교 분석," 『한국기후변화학회지』, 15(5), pp.799-814.

홍민아·송민경, 2026, "글로벌 적응목표 달성을 위한 주요 국가의 산림부문 적응 정책 분석 연구," 『환경정책』, 34(1), pp.67-82.

Blei, D. M., A. Y. Ng, and M. I. Jordan, 2003, "Latent Dirichlet allocation," *Journal of Machine Learning Research*, 3, pp.993-1022.

Butchart, S. H. M., M. Walpole, B. Collen, A. van Strien, J. P. W. Scharlemann, R. E. A. Almond et al., 2010, "Global biodiversity: Indicators of recent declines," *Science*, 328(5982), pp.1164-1168.

Convention on Biological Diversity (CBD), 2022a, *Decision 15/4: Kunming-Montreal Global Biodiversity Framework*, Montreal: SCBD.

_____, 2022b, *Decision 15/5: Monitoring framework for the Kunming-Montreal Global Biodiversity Framework*, Montreal: SCBD.

_____, 2022c, *Decision 15/6: Mechanisms for planning, monitoring, reporting and review*, Montreal: SCBD.

Costanza, R., R. d'Arge, R. de Groot, S. Farber, M. Grasso, B. Hannon et al., 1997, "The value of the world's ecosystem services and natural capital," *Nature*, 387(6630), pp.253-260.

Denny, M. J. and A. Spirling, 2018, "Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it," *Political Analysis*, 26(2), pp.168-189.

Harris, M., 2025, "The IPCC uncertainty framework: What some decision makers want (and why they shouldn't)," *Climatic Change*, 178(5), 105.

Hase, V., D. Mahl, M. S. Schäfer, and T. R. Keller, 2021, "Climate change in news media across the globe: An automated analysis of issue attention and themes in climate change coverage in 10 countries (2006-2018)," *Global Environmental Change*, 70, 102353.

Helgeson, C., R. Bradley, and B. Hill, 2018, "Combining probability with qualitative degree-of-certainty metrics in assessment," *Climatic Change*, 149(3), pp.517–525.

IPBES, 2019, *Global assessment report on biodiversity and ecosystem services*, Bonn: IPBES Secretariat.

IPCC, 2023, *AR6 synthesis report: climate change 2023*, Geneva: IPCC.

Janzwood, S., 2020, "Confident, likely, or both? The implementation of the uncertainty language framework in IPCC special reports," *Climatic Change*, 162(3), pp.1655–1675.

Kause, A., W. Bruine de Bruin, J. Persson, H. Thorén, L. Olsson, A. Wallin et al., 2022, "Confidence levels and likelihood terms in IPCC reports: A survey of experts from different scientific disciplines," *Climatic Change*, 173(2), pp.1–18.

Mach, K. J., P. T. Freeman, M. D. Mastrandrea, and C. B. Field, 2016, "A multistage crucible of revision and approval shapes IPCC policymaker summaries," *Science Advances*, 2(8), e1600421.

Mach, K. J., M. D. Mastrandrea, P. T. Freeman, and C. B. Field, 2017, "Unleashing expert judgment in assessment," *Global Environmental Change*, 44, pp.1–14.

Mastrandrea, M. D., C. B. Field, T. F. Stocker, O. Edenhofer, K. L. Ebi, D. J. Frame et al., 2010, *Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties*, Geneva: IPCC.

Oppenheimer, M., B. C. O'Neill, M. Webster, and S. Agrawala, 2007, "The limits of consensus," *Science*, 317(5844), pp.1505–1506.

Oreskes, N., 2004, "The scientific consensus on climate change," *Science*, 306(5702), p.1686.

Roberts, M. E., B. M. Stewart, D. Tingley, C. Lucas, J. Leder-Luis, S. K. Gadarian et al., 2014, "Structural topic models for open-ended survey responses," *American Journal of Political Science*, 58(4), pp.1064–1082.

Schmidt, A., A. Ivanova, and M. S. Schäfer, 2013, "Media attention for climate change around the world: A comparative analysis of newspaper coverage in 27 countries," *Global Environmental Change*, 23(5), pp.1233–1248.

Schneider, S. H. and R. H. Moss, 1999, "Uncertainties in the IPCC TAR: Recommendations to lead authors for more consistent assessment and reporting," Geneva: IPCC.

Secretariat of the CBD, 2020, *Global Biodiversity Outlook 5*, Montreal: SCBD.

_____, 2024, *Guidance and Template for the Seventh and Eighth National Reports*, Montreal: SCBD.

Slade, R., M. Pathak, S. Connors, M. Tignor, A. E. Okem, and N. Leprince-Ringuet, 2024, "Back to basics for the IPCC: Applying lessons from AR6 to the Seventh Assessment Cycle," *npj Climate Action*, 3(1), 48.

Taskforce on Nature-related Financial Disclosures (TNFD), 2023, *Recommendations of the Taskforce on Nature-related Financial Disclosures*.

The Economics of Ecosystems and Biodiversity (TEEB), 2010, *Mainstreaming the Economics of Nature: A Synthesis of the Approach, Conclusions and Recommendations of TEEB*, Nairobi: UNEP.

Victor, D. G., R. Gerlagh, and G. Baiocchi, 2014, "Getting serious about categorizing countries," *Science*, 345(6192), pp.34-36.

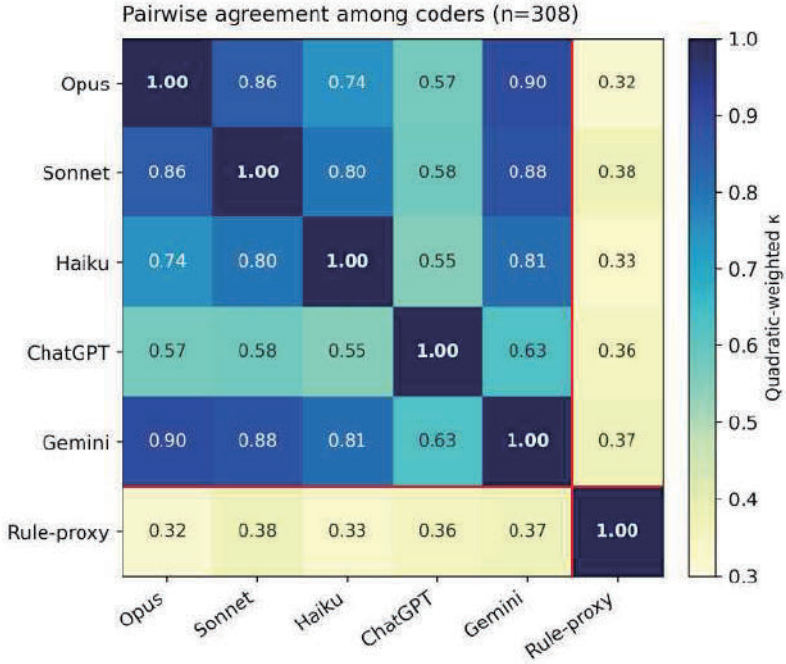
웹사이트

제7차 국가보고 포털, 2006, https://ort. cbd.int/national-reports/nr7, [2026.4.10.]

▌부록 ▌ 데이터 구축 및 한계

본 연구의 데이터셋은 134개국 국가보고서에서 자동 추출한 국가목표 3,120건으로 구성되며, 국가 단위(NBSAP 상태, 국가목표 수, 평균 진척도·증거수준)와 목표 단위(진척도, GBF 목표 연계, 증거수준) 두 파일로 정리하였다. 진척도 범주 커버리지는 110개국 2,558개 목표이다. 증거수준 프록시는 45개 표본의 저자 수기 코딩(이차 가중 κ 0.50), 이어 312개(전체의 10%) 표본에 대한 코드북 기반 독립 재코딩(이차 가중 κ 0.36, 95% CI 0.24-0.48; ±1단계 이내 일치율 82.2-84.0%)으로 검증하였다(Ⅲ장). 코더 간 일치는 코더 이질성이 커질수록 낮아졌다: 동일 모델 3인스턴스 Fleiss κ=0.94(편향으로 과대) → 이종 3모델(Opus·Sonnet·Haiku) 0.77 → 타 벤더 2모델(GPT-5.6·Gemini 3.1 Pro)을 더한 5모델 앙상블 0.70('상당한 일치'). OpenAI GPT-5.6 모델이 다른 모델과 상대적으로 낮게 일치(ChatGPT 포함 쌍 평균 κ 0.58 대 그 외 0.83)하여 벤더 간 차이가 관찰되었다. 이는 코드북의 재현성을 뒷받침하되 연구자 신뢰도의 대체물은 아니다. 자동 프록시는 개별 목표의 세밀한 수준 판정에서 연구자 코딩을 대체하기 어려우므로 결과 해석은 총량·방향에 한정하였다. 남은 한계로 (1)비영어 인터페이스 보고서의 진척도 표준필드 미추출, (2)증거수준 4·5단계 세부 구분의 신뢰도가 상대적으로 낮다는 한계가 있다. 개별 목표 수준의 확정적 신뢰도는 별첨 코드북에 따라 서로 독립적인 연구자 코더 2인 이상이 전수 재코딩하여 산정할 수 있으며, 이는 후속 과제로 남긴다. 5개 언어모델과 규칙 기반 프록시 간의 쌍별 일치도(이차 가중 κ) 행렬은 부록 〈그림 A1〉과 같다.

〈부록 그림 A1〉 코더 간·자동 프록시 간 일치도(이차 가중 κ) 히트맵



주: 붉은 선 안쪽 5개 대규모 언어모델 대비 규칙 기반 프록시(맨 아래, 오른쪽 행)의 낮은 일치가 대비됨

〈부록 표 A1〉 분석대상 134개국의 지역·소득군 분류 및 기본 특성

국가	지역	소득군	OECD	NBSAP	목표수
Albania	중·동유럽	상위중소득	–	개정중	29
Algeria	아프리카	상위중소득	–	개정완료	24
Angola	아프리카	하위중소득	–	–	23
Armenia	중·동유럽	상위중소득	–	개정완료	30
Australia	아시아·태평양	고소득	○	개정완료	9
Austria	서유럽·기타	고소득	○	개정완료	37
Bahamas	중남미	고소득	–	개정중	7
Bahrain	아시아·태평양	고소득	–	개정중	23
Bangladesh	아시아·태평양	하위중소득	–	미개정	23
Belarus	중·동유럽	상위중소득	–	개정완료	19
Belgium	서유럽·기타	고소득	○	개정완료	104
Belize	중남미	상위중소득	–	개정완료	30
Benin	아프리카	하위중소득	–	개정완료	26
Bhutan	아시아·태평양	하위중소득	–	개정완료	20

Bolivia	중남미	하위중소득	–	개정중	17
Botswana	아프리카	상위중소득	–	개정중	23
Brazil	중남미	상위중소득	–	개정완료	25
Brunei Darussalam	아시아·태평양	고소득	–	개정완료	18
Bulgaria	중·동유럽	고소득	–	개정완료	11
Burkina Faso	아프리카	저소득	–	개정완료	23
Burundi	아프리카	저소득	–	개정완료	25
Cambodia	아시아·태평양	하위중소득	–	개정완료	23
Cameroon	아프리카	하위중소득	–	개정완료	28
Canada	서유럽·기타	고소득	○	개정완료	23
Chad	아프리카	저소득	–	개정완료	20
Chile	중남미	고소득	○	개정완료	39
China	아시아·태평양	상위중소득	–	–	27
Colombia	중남미	상위중소득	○	–	6
Congo	아프리카	하위중소득	–	개정완료	25
Cook Islands	아시아·태평양	고소득	–	–	16
Costa Rica	중남미	상위중소득	○	개정중	23
Cote d'Ivoire	아프리카	하위중소득	–	개정완료	21
Cuba	중남미	상위중소득	–	개정완료	24
Cyprus	서유럽·기타	고소득	–	–	14
Czechia	중·동유럽	고소득	○	개정완료	10
Dominican Republic	중남미	상위중소득	–	개정중	23
DR Congo	아프리카	저소득	–	개정완료	25
Ecuador	중남미	상위중소득	–	개정완료	23
El Salvador	중남미	상위중소득	–	–	10
Equatorial Guinea	아프리카	상위중소득	–	개정완료	25
Estonia	중·동유럽	고소득	○	미개정	24
Eswatini	아프리카	하위중소득	–	개정중	57
Ethiopia	아프리카	저소득	–	개정완료	19
European Union	서유럽·기타	고소득	–	개정완료	45
Finland	서유럽·기타	고소득	○	개정중	23
France	서유럽·기타	고소득	○	개정완료	40
Gabon	아프리카	상위중소득	–	개정중	23
Gambia	아프리카	저소득	–	개정중	24
Germany	서유럽·기타	고소득	○	개정완료	59
Ghana	아프리카	하위중소득	–	개정완료	23
Greece	서유럽·기타	고소득	○	–	13
Guatemala	중남미	상위중소득	–	개정중	1

Guinea	아프리카	하위중소득	–	개정완료	4
Guyana	중남미	고소득	–	개정중	21
Honduras	중남미	하위중소득	–	–	21
Hungary	중·동유럽	고소득	○	–	78
Iceland	서유럽·기타	고소득	○	–	21
India	아시아·태평양	하위중소득	–	개정완료	23
Indonesia	아시아·태평양	상위중소득	–	개정완료	20
Iraq	아시아·태평양	상위중소득	–	개정중	7
Ireland	서유럽·기타	고소득	○	–	23
Israel	서유럽·기타	고소득	○	–	20
Italy	서유럽·기타	고소득	○	개정완료	21
Japan	아시아·태평양	고소득	○	개정완료	40
Jordan	아시아·태평양	하위중소득	–	개정완료	23
Kazakhstan	아시아·태평양	상위중소득	–	–	8
Kenya	아프리카	하위중소득	–	–	25
Kyrgyzstan	아시아·태평양	하위중소득	–	개정완료	35
Lesotho	아프리카	하위중소득	–	개정완료	19
Liberia	아프리카	저소득	–	개정완료	21
Luxembourg	서유럽·기타	고소득	○	개정완료	18
Madagascar	아프리카	저소득	–	개정완료	23
Malawi	아프리카	저소득	–	개정완료	23
Malaysia	아시아·태평양	상위중소득	–	개정완료	17
Mali	아프리카	저소득	–	개정중	16
Malta	서유럽·기타	고소득	–	–	22
Marshall Islands	아시아·태평양	상위중소득	–	–	3
Mexico	중남미	상위중소득	○	개정완료	47
Micronesia	아시아·태평양	하위중소득	–	개정완료	35
Moldova	중·동유럽	상위중소득	–	–	23
Mongolia	아시아·태평양	하위중소득	–	개정완료	20
Morocco	아프리카	하위중소득	–	–	25
Mozambique	아프리카	저소득	–	–	28
Namibia	아프리카	상위중소득	–	개정완료	23
Nepal	아시아·태평양	하위중소득	–	개정중	44
Netherlands	서유럽·기타	고소득	○	개정완료	60
New Zealand	아시아·태평양	고소득	○	개정중	13
Niger	아프리카	저소득	–	–	18
Nigeria	아프리카	하위중소득	–	–	23
Norway	서유럽·기타	고소득	○	개정완료	23

Oman	아시아·태평양	고소득	–	개정중	0
Pakistan	아시아·태평양	하위중소득	–	개정중	23
Palau	아시아·태평양	상위중소득	–	개정완료	23
Panama	중남미	고소득	–	–	30
Paraguay	중남미	상위중소득	–	개정완료	30
Peru	중남미	상위중소득	–	개정완료	32
Philippines	아시아·태평양	하위중소득	–	개정완료	23
Poland	중·동유럽	고소득	○	–	1
Qatar	아시아·태평양	고소득	–	–	15
Republic of Korea	아시아·태평양	고소득	○	개정완료	21
Russian Federation	중·동유럽	고소득	–	개정중	24
Saint Vincent & Grenadines	중남미	상위중소득	–	개정완료	14
Samoa	아시아·태평양	하위중소득	–	개정중	22
Saudi Arabia	아시아·태평양	고소득	–	개정완료	17
Senegal	아프리카	하위중소득	–	개정중	4
Seychelles	아프리카	고소득	–	개정중	23
Sierra Leone	아프리카	저소득	–	–	23
Singapore	아시아·태평양	고소득	–	개정중	20
Slovakia	중·동유럽	고소득	○	개정완료	6
Somalia	아프리카	저소득	–	–	23
South Africa	아프리카	상위중소득	–	개정중	19
South Sudan	아프리카	저소득	–	개정중	24
Spain	서유럽·기타	고소득	○	개정완료	22
Sri Lanka	아시아·태평양	하위중소득	–	개정중	1
Sudan	아프리카	저소득	–	개정완료	23
Suriname	중남미	상위중소득	–	개정완료	19
Sweden	서유럽·기타	고소득	○	개정완료	60
Switzerland	서유럽·기타	고소득	○	개정완료	10
Tanzania	아프리카	하위중소득	–	개정완료	23
Thailand	아시아·태평양	상위중소득	–	개정완료	12
Togo	아프리카	저소득	–	–	25
Trinidad and Tobago	중남미	고소득	–	개정완료	23
Tunisia	아프리카	하위중소득	–	개정완료	23
Turkiye	서유럽·기타	상위중소득	○	개정중	7
Uganda	아프리카	저소득	–	개정완료	37
United Arab Emirates	아시아·태평양	고소득	–	개정완료	15
United Kingdom	서유럽·기타	고소득	○	개정완료	23
Uruguay	중남미	고소득	–	개정중	23

Uzbekistan	아시아·태평양	하위중소득	–	개정중	23
Vanuatu	아시아·태평양	하위중소득	–	개정완료	23
Venezuela	중남미	상위중소득	–	개정완료	11
Yemen	아시아·태평양	저소득	–	개정완료	18
Zambia	아프리카	하위중소득	–	–	23
Zimbabwe	아프리카	하위중소득	–	개정완료	23

주: 소득군은 세계은행 소득분류, 지역은 유엔 지역그룹 구분에 따름. OECD ○=가입국, NBSAP는 제7차 국가보고서 기준(– = 판독불가). 목표수는 추출된 국가목표 수. 국가명은 영문 표기.

이후승: 영국 글라스고우대학교에서 생태학 박사학위를 취득하고 캐나다 야생생물센터(Center for Wildlife Ecology)와 미국 해양대기청(NOAA)을 거쳐 현재 한국환경연구원 자연환경연구실 실장으로 재직 중이다. 생활사 전략(life-history strategy), 생물다양성, 기후변화, 재생에너지, 국제협약, 생태모델 및 환경영향평가 등에 관한 연구를 계량, 통계 및 모델링 등 다양한 방법론을 사용하여 수행하고 있다(wslee@kei.re.kr).

투 고 일: 2026년 07월 15일
심 사 일: 2026년 07월 20일
게재확정일: 2026년 07월 23일