



International Baccalaureate®
Baccalauréat International
Bachillerato Internacional

평가 원칙 및 수행: 디지털 시대에서의 양질의 평가



International Baccalaureate®
Baccalauréat International
Bachillerato Internacional

평가 원칙 및 수행: 디지털 시대에서의 양질의 평가

평가 원칙 및 수행: 디지털 시대에서의 양질의 평가

2026년 4월 발행된 “Assessment principles and practices:
Quality assessment in a digital age”의 한국어 번역본
발행 2026년 5월

이 문서를 번역하고 발행할 수 있도록 지원해주신 대한민국 대구광역시교육청과
제주특별자치도교육청에 감사드립니다.

본 간행물은 스위스 제네바(주소: Rue du Pré-de-la-Bichette 1, 1202 Genève,
Switzerland)에 본부를 둔 비영리 교육재단 국제 바칼로레아 기구(International
Baccalaureate Organization)에 의하여 출판되었습니다.

홈페이지: ibo.org

© International Baccalaureate Organization 2026

국제 바칼로레아 기구(International Baccalaureate Organization, 이하 IB)는 더 나은
평화로운 세상의 구현을 목표로 하여 네 개의 도전적인 높은 수준의 교육을 전 세계
학교 사회에 제공합니다. 본 간행물은 이러한 프로그램을 지원하기 위해 제작된 여러
자료 중 하나입니다.

IB의 출판물은 위키백과(Wikipedia)와 같은 커뮤니티 기반 자료를 비롯한 다양한
정보를 사용할 수 있으며, 이에 따라 정보의 정확성 및 신뢰성을 검증하기 위해 엄격히
점검합니다. IB는 지식재산의 원칙을 존중하며, 사용된 모든 자료의 지식 재산권
소유자로부터 허가를 받기 위해 최선을 다하고 있습니다. IB는 본 간행물에 사용된
자료에 대한 허가를 받은 것에 감사하며, 정보의 오류 및 누락이 있을 시 최대한 신속히
수정하겠습니다.

모든 권리 보유. **IB 지식 재산권 사용에 대한 규칙**에 명시적으로 규정된 경우를
제외하고, IB의 사전 서면 동의 없이 본 간행물의 일부를 어떤 방식으로든 검색 시스템에
저장하거나 재생산, 또는 전송해서는 안 됩니다.

IB 상품 및 간행물은 **IB 온라인 스토어**를 통해 구입할 수 있습니다. (이메일: sales@ibo.org)
ibo.org) IB와의 공식적인 관계없이 IB 환경 전반에서 활동하는 제3자는, 별도의 사용료를
지급한 경우를 포함하여 IB의 서면 라이선스 없이 그 어떤 상업적인 용도로도 IB
출판물을 사용할 수 없습니다. 여기서 제3자는 교육기관, 교원 연수 개발자, 교육 출판사,
교육 과정 지도 운영자, 온라인 교육자료 데이터베이스 등을 포함하며 이에 국한되지
않습니다. 라이선스 요청은 copyright@ibo.org에 문의 바라며, 자세한 내용은 **IB 공개
홈페이지**에서 확인할 수 있습니다.

국제 바칼로레아(IB)의 교육목표

IB의 목표는 서로 다른 문화를 이해하고 존중하며, 더 나은 평화로운 세상을 실현하는데 기여할 수 있는, 지식이 풍부하고 탐구심과 배려심이 많은 청소년을 기르는 것입니다.

이를 위해 본 기관은 학교, 정부 및 국제기구와 협력하여 국제적 수준의 교육과 엄격한 평가 시스템을 갖춘 도전적인 교육 프로그램을 개발하고 있습니다.

IB 프로그램은 전 세계 학생들이 적극적으로 공감할 줄 알며, 서로 다름을 이해하고 존중하는 평생 학습자가 될 것을 장려합니다.



IB 학습자상

모든 IB 프로그램의 목표는 인류의 공통 과제에 관심을 두고 세계를 함께 지켜나갈 책임을 다하는 청소년들이 국제적 소양을 갖춘 인재로 성장하여, 더 평화롭고 보다 나은 세상을 만들어나갈 수 있도록 돕는 것입니다.

IB 학습자로서 우리는 다음 역량을 추구합니다.

탐구하는 사람

우리는 호기심을 키워 탐구하고 연구하는 능력을 향상시킵니다. 우리는 독립적으로 또 다른 사람과 함께 배우는 법을 압니다. 우리는 열정을 가지고 배움에 임하며, 학습에 대한 열의를 늘 잃지 않습니다.

지식이 풍부한 사람

우리는 개념적 이해를 통한 성장을 지향하며, 다양한 학문의 지식을 탐구합니다. 우리는 지역적이고 세계적으로 중요한 사안들과 의견에 관심을 기울입니다.

사고하는 사람

우리는 비판적이고 창의적인 사고력으로 복잡한 문제를 분석하며 책임 있게 행동합니다. 우리는 합리적이고 윤리적인 의사결정을 주도합니다.

소통하는 사람

우리는 하나 이상의 언어와 다양한 방법으로 창의적이고 자신 있게 우리 자신을 표현합니다. 우리는 다른 개인과 집단의 의견을 경청하며 효과적으로 협력합니다.

원칙을 지키는 사람

우리는 공정성과 정의감을 바탕으로 인간의 존엄성 및 권리를 존중하며, 성실하고 정직하게 행동합니다. 우리는 우리 자신의 행동과 그 결과에 따른 책임을 집니다.

열린 마음을 지닌 사람

우리는 비판적인 사고를 통해 우리 고유의 문화와 역사를 바라보고 타인의 가치관과 전통을 수용합니다. 우리는 다양한 관점을 추구하고 평가하며, 경험을 통해 성장합니다.

배려하는 사람

우리는 서로 공감하고 격려하며 존중합니다. 우리는 봉사 정신을 갖고, 타인의 삶과 지역 사회에 긍정적인 변화를 도모합니다.

도전하는 사람

우리는 철저하게 계획하고 의사결정을 내려 불확실성에 도전하며, 독립적으로 또 협력을 통해 새로운 아이디어와 혁신적인 전략을 모색합니다. 우리는 도전과 변화에 맞서 굴복하지 않고 슬기롭게 대처해 나갑니다.

균형 잡힌 사람

우리는 자신과 타인의 행복을 위해 삶의 지적, 신체적, 정서적 균형을 이루는 것이 중요하다는 것을 알고 있습니다. 우리는 타인뿐 아니라 우리가 살아가는 세상과도 상호 의존함을 인지하고 있습니다.

성찰하는 사람

우리는 세상과 자기 생각 및 경험에 대해 깊게 생각합니다. 우리는 개인의 학습과 성장에 도움이 되도록 우리 자신의 강점과 약점을 이해하려고 노력합니다.

IB 학습자상은 IB 월드 스쿨이 중요시하는 열 가지 자질을 나타냅니다. 우리는 이러한 자질들이 개인과 단체가 지역과 국가를 넘어 글로벌 커뮤니티에서 책임감 있는 구성원으로 자라는 데 도움을 준다고 믿습니다.

목차

소개 및 개요	1
소개	1
추가 자료	3
본 자료 활용하기	5
평가 원칙 및 수행	13
‘좋은’ 평가	14
IB 평가에서 테크놀로지의 역할	23
섹션 A: 평가 원칙	27
교육에서의 평가	27
타당성	32
타당성 사슬의 요소	35
기준	53
총괄 평가에서 학생의 성취 설명하기	57
IB 평가 채점하기	60
윤리적 사고방식 기르기	63
섹션 B: IB 평가 수행	66
수행 정의하기	66
학생 성취도 보고하기	67
평가 과정: 역할과 책임	75
평가 공정성	80
모든 학생을 위한 공정성	84
평가 주기	88
평가 제작	90
채점	97
조정	108
조정: 교사에게 기대되는 사항	110
성적 부여 및 종합	114
품질 평가	123
IB 최종 학위 수여 위원회	124
결과 발표 준비	126
성적에 대한 문의 및 평가에 대한 이의 제기	129

학교에 제공되는 평가 피드백	134
섹션 C: IB 프로그램별 과정	136
프로그램별 과정 정의하기	136
모든 프로그램의 공통 요소	137
IB 학습자상	139
프로그램별 요구와 해결책	141
디플로마 프로그램	142
직업연계 프로그램	147
중등교육 프로그램	151
초등교육 프로그램	156
부록	159
내부 평가의 조정: 자세히 알아보기	159
참고 문헌	165
용어 해설	168
인쇄 가능한 자료	178

소개

국제 바칼로레아(IB) 프로그램은 다양한 교육적 배경과 전통을 아우르며, 150개가 넘는 국가와 지역에서 시행되고 있습니다. IB 평가 철학과 방법이 익숙한 사람도 있는 반면, 어떤 이들에게는 이 체계가 복잡하고 낯설게 느껴질 수도 있습니다.

IB는 단순히 측정하기 쉬운 것을 평가하기보다, 진정으로 중요한 것을 평가하겠다는 원칙에 따라 도전적인 과제를 수행하고 있습니다. IB의 초대 사무총장인 알렉 피터슨(Alec Peterson)은 이러한 교육 생태계의 도전 과제는 균형을 이루어야 한다고 설명했습니다.






필요한 것은 가능한 한 타당한 평가 과정이다. 즉, 학생이 삶의 다음 단계에서 의미를 갖는 방식으로 학생의 전반적인 역량과 잠재력, 그리고 인성을 진정으로 평가하면서도, 동시에 학생, 학부모, 교사 및 진학 기관 모두 평가가 공정하게 이루어졌다고 확신할 수 있을 만큼 충분히 신뢰할 수 있는 평가가 필요하다. 그러나 이러한 평가 과정이 역류 효과로 인해 좋은 교수법을 왜곡하거나, 너무 느리거나, 부족한 교육 자원을 지나치게 소모해서는 안 된다.

(Peterson, 1971, pp. 27-55)

본 간행물은 IB가 프로그램 전반에서 평가가 의미 있고 공정하며 학생들에게 실질적으로 도움이 될 수 있도록 하기 위해 사용하는 원칙들을 알기 쉽게 설명하는 것을 목표로 합니다. 이는 학습 및 교수 과정을 반영하고, 실제 학습 맥락과 밀접하게 연결된 방식으로 이루어집니다.

IB에서 평가는 [그림 1]과 같이 크게 2가지 유형으로 구분됩니다.

[그림 1] 총괄 평가 및 형성 평가

목적	 총괄 평가: 숙달 정도를 입증하거나 학생 능력을 측정하기 위한 형성 평가: (학년 차원 또는 개별적으로) 개발이 필요한 영역 또는 중점을 두어야 할 영역을 파악하기 위한
결과	 총괄 평가: 결과물 형성 평가: 과정
평가 시기	 총괄 평가: 학습의 마무리 단계 형성 평가: 학습 과정 전반에 걸쳐
근거 제시 방법	 총괄 평가: 포트폴리오, 시험, 과제물(보통 개별적) 형성 평가: 관찰, 종료 티켓, 짝과의 공유 활동, 자기 평가(보통 협력적)
평가 도구	 총괄 평가: 기준표 및/또는 점수 구간표 형성 평가: 학생의 발전과 학습 진행을 체계적으로 지원하는 루브릭

이 두 평가의 각각의 역할을 이해하는 것은 IB 공동체에게 매우 중요합니다.

평가 결과는 학생의 향후 진학 또는 진로에 영향을 미치며, 그들의 삶에 중대한 영향을 끼칩니다. 평가는 긍정적인 결과와 부정적인 결과 모두를 초래할 수 있는 도구입니다. 교사, 학부모, 법적 보호자, 또는 학생으로서 평가의 강점과 약점, 그리고 그 배경을 이해하는 것은 매우 중요합니다. 본 간행물은 다음과 같이 자주 제기되는 질문과 우려 사항을 다룹니다.

- “IB가 그렇게 하는 이유는 무엇입니까?” - IB 평가 수행 이면의 원칙 이해하기
- “불공평해요. 저는 더 높은 성적을 받았어야 해요.” - 평가 결정의 공정성과 그 영향 설명하기
- “그렇게 생각해 본 적이 없습니다” - 평가 과정 분석하기
- “더 나은 교수를 위해서만 평가를 활용하고 싶습니다” - 형성 및 총괄 평가 모두에서 양질의 평가의 중요성 강조하기

과학 기술은 끊임없이 변화하고 있으며, IB 평가는 이러한 변화를 적극적으로 수용함으로써 학생의 참여, 교육, 학습 방식이 진정성 있게 반영되도록 유지하는 것이 매우 중요합니다. 본 문서의 제목인 ‘디지털 시대에서의 양질의 평가’라는 문구에서도 나타나듯이 IB는 평가가 적합성을 유지하려면 학생이 익숙하고 편안한 방식으로 평가가 진행되어야 함을 인식하고 있습니다. 더욱 중요한 점은, 이러한 디지털 전환의 과정에서도 IB는 IB의 근간이 되는 기본 원칙에 충실해야 함을 인식하고 있다는 것입니다. 즉, 과학 기술은 학습과 평가를 지원함으로써 보다 타당한 평가를 가능하게 하는 수단으로 활용됩니다. ‘디지털 시대에서의 양질의 평가’라는 문구는 변화하는 교육 환경에 대한 IB의 적극적이고 신속한 대응과 함께 평가의 근본적인 원칙에 대한 확고한 헌신을 강조합니다.

본 간행물은 IB의 초등교육 프로그램(PYP), 중등교육 프로그램(MYP), 디플로마 프로그램(DP), 그리고 직업연계 프로그램(CP)을 포함한 모든 IB 프로그램에 적용되는 평가 원칙을 상세히 설명합니다. 이러한 원칙은 모든 유형의 IB 평가를 포괄하며, 평가가 현재와 같은 방식으로 설계된 이유를 설명합니다. 일부 섹션은 IB에 제출되어 외부 채점을 받는 평가를 포함하여 총괄 평가에 특히 초점을 두고 있지만, 이 문서에서 제시하는 통찰은 프로그램 전반의 모든 교육 실천을 향상시키는 데 목적이 있습니다. 외부 총괄 평가가 시행되지 않는 경우라도, 효과적이고 공정한 평가의 요소를 이해하는 것은 수업의 질을 높이고 학생들이 향후 평가에 더 잘 대비하는 데 도움이 됩니다. 또한 본 간행물은 교사, 교장단, 채점관, 대학교 및 대학 입학 사정관을 포함한 IB 공동체 전체를 위해 평가 원칙을 쉽게 풀어 설명하는 것을 목표로 합니다. 이는 IB와 그 평가 방식에는 익숙하지만 평가의 기저 이론을 깊이 있게 다뤄본 경험이 없는 이들을 위해 작성되었습니다. 이 자료를 통해 IB 평가 원칙이 학생들에게 세계적 수준의 경험을 제공하고, IB의 더 넓은 교육적 의도를 어떻게 뒷받침하는지 이해하는 데 도움이 되기를 바랍니다.

용어

IB 평가에서는 교사와 학생에게 다소 생소할 수 있는 다양한 용어가 사용됩니다. IB의 모든 간행물은 혼동을 최소화하면서도 정확성을 유지하기 위해 세심한 노력을 기울이고 있습니다.

이러한 노력의 일환으로 본 문서의 각 섹션에서는 약어가 처음 등장할 때마다 풀어서 씬으로써 명확한 의미 전달을 도모하고 있습니다. 마찬가지로 중요한 용어를 처음 사용할 때는 용어에 대한 설명을 제시하거나 링크를 통해 제공합니다.

또한 본 간행물에는 평가 관련 용어 전체를 수록한 용어 해설이 포함되어 있습니다. 일반적인 평가 용어 외에 과목별 용어에 관한 설명은 각 IB 교과 가이드에서 확인할 수 있습니다.

추가 자료

이 자료는 IB가 교육에 대한 접근 방법을 설명하기 위해 간행한 여러 자료 중 하나입니다. 이러한 주요 자료들 간의 상호 관계를 이해하는 것은 종종 혼란스러울 수 있습니다. <표 1>과 <표 2>에는 주요 자료의 요약과 각 자료의 목적 및 역할에 대한 설명을 제공합니다.

<표 1> 모든 프로그램에 공통으로 활용되는 주요 IB 간행물

주요 IB 간행물 – 프로그램 공통	
「평가 원칙 및 수행: 디지털 시대에서의 양질의 평가」	이 간행물은 IB가 평가에 대해 전반적으로 취하는 접근 방법과 이를 실제로 어떻게 적용하는지를 설명합니다. IB 교육에서의 총괄 평가(공식 시험) 요소에 중점을 둡니다.
「IB 교육이란 무엇인가?」	이 간행물은 IB 교육의 핵심을 명확히 전달하기 위해 작성되었으며, 모든 IB 프로그램의 토대를 이루는 교육적 이상을 설명합니다. 또한 IB의 교육철학을 설명함으로써, IB를 도입하는 학교나 오랜 기간 IB 월드 스쿨로서 IB 프로그램을 시행해온 학교 모두에게 실질적인 지원을 제공합니다.
「프로그램 기준 및 운영방침」	이 간행물은 학교와 IB가 네 가지 IB 프로그램 운영 성과를 평가하기 위한 일련의 기준을 제시합니다. 이 문서에는 프로그램 기준(모든 IB 프로그램을 시행하기 위해 학교가 충족해야 하는 일반 요건), 운영방침(모든 프로그램에 공통으로 적용되는 기준의 구체적 정의) 및 요구 사항(각 프로그램별 세부 요건)이 포함되어 있습니다.

<표 2> 프로그램별 주요 IB 간행물

주요 IB 간행물 – 프로그램별	
「초등교육 프로그램: 원리부터 실천까지」 「중등교육 프로그램: 원리부터 실천까지」 「디플로마 프로그램: 원리부터 실천까지」 “Career-related Programme: From principles into practice”(영문 제공)	이러한 간행물은 특정 IB 프로그램의 맥락에서 학습 및 교수에 중점을 두고 있습니다. 또한 각 프로그램의 요구 사항을 설명합니다.

주요 IB 간행물—프로그램별	
교과 가이드(MYP 및 DP에만 해당)	교과 가이드는 특정 과목, 과정 또는 학문(예: 지리, 시각 예술)에 대한 상세한 정보를 제공하며, 해당 과목의 목적, 목표, 교수요목, 내부 평가(IA) 기준 등을 제시합니다. 또한 일반적으로 학습 및 교수에 대한 과목별 추가 지침도 함께 제공합니다.
과목 연계성(PYP에만 해당)	과목 연계성은 학교가 자체 범위 및 학습 순서를 설계하는 데 활용할 수 있도록 마련된 예시 자료입니다. 필요에 따라 지역 공동체의 맥락과 요구 사항에 맞게 조정하여 사용할 수 있고, 또한 각 과목 및 구성요소에 대한 개념적 이해와 개념 예시를 제공합니다. 함께 수록된 과목 개요는 교과군 맥락에서 PYP 필수 요소를 명확히 드러내며, 초학문적 프로그램에서 교과군의 역할을 구체적으로 설명합니다.
“Middle Years Programme Assessment procedures”(영문 제공) 「디플로마 프로그램의 평가 절차」 “Career-related Programme Assessment procedures”(영문 제공)	이러한 간행물은 특정 IB 프로그램 평가를 시행할 때 반드시 준수해야 하는 규칙, 규정, 세부 절차를 설명합니다. 또한 IB 월드 스쿨이 따라야 하는 규칙의 근간이 되는 일반 규정도 포함하고 있습니다.
「시험 준비 정책」(MYP, DP, CP에만 해당)	이 간행물은 시험 준비에 관한 IB의 정책을 설명하며, 학교 내 IB 기밀 자료의 안전한 보관 및 시험 일정 변경이 포함됩니다.
「IB 시험 시행 지침서」(DP에만 해당하며, CP는 DP 버전을 사용) 및 “The conduct of IB Middle Years Programme on-screen examinations”(MYP, 영문 제공)	이 간행물은 코디네이터와 감독관에게 프로그램의 시험 운영 및 시행과 관련된 규정을 안내합니다. 모든 교사장에는 반드시 본 지침서의 사본이 비치되어야 합니다.

이러한 포괄적인 간행물 외에도 IB는 교사, 코디네이터 및 기타 이해관계자를 위해 보다 구체적인 지침이 담긴 자료를 제공합니다. 이 자료들은 [IB 웹 사이트](#) 및 [프로그램 자료실](#)에서 확인할 수 있습니다.

본 자료 활용하기

「평가 원칙 및 수행: 디지털 시대에서의 양질의 평가」에서 다루는 주제의 범위는 매우 넓습니다. 본 간행물은 독자의 관심과 필요에 따라 다양한 방식으로 읽고 활용할 수 있도록 설계되었습니다. 이 섹션의 **주제 질문**은 각 독자가 자신에게 가장 중요한 측면을 파악하기 위한 출발점이 될 수 있습니다.

주제 질문

「평가 원칙 및 수행: 디지털 시대에서의 양질의 평가」는 IB의 평가 접근 방법에 대한 포괄적인 개요를 제공하는 것을 목적으로 합니다. 그러나 많은 독자들은 평가 과정의 특정 측면에 대해 보다 구체적인 정보를 찾고자 할 것입니다. 이를 위해 주제 질문 뒤에는 독자가 궁금해할 만한 특정 주제에 초점을 둔 섹션 목록이 제시되어 있습니다. 이 목록은 참고용으로만 제공되며, 각 섹션을 읽어야 할 권장 순서를 의미하지는 않습니다.

본 간행물 마지막에 수록된 용어 해설을 활용하여 관심 있는 용어를 찾아볼 수도 있습니다. 또한 PDF 버전의 ‘찾기’ 기능을 사용하면 용어 해설 속의 관심 용어를 보다 쉽게 찾아볼 수 있습니다.

‘평가에 대한 IB의 접근 방법은 무엇인가?’

IB 평가의 이론적 기반에 관심이 있는 대학 입학 사정관, 교사, 이해관계자들을 위해 IB가 ‘좋은’ 평가를 어떻게 정의하는지, 평가 절차 수립 시 어떤 요인들이 의사 결정에 영향을 미치는지, 그리고 평가 과정에서 상충하는 요구를 어떻게 조율하는지에 관한 정보를 제공합니다.

자주 제기되는 질문은 다음과 같습니다.

- IB 평가란 무엇입니까? IB 평가를 특별하게 만드는 요소는 무엇입니까?
- 인공지능(AI)은 평가를 어떻게 변화시키고 있습니까?

이러한 질문들과 특히 관련이 있는 섹션은 다음과 같습니다.

소개 및 개요

- ‘좋은’ 평가
- 좋은 평가는 교육과정의 목표를 달성하는 데 도움을 준다
- 좋은 예측 가능성
- 좋은 평가는 다양한 평가 과제를 활용한다
- 교실 기반 평가 및 내부 평가의 역할
- 협력 작업 및 개별 점수
- 좋은 평가는 학생의 다양한 역량과 고차원적 사고 기능을 고려한다

- 고차원적 사고 기능
- IB 학습자상 및 기능 개발
- 국제적 소양과 문화 간의 이해
- 좋은 디지털 평가는 어떤 모습인가?

섹션 A: 평가 원칙

- 교육에서의 평가
- 평가 정의하기
- 평가 접근 방법
- 역류 효과 및 학습
- 타당성
- 타당성 정의하기
- 타당성 주장 만들기
- 타당성 유지하기
- 타당성 사슬의 요소
- 타당성의 여러 측면 간의 균형 맞추기
- 신뢰성
- 일관된 결과와 ‘올바른’ 결과
- 구성개념 적합성 및 진위성
- 관리 가능성
- 공정성과 편향
- 비교 가능성
- 타당성에 대한 IB의 접근 방법
- 디지털 평가가 타당성에 미치는 이점
- 기준
- 기준의 세 가지 측면
- 상대 평가 및 절대 평가
- 상대 평가
- 절대 평가
- 성취 기준 평가

- 기준 유지하기
- 총괄 평가에서 학생의 성취 설명하기
- 성적의 영향
- 전문적 판단의 중요성
- IB 평가 채점하기
- 채점 정의하기
- 채점 방법
- 형성 평가 채점하기

섹션 B: IB 평가 수행

- 학생 성취도 보고하기
- IB 성적의 의미
- 점수와 성적의 차이점
- 성공적인 시험 기수
- 성적과 성취도
- 학문적 진실성 정의하기
- 모든 학생을 위한 공정성

인쇄 가능한 자료

- ‘IB의 평가 원칙’(PDF)

“저는 PYP 교사라서 평가를 수행하지 않습니다. 이 간행물이 저에게 어떤 도움이 되나요?”

학교 기반의 총괄 평가 및 형성 평가 기회에 대해 이해하고자 하는 PYP 교사를 위해 작성되었습니다.

PYP에는 IB에 제출하는 평가는 없지만, 학교 기반의 총괄 평가와 형성 평가가 있으며, 이 간행물은 이러한 영역에서의 모범 사례를 제시합니다. 학문적 진실성과 윤리적 사고방식 함양에 관한 정책 역시 유아·유치부 단계부터 학생의 발달 단계에 맞는 방식으로 다루어야 할 중요한 주제입니다.

이러한 질문들과 특히 관련이 있는 섹션은 다음과 같습니다.

섹션 A: 평가 원칙

- 평가 접근 방법
- 형성 평가 채점하기
- 윤리적 사고방식 기르기
- 평가에서 인공지능의 윤리적 사용

섹션 B: IB 평가 수행

- 학문적 진실성 정의하기
- 모든 학생을 위한 공정성

섹션 C: IB 프로그램별 과정

- 모든 프로그램의 공통 요소
- 초등교육 프로그램

“저는 MYP 교사이며, 이미 디지털 방식으로 평가하고 있습니다. 이 간행물이 저에게 어떤 도움이 되나요?”

본 간행물은 디플로마 프로그램 및 직업연계 프로그램(DP/CP)에서 디지털 평가 전환을 다루는 요소가 있지만, MYP 컴퓨터 기반 평가의 실제 가치를 뒷받침하는 내용도 포함하고 있습니다. 따라서 이 간행물은 MYP에도 그대로 적용됩니다. 다음의 주제 질문을 참고하십시오. “저는 IB 제출형 평가의 채점 방식과 성적 부여 방식에 관심이 있는 MYP, DP 또는 CP 교사입니다. 학교는 어떤 피드백을 받을 수 있나요?”

“디지털 평가로 인해 DP/CP 평가는 어떻게 달라지나요?”

IB 평가에 익숙하며 IB가 디지털 평가로의 전환을 어떻게 관리하고 있는지에 관심이 있는 교사를 위해 작성되었습니다.

자주 제기되는 질문은 다음과 같습니다.

- 디지털 평가란 무엇입니까?
- 평가 과제는 컴퓨터 기기로 채점됩니까?
- 디지털 평가가 더 나은 이유는 무엇입니까?
- 동일한 기준이 적용됩니까?

이러한 질문들과 특히 관련이 있는 섹션은 다음과 같습니다.

소개 및 개요

- ‘좋은’ 평가
- IB 평가에서 테크놀로지의 역할
- 좋은 디지털 평가는 어떤 모습인가?
- 디지털 평가와 관련된 위험 요소

섹션 B: IB 평가 수행

- 모든 학생을 위한 공정성
- 평가 주기
- 디지털 평가가 평가 주기에 미치는 영향

“저는 IB 제출형 평가의 채점 방식과 성적 부여 방식에 관심이 있는 MYP, DP 또는 CP 교사입니다. 학교는 어떤 피드백을 받을 수 있나요?”

학생들의 평가 과제가 IB에 채점을 위해 제출된 이후 어떤 과정을 거쳐 성적이 산출되는지 알고자 하는 교사를 위해 작성되었습니다. 또한 IB가 평가 과정에 교사의 평가를 포함하면서도, 왜 교사의 판단을 항상 그대로 수용하지는 않는지에 대한 이유를 이해하고자 하는 교사에게도 적합합니다.

자주 제기되는 질문은 다음과 같습니다.

- 이번 연도에 성적 급간이 달라진 이유는 무엇입니까?
- 다른 채점관을 선택할 수 없는 이유는 무엇입니까?
- 누가 답안지를 채점합니까?
- 채점관은 어떻게 검증됩니까?
- 어떻게 이의를 제기할 수 있습니까?
- 저는 학생에게 다른 점수를 부여했을 텐데, 이 결과가 어떻게 공정합니까?
- 채점 기준표에 따르면 학생의 답안은 점수를 받을 수 있는데, 이 결과가 어떻게 공정하다고 할 수 있습니까?
- ‘점수 없는 교실’이 학생에게 최선의 방법이 아닙니까?
- 제 점수가 변경된 이유는 무엇입니까?
- 해마다 조정 인자가 변하는 이유는 무엇입니까?

이러한 질문들과 특히 관련이 있는 섹션은 다음과 같습니다.

소개 및 개요

- 교실 기반 평가 및 내부 평가의 역할

섹션 A: 평가 원칙

- 교육에서의 평가
- 평가 정의하기
- 평가 접근 방법
- 역류 효과 및 학습
- 타당성 사슬의 요소
- 성적의 영향
- IB 평가 채점하기

섹션 B: IB 평가 수행

- 학생 성취도 보고하기
- IB 성적의 의미

- 점수와 성적의 차이점
- 성공적인 시험 기수
- 성적과 성취도
- 예측 성적
- 평가 과정: 역할과 책임
- 책임 채점관 및 수석 채점관
- 기타 채점관 역할
- IB 직원의 책임
- 채점관 위계 체계
- 모든 학생을 위한 공정성
- 점수와 성적
- 채점 접근 방법
- 분석적 채점 기준표
- 전체론적 기준: 점수 구간표
- 표준화
- 품질 모델
- 연습 답안지
- 채점 자격 취득용 답안지
- 실제 채점 및 시드 답안지
- 성공적인 표준화의 지표
- 허용 오차
- 채점이 어렵거나 이례적인 답안지
- 학교와의 관계
- 문제 문항 그룹
- 종합
- 조정
- 조정 및 교사에게 기대되는 사항
- 학생 과제의 선택
- 특수한 상황

- 조정 인자를 산출할 수 없는 경우
- 역동적 표본 추출
- 성적 부여 및 종합
- 판단 기반 성적 급간과 보간된 성적 급간
- 성적 부여 과정에서 활용되는 근거
- 이번 연도의 학생 집단 고려하기
- 평가에 대한 피드백
- 답안지 근거 검토하기
- 결과에 대한 통계 검토하기
- 근거 간의 균형 맞추기
- 고정된 성적 급간
- 성적 급간 권장 사항 검토 및 승인
- 프로그램 수수료증 부여하기
- 교사 관찰자
- 성적 부여 과정의 원칙
- 품질 평가
- 학교에 제공되는 평가 피드백
- 내부 평가 피드백 양식

섹션 C: IB 프로그램별 과정

- 모든 프로그램의 공통 요소

“IB는 시험을 어떻게 고안하고 개발하나요?”

이 섹션은 시험이 어떻게 출제되는지 이해하고자 하는 교사와 시험의 수준에 대한 우려가 있는 이들을 위해 작성되었습니다.

자주 제기되는 질문은 다음과 같습니다.

- 누가 시험을 출제합니까?
- 출제된 시험은 어떻게 검증됩니까?
- 다른 언어로 된 시험지의 내용과 동일합니까?
- 시험은 과연 평가해야 할 것을 평가합니까?

이러한 질문들과 특히 관련이 있는 섹션은 다음과 같습니다.

소개 및 개요

- ‘좋은’ 평가
- 좋은 평가는 교육과정의 목표를 달성하는 데 도움을 준다
- 좋은 예측 가능성
- 좋은 평가는 다양한 평가 과제를 활용한다
- 교실 기반 평가 및 내부 평가의 역할

섹션 A: 평가 원칙

- 교육에서의 평가
- 평가 정의하기
- 평가 접근 방법
- 역류 효과 및 학습
- 타당성
- 타당성 정의하기

섹션 B: IB 평가 수행

- 시험 출제자의 역할
- 모든 학생을 위한 공정성
- 평가 제작
- 시험 출제 과정 개요
- 내용 승인까지의 과정
- 출제 및 내용 승인
- 구성 및 교정
- 편의성 승인
- 품질 관리
- 번역
- 특수 변형 시험지
- 점수와 성적

평가 원칙 및 수행

평가 원칙이란 IB에서 자격 요건과 평가를 설계, 수행, 채점하고 성적을 부여할 때 중요하다고 판단되는 기준을 의미합니다. 이러한 원칙은 IB 교육에서 무엇을 중요하게 여기는지에 대한 관점을 바탕으로 수립되며, 가장 중요한 원칙은 평가가 교육을 왜곡하지 않고 오히려 지지해야 한다는 것입니다.

평가 수행이란 IB가 이러한 원칙을 의미 있고 실질적으로 실현하는 방법을 말합니다. 평가 수행에는 ‘원칙을 지키자’는 IB의 철학을 유지하면서도, 실제 교육 현장에서 마주하는 상충하는 요구와 현실적인 제약을 함께 고려합니다.

다음은 IB 평가 원칙의 주요 내용을 요약한 것입니다.

총괄 평가와 형성 평가는 모두 다음을 충족해야 합니다.

- 의도한 목적에 타당해야 합니다. 이는 ‘구성개념 적합성’(올바른 대상을 평가하는 것), 신뢰성, 공정성(편향 없음), 다른 대안과의 비교 가능성, 학생, 학교, IB의 관리 가능성 간의 상충하는 요구를 균형 있게 고려해야 함을 의미합니다.
- 긍정적인 역류 효과를 가져와야 합니다. 즉, 평가는 양질의 학습과 교수를 촉진하는 방식으로 설계되어야 합니다.
- 가능한 한 넓은 범위의 학생에게 적합해야 하며 학생이 자신의 성취 수준을 드러낼 수 있어야 합니다.
- 모든 학생의 응시 기회가 보장될 수 있도록 필요한 모든 조치가 마련될 수 있게 설계되어야 합니다.
- 보다 넓은 IB 프로그램의 맥락 속에서 고려되어야 하며, 분리된 요소로 다루어져서는 안 됩니다.
- 학습의 동시성과 학생의 전반적인 학습 경험을 지원해야 합니다.
- IB의 넓은 교육 목표를 지원하고, 탐구하는 사람, 지식이 풍부한 사람, 사고하는 사람, 소통하는 사람, 열린 마음을 지닌 학습자를 기르는 데 기여해야 합니다.

평가 수행은 일반적으로 거시적 수준에서 설명되며 개별 과목, 학문, 프로그램의 맥락에 맞게 적절히 실행되어야 합니다. 이는 무엇을 해야 하는지 포괄적으로 제시하지만 일상적인 절차나 프로그램 및 과목별 시행 차이에 대해서는 구체적인 지침을 제공하지는 않습니다.

교사를 위해 인쇄 가능한 자료인 ‘IB의 평가 원칙’(PDF)을 참고하십시오.

‘좋은’ 평가

- ‘좋은 평가’에 대한 단일한 정의는 존재하지 않습니다. 좋은 평가는 주어진 평가의 목적과 여러 우선순위에 얼마만큼의 비중을 두는지에 따라 달라집니다.
- IB의 기본 원칙은 평가 가능한 것을 중요하다고 판단하기보다 중요한 것을 평가하는 것입니다.
- IB 평가는 학습과 교수에 긍정적인 역류 효과(backwash effect)를 가져오는 것을 목표로 합니다.
- 일반적으로 좋은 평가는 학생의 학습을 종합적이고 다각도로 입증하는 데 도움이 되는 다양한 수행 과제를 포함해야 하며, 시험뿐 아니라 보다 심화된 수업 기반 활동의 기회도 함께 제공해야 합니다.
- 다양한 학습 유형과 선호를 지닌 모든 학생을 포용하기 위해, 좋은 평가는 보편적 평가 설계 원칙에 기반하여 설계되고 개발되어야 합니다. 평가는 학습자의 참여, 정보 제시 방식, 수행 방식, 표현 방식 등 다양한 수단을 모두 고려해야 합니다.

IB의 관점에서 **좋은 평가는 유의미한 학습 성과를 우선시합니다.** 이러한 평가는 교육과정의 목표와 긴밀하게 부합하며 학생이 폭넓은 기능과 역량을 기를 수 있도록 지원합니다.

IB의 기본 원칙은 평가 가능한 것을 중요하다고 판단하기보다 **중요한 것을 평가하는 것**입니다. 이는 신뢰성 및 학생의 과제 부담 등 다른 고려 사항들과 균형을 이루어야 하며, 한 가지 접근 방법만으로는 모든 우선순위를 충족시키는 것은 어렵습니다. 특히 좋은 평가 설계는 총괄 평가 및 형성 평가에 따라 다르게 접근되어야 합니다. 이러한 원칙을 바탕으로 IB는 다음과 같은 평가를 좋은 평가로 간주합니다.

- 교육과정 목표 달성에 도움을 준다.
- 다양한 평가 과제를 활용한다.
- 학생의 다양한 역량과 고차원적 사고 기능을 고려한다.

다음 섹션에서는 이러한 특징들을 차례로 살펴봅니다.

좋은 평가는 교육과정의 목표를 달성하는 데 도움을 준다

- 평가는 양질의 학습과 교수법을 장려하는 긍정적인 역류 효과를 가져야 합니다.
- 평가는 예측 가능성이 높아야 합니다.

평가는 학습 및 교수와 분리된 요소로 여겨져서는 안 됩니다. IB 평가 결과는 총괄 평가에 기반하며 학습 및 교수에 직접적인 피드백을 제공하기 위한 목적이 아닙니다. 그러나 평가에 무엇이 포함되느냐가 실제 수업 내용에 영향을 미친다는 점은 잘 알려져 있으며, 이를 ‘역류 효과’라고 합니다.

IB는 모든 학생에게 가장 바람직한 교육적 결과를 이끌어내도록 평가를 설계해야 합니다. 학생 학습에 미치는 영향은 IB 평가 설계에서 필수적으로 고려되며, 구성개념 적합성과 더불어 IB가 타당성의 다양한 요소 간 균형을 이루는 데 중요한 기준이 됩니다.

교부담 평가가 학습과 교수에 미치는 영향을 긍정적으로 활용하기 위해, 좋은 교수법과 학생의 건설적 학습 참여를 유도하는 방식으로 평가를 설계할 수 있습니다(예: Murphy, 1999 참조).

IB 프로그램 전반에서 평가는 학습 경험과 긴밀히 연계되어 있으며, 이를 반영합니다. 구성주의적 일체화 접근 방법(Biggs, 1996)을 적용하여, 모든 요소는 의도한 학습 성과를 중심으로 설계됩니다. 이러한 학습 성과는 학생들을 가르치는 방식과 평가하는 방식 모두를 형성하고 이끌어, 학생들이 성공을 위한 준비를 잘 마칠 수 있도록 합니다. 자세한 내용은 「[IB에서의 평가 접근 방법](#)」에서 확인하시기 바랍니다.

IB 교육 목표에 명시되고 IB 학습자상에서 탐구되는 이상적인 학생의 특성은 **학생 학습에 관한 구성주의 이론**과 잘 부합합니다. 이 프레임워크에서 학생은 적극적으로 학습 과정에 참여하고, 자신의 학습에 책임을 지며, 탐구를 통해 지식, 이해 및 기능을 발전시킵니다.

예를 들어, 학생 자신의 문화가 아닌 다른 문화적 관점에 적극적이고 열린 태도로 참여하는 것은 여러 과목의 평가 요구 사항에서 명시적으로 기대되는 요소입니다. 배려와 공감과 같은 정서적 자질은 공식 평가에 포함하기 어렵지만, 전반적인 평가 시스템 내에서 반드시 반영되어야 합니다. 이는 PYP의 발표회, MYP의 공동체 참여 요소, DP의 창의, 활동, 봉사(CAS) 요구 사항, CP의 공동체 참여와 같이 평가되지 않는 프로그램별 교육과정 요소를 통해 부분적으로 달성할 수 있습니다.

특히 DP/CP 과정 설계에 있어 IB는 평가가 수행되는 방식이 교내 IB 과정의 학습에 상당한 영향을 미친다는 것을 알고 있으며, 예측 가능한 타당성(결과가 미래의 성과를 예측하는 정도)에 중점을 둡니다. 각 과목에 적용되는 평가 모델은 다양한 유형의 근거를 포함하도록 폭넓게 설계되며, 이를 통해 학생 성취와 학습을 뒷받침할 광범위한 근거를 제시함으로써 구성개념 적합성을 지원합니다.

IB는 평가가 학습 및 교수에 미치는 역류 효과를 충분히 인지하고 있으며, 동시에 학교가 IB 프로그램의 목표와 철학을 학생에게 심어줄 수 있는 교수법을 적용하도록 장려합니다.

IB는 주기적으로 연구를 수행하여 IB 프로그램이 향후 학습을 잘 대비하도록 설계가 성공적으로 이루어졌는지 평가합니다. IB의 다양한 연구에 대한 자세한 정보는 IB 웹사이트의 ‘[Research](#)’(연구) 섹션을 참고하십시오.

높은 예측 가능성

‘예측 가능성 (predictability)’이란 무엇이 발생할지 또는 언제 발생할지를 가늠할 수 있음을 의미합니다. 평가에서 예측 가능성이란 학교가 주어진 시험에서 어떤 유형의 문제가 언제 출제될지를 파악할 수 있는 능력을 의미합니다. 평가의 예측 가능성이 높으면 시험 문제가 교사가 실제로 가르치는 교육과정과 일치하게 되어 평가의 공정성을 높일 수 있습니다. 이를 통해 학생들은 문제의 내용이 달라질 수 있더라도 어떤 유형의 문제가 나올지 대략적으로 예상할 수 있습니다.

IB는 특정 교육과정이 개정되기 전까지, 학교가 교수법 선택에 기울인 노력이 충분히 보답받을 수 있도록 하고자 합니다. 전체 교수요목은 해당 평가가 요구하는 방식에 따라 검토되어야 합니다. 과정이 마무리될 무렵에는 낮은 예측 가능성(아직 출제되지 않아 시험에 나올 가능성이 높다고 학교가 판단하는 상황)이 불가피하게 발생하지 않도록 주의를 기울여야 합니다.

기본 원칙은 어떤 시험에서든 학생이나 학교 모두에게 전혀 예상치 못한 문제가 나오지 않아야 한다는 것입니다. 따라서 어떤 항목이든 출제되는 문제는 교과 가이드에 의해 명시적으로 뒷받침되어야 합니다. IB는 평가에서 가능하고 적절한 경우, 예측 가능성으로 인해 발생할 수 있는 문제를 최소화하고자 합니다. 이를 위해 IB는 사전에 준비한 답변을 가진 학생이 유리해지지 않도록 평가를 설계합니다.

궁극적으로, 낮은 예측 가능성을 줄이기 위해 모든 주제, 옵션, 텍스트를 시험할 충분한 방법을 확보하는 것이 평가 설계의 핵심입니다.

<표 3> 높은 예측 가능성과 낮은 예측 가능성

높은 예측 가능성	낮은 예측 가능성
문제의 모든 가능한 조합을 주제, 옵션, 텍스트 그리고 지시어를 기준으로 구성한다면 해당 과목에서 IB가 출제하는 문제 유형은 그 결과 안에 모두 포함될 것입니다.	교사는 반복되는 시험 기수에서 유사한 문제가 자주 출제되는 경향을 파악할 경우, 특정 유형의 문제에 대해 학생들을 과도하게 준비시키는 경향이 있을 수 있습니다. 이는 지식, 분석 및 평가 등 다양한 기능을 입증해야 하는 비교적 긴 답변을 요하는 문제에서 특히 문제가 됩니다.
문제 재사용에 예측할 수 없는 접근 방법.	비효과적인 평가 설계로 인해 출제 가능한 문제의 범위가 제한되는 경우가 있습니다(예: 질문 가능한 주요 주제가 하나뿐인 과정에서 특정 텍스트를 필수로 지정하는 경우). 이러한 설계는 구조적으로 낮은 예측 가능성을 초래합니다.

좋은 평가는 다양한 평가 과제를 활용한다

선다형 문제, 단답형 문제, 서술형 문제, 에세이, 프로젝트, 포트폴리오의 단일 과제, 연구 과제는 모두 평가 과제의 예시입니다.

평가 도구 또는 항목은 주제 또는 내용의 연속성에 기반하여, 혹은 편의를 위해 통합되는 하나 이상의 과제로 구성됩니다. 총괄 평가, 학생 과제 포트폴리오, 심화 협업 프로젝트, 연구 과제는 평가 도구 또는 항목의 예시입니다. 평가 과제의 개념과 항목의 개념은 일부 겹치는 부분이 있습니다. 때로는 학생이 항목에서 선택 가능한 여러 과제 중 하나만 수행할 수도 있습니다.

IB가 다양한 평가 과제와 항목을 사용하는 데에는 여러 가지 이유가 있습니다. 우선 역사적, 실용적 관점에서 Peterson(2003)은 DP 평가가 처음 개발될 당시를 두고 “IB 학생들이 가장 진학을 희망하던 여러 나라와 지역(영역)의 교육 기관에서 사용되던 다양한 평가 기법을 고려해야 할 의무와 동시에 기회도 있었다.”고 말하며, 이러한 원칙은 이후 MYP와 CP에도 확장되었습니다. 또한 평가의 목적 적합성과 관련된 타당성 측면에서도 다양한 평가 방식이 요구됩니다. 마지막으로 평가 기법의 다양성은 평가에서 발생할 수 있는 불공정성의 가능성을 줄이는 데 도움이 됩니다(Linn, 1992; Brown, 2002 참고). 다양한 항목과 그 안에 설정된 과제들은 과목 전체의 평가 모델을 종합적으로 고려할 때, 해당 과목의 모든 목표에 대해 학생 성취가 충분히 반영되도록 합니다.

교실 기반 평가 및 내부 평가의 역할

교실 기반 평가는 시험 방식으로는 평가하기 어려운 영역에서 학생을 평가할 수 있는 다양한 기회를 제공합니다. 교실 기반 평가의 가장 중요한 측면은 시험의 특징인 시간의 압박 없이 학생들이 문제를 탐구하고 사고 과정을 발전시키는 심화 과제를 수행할 수 있는 기회를 제공한다는 점입니다. 이는 교실 기반 평가를 통해서만 수행할 수 있는 다양한 유형의 평가 과제가 존재함을 의미합니다.

교실 기반 평가에 적합한 과제의 예로는 프로젝트 과제, 현장 연구, 실험실 실습, 공연, 수학적 조사 등이 있습니다.

국제적 수준의 자격의 맥락에서 볼 때, 내부 평가(IA)에는 이 밖에도 여러가지 **이점**이 있습니다. 내부 평가는 공통된 일련의 기능을 계속해서 다루면서도 주제를 훨씬 더 **유연하게** 선택할 수 있습니다. 이를

통해 학교는 **지역적, 문화적, 지리적 맥락**에서 학습하거나 교실과 교실 밖 세상을 더욱 긴밀하게 연결할 수 있습니다. 학생들의 문화적 배경이 학교가 위치한 지역의 문화와 다른 경우가 많은 국제 학교에서는 내부 평가 과제를 활용하여 지역 사회나 환경과의 관계를 형성할 수 있습니다.

또한 내부 평가는 개별 학생들에게 주제나 이슈를 스스로 선택할 기회를 제공하는 경우가 많습니다. 이를 통해 학생들은 특정 관심사를 추구하고 자신의 학습에 대해 더 큰 주도권을 갖게 됩니다. 이와 같은 유연한 접근 방법은 내부 평가를 학생 교육의 가치 있는 요소로 만들며, 평가 과정뿐 아니라 전체 학습 경험의 타당성까지도 높여 줍니다.

한편, 교실 기반 평가에는 **어려움** 또한 존재합니다. 가장 큰 어려움 중 하나는 **학문적 진실성**을 보장하는 것입니다. 표준화된(감독된) 조건 없이 과제가 수행될 때, 예를 들어 교실 밖에서 학생들이 IB의 학문적 진실성 정책을 위반하는 행위(다른 사람에게 자신의 과제를 대신 작성하게 하는 등)를 하지 않도록 보장하기가 훨씬 어렵습니다. 인터넷을 자유롭게 사용할 수 있게 되면서 이러한 문제는 기하급수적으로 커졌습니다. 이러한 이유로 IB는 학교 교직원과 교사가 학생 과제물의 진위를 파악하고, 학생이 스스로 작성하지 않은 과제를 구별하는 데 가장 적합한 위치에 있다고 생각합니다. 학문적 진실성에 관한 자세한 정보와 자료는 **IB 웹사이트**의 ‘Academic integrity(학문적 진실성)’ 페이지에서 확인할 수 있습니다.

교실 기반 평가가 직면한 또 다른 어려움은 교사와 학생 모두에게 상당한 **부담**을 초래할 수 있다는 점입니다. 교실에서 평가를 수행하면 가용 수업 시간이 줄어들고, 내부적으로 설정된 과제는 학생에게 상당한 시간 투자를 요구합니다. 교사가 내부 평가에 필요한 기능과 절차를 학생에게 지도하는 데 시간을 할애하는 것은 적절하지만, 학생과 교사 모두 필요 이상으로 내부 평가 과제를 반복 연습하려는 강한 유혹을 느낄 수 있으며, 이는 수업 시간을 더욱 감소시킵니다. DP 교과 가이드에서 내부 평가 과제를 지원하는 교사의 역할에 관한 자세한 지침을 제공합니다.

교실 기반 평가는 **외부적으로** 또는 **내부적으로** 평가할 수 있습니다. 외부 평가의 경우 학생이 수행한 과제가 IB로 제출되어 채점관에 의해 평가됩니다. 채점관의 판단이 적절히 이루어지는지 IB에서 확인합니다(본 간행물의 ‘**채집**’ 섹션 참고).

내부 평가의 경우 학교(보통 학생의 교사)가 학생의 과제를 채점하며, 이후 IB는 조정 과정을 통해 교사가 세계적 기준을 올바르게 적용했는지 확인합니다(본 간행물의 ‘**조정**’ 섹션 참고).

내부 평가에 대한 관점은 세계적으로 다양합니다. 일부 교육 시스템에서는 교사가 단일 평가에서 얻은 ‘**단편적인 결과**’가 아닌 학생의 성과에 대한 전인적인 의견을 제시할 수 있는 최적의 위치에 있다는 사실에 큰 가치를 둡니다. 반면 다른 교육 시스템에서는 학생의 성과를 기준으로 교사의 성과를 평가하기 때문에, 교사가 모든 내부 평가에서 높은 점수를 부여하려는 강한 동기가 생깁니다.

또한 교사의 판단은 학생의 이전 과제 경험에 의해 영향을 받을 수 있으며, 이로 인해 다양한 형태의 편향이 발생할 수 있습니다. 교사는 때때로 학생의 내부 평가를 지도하고 지원하는 과정에서 자신의 역할의 범위와 한계가 어디까지인지 명확하지 않을 수 있으며, 해당 교과군에서의 세계적 성취 기준을 충분히 이해하지 못하는 경우도 많습니다. 자신의 학생 과제를 평가할 때 교사는 본교의 일반적인 기준에 크게 영향을 받을 수도 있습니다. 설령 제출된 과제에 과도한 점수를 부여할 동기가 없더라도 교사와 학생 간에 형성된 사제 관계 자체가 객관적인 판단을 어렵게 만들 수 있습니다. 이러한 무의식적 편향에 관한 연구는 눈에 띄는 영향을 미칠 수 있음을 보여 줍니다(참고 예시: Zanga, De Gioannis, 2023).

IB는 학생의 발달 과정을 직접 관찰해 온 사람이 교실 기반 과제를 통해 의미 있는 평가 과제를 설계함으로써 얻는 이점이 내부 평가로 인해 발생할 수 있는 문제보다 더 크다고 믿습니다. 따라서 내부 평가 과제는 통상적으로 과목 평가 모델의 핵심 구성요소가 됩니다.

협력 작업 및 개별 점수

까다로운 문제 중 하나는 학생들이 함께 협력하여 완수하는 협력 과제를 어떻게 평가할 것인가입니다. 이는 학생이 미래의 직장에서 필연적으로 마주할 중요한 측면이지만 학업 평가에서 종종 간과됩니다.

어떤 상황에서는 그룹 활동에서 개인이 맡은 역할을 파악할 수 있습니다. 예를 들어 댄스 공연에서는 각 댄서가 직접 보여주는 기술을 관찰하여 개별적으로 평가할 수 있습니다. 이러한 경우 그룹 과제는 논쟁의 여지가 없습니다.

그러나 최종 결과물에서 각 구성원이 어떤 부분을 담당했는지 구분하기 어려운 경우도 있습니다. 이러한 경우 그룹 전체에 단일 점수를 부여할 수 있으나 이는 개별 기여도의 차이를 반영하지 못합니다. 한 학생이 실제로 과제에 더 많이 기여했다라도 다른 학생과 같은 점수를 받게 될 수 있습니다. IB 성적이 개별 학생의 선발 및 진로 결정에 활용된다는 점을 고려할 때, IB는 이러한 접근 방법이 공정하지 않다고 봅니다. 따라서 IB는 개별 학생의 성취를 분리하여 평가할 수 없는 그룹 평가 방식은 가급적 지양합니다.

좋은 평가는 학생의 다양한 역량과 고차원적 사고 기능을 고려한다

IB 교육은 학생이 단순히 ‘사실’을 습득하는 데 그치는 것을 목표로 하지 않습니다. 여러 정부가 학생들에게 21세기형 **기능, 직업 역량** 또는 유사한 **진취적 계획**을 제공해야 한다는 현재의 인식은 IB가 오랫동안 추구해 온 목표를 잘 반영합니다.

학생 역량에 대한 IB의 철학과 접근 방법은 **IB 학습자상**과 **국제적 소양**의 연계성에 초점을 맞춥니다. 프로그램 간 연속성을 탐구하는 ‘섹션 C: IB 프로그램별 과정’에서 이를 더 자세히 다룹니다. 유념해야 할 핵심 개념은 IB의 평가 방식이 단순한 학생의 교과 내용 지식만을 평가하지 않고, 현실 세계의 적용을 반영하는 다양한 역량을 평가한다는 점입니다. 신중한 설계를 통해 IB 평가는 미래의 도전 과제에 대비할 수 있는 **평생 학습자**를 기르는 것을 목표로 합니다.

따라서 좋은 평가는 각 과정에서 달성하고자 하는 **모든 범위의 결과물**을 고려하며, 학생이 다양한 영역에서 자신의 능력을 발휘할 수 있도록 합니다. 다만 실제 평가에서는 이러한 모든 결과물을 동시에 측정하는 것이 항상 바람직하거나 현실적으로 가능한 것은 아니지만, 좋은 평가는 이러한 한계를 균형 있게 고려합니다. IB는 이러한 결과물을 **고차원적 사고 기능**과 폭넓은 **학생 역량**, 그리고 **국제적 소양**과의 연계로 가장 효과적으로 분류합니다.

고차원적 사고 기능

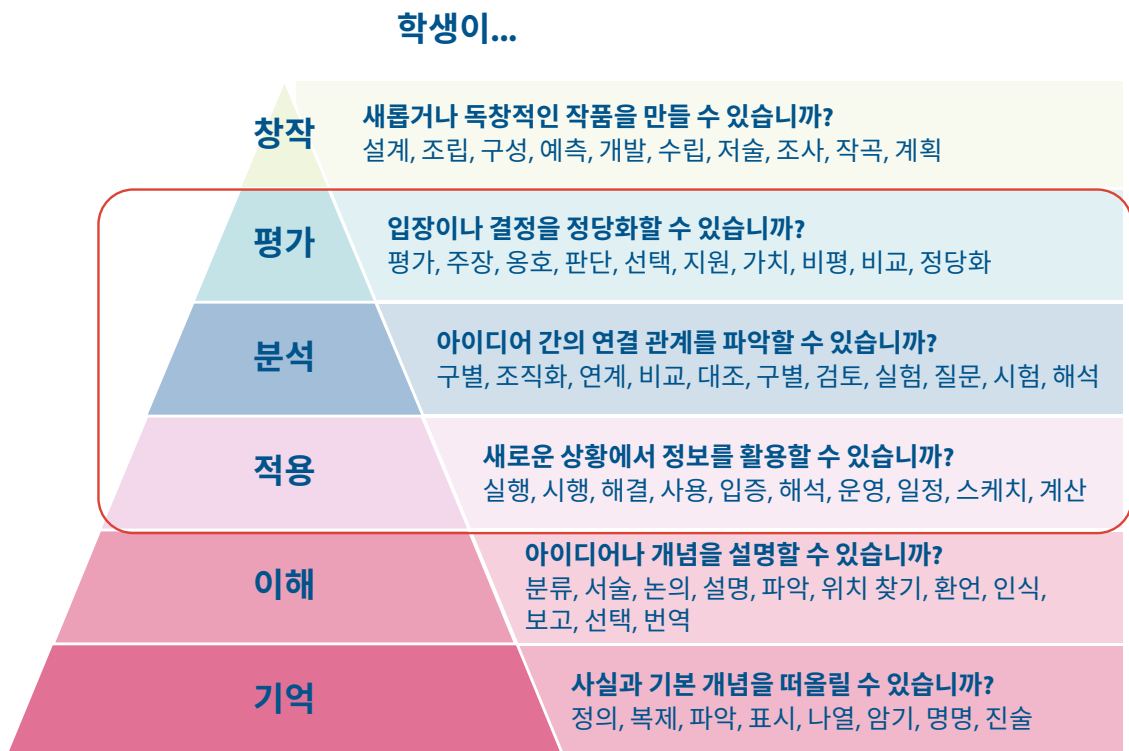
IB 평가는 단순한 지식 암기를 넘어 **평가와 분석**과 같은 고차원적 사고 기능을 측정하도록 설계되었습니다.

이는 IB 교육철학 형성에 결정적 영향을 미친 Alec Peterson(2003)의 주장에서도 확인할 수 있습니다. 그는 “사실 또는 기존에 해석된 사실을 그대로 받아들이거나 되새기는 것이 아니라 새로운 상황과 새롭게 제시되는 사실에 적용할 수 있는 사고력과 사고 방식을 발전시키는 것이 중요하다”고 밝혔습니다. 수가타 미트라(Sugata Mitra)는 2011년 싱가포르에서 열린 IB 학교장 컨퍼런스(IB Heads Conference)에서 피터슨의 주장을 한 단계 발전시켜, 인터넷을 통한 지식의 즉각적이고 포괄적인 접근성으로 인해 지식 자체의 가치가 크게 떨어졌다고 주장하며, 오히려 21세기 시민에게 요구되는 것은 지식을 분석하고 해석하며 새롭게 창출하는 능력이라고 강조합니다(Mitra, 2011).

IB 평가는 블룸(Bloom)의 교육목표 분류학(Bloom et al., 1956; Anderson, Krathwohl, 2001)과 같은 기존의 분류학을 기반으로 평가 과제의 성취 기대치를 설정함으로써 고차원적 사고 기능에 중점을

됩니다. 실제로 블룸의 교육목표 분류학의 목표는 평가 과제를 성공적으로 완수하고 만점을 받기 위해 필요한 기능과 인지 과정의 다양성을 설명하는 데 매우 유용한 프레임워크를 제공합니다. 블룸의 고차원적 기능은 반드시 다른 유형의 평가를 요구합니다. 학생의 분석, 종합 및 평가 기능은 어느 정도 충분한 분량의 분석, 종합 및 평가를 요구해야만 제대로 측정할 수 있습니다. 성과 평가는 이러한 영역에서의 학생 성과를 측정할 수 있는 유일한 현실적인 수단입니다. 이러한 활동의 결과를 엄격히 규정될 수 없으므로 평가는 상대적으로 비구조적이고 개방형이어야 하며, 다양하지만 올바른 답변을 허용해야 합니다.

[그림 2] 블룸(Bloom)의 교육목표 분류학



신뢰성과 타당성의 다른 측면에서 여러 도전 과제가 있음에도 불구하고, IB 평가가 의미 있으려면 이러한 기능에서 학생의 성과를 인정하고 점수를 부여할 수 있는 것은 매우 중요합니다. 지식, 개념, 자주 사용되는 기법의 암기만을 보상하는 시험은 이처럼 매우 중요한 고차원적 사고 기능을 충분히 다루지 못하므로 IB의 교육 목표에 부합하지 않습니다.

IB 학습자상 및 기능 개발

오늘날의 교육은 창의적이고 비판적인 접근을 통해 문제를 해결하고 의사결정을 내리는 사고방식에 더욱 초점을 두고 있다. 또한 교육은 의사소통, 협업과 같은 여러 업무수행 방식, 새로운 테크놀로지의 잠재력을 인식하고 활용하는 능력, 위험에 대비하는 능력 등 필요한 도구에 대한 설명이기도 하다. 마지막으로 교육은 다변화된 세상에서 활동적이고 참여도 높은 시민으로 살아가는 역량과도 관련성이 있다. 이러한 시민은 무엇을 배우고자 하는지, 그리고 어떻게 학습하고자 하는지에 영향을 미치며, 이는 교육자의 역할을 형성한다.

(Schleicher, 2016)

21세기에 요구되는 기능이 이전 세대와는 근본적으로 다르다는 주장이 늘고 있지만, 이러한 기능들을 뒷받침하는 탐구 접근 방법은 이미 소크라테스 시대부터 중요하게 여겨져 왔다는 견해도 있습니다.

그럼에도 불구하고, 학생들이 삶을 준비하는 데 필요한 다양한 자질을 기를 수 있도록 지원하는 것이 중요하다는 점에는 대체로 동의합니다. Llewellyn(2014)의 주장을 예로 들 수 있습니다.

이러한 기능을 분류하는 방법은 다양하며, 경제협력개발기구(Organisation for Economic Co-operation and Development, OECD)의 21세기 역량, 랜드 연구소(RAND) 교육 부문 및 전미연구평의회(National Research Council, NRC) 프레임워크 등이 대표적입니다. 그러나 IB에서는 이러한 역량들을 IB 학습자상 내에서 정의하고 설명합니다.

IB 학습자상의 모든 측면을 총괄 평가로 측정하는 것은 적절하지 않지만, 일부 측면은 고차원적 사고 기능의 개념에 내포되어 있습니다. 좋은 평가는 이러한 자질의 중요성을 인지하고, 형성 평가와 총괄 평가의 기회를 통해 학생들이 이러한 역량을 개발할 수 있도록 합니다. 대표적인 예로는 심리학 및 과학 과목에서 설문조사나 실험을 수행할 때 윤리적인 접근 방법을 장려하는 것, 성찰 프로젝트에서 윤리적 딜레마를 심층적으로 탐구하는 것, 적절한 동료평가를 지원하는 것, 그리고 학생들에게 예상치 못한 맥락을 제시하는 것 등이 포함될 수 있습니다.

IB의 학생 역량 접근 방법에 대한 자세한 정보는 [IB 웹사이트](#) 또는 관련 프로그램의 「원리부터 실천까지」 간행물에서 확인할 수 있습니다.

국제적 소양과 문화 간의 이해

IB의 교육 목표는 학생들이 ‘서로 다른 문화를 이해하고 존중하며, 더 나은 평화로운 세상을 실현하는데 기여할 수 있는, 지식이 풍부하고 탐구심과 배려심이 많은 청소년을 기르는 것’입니다(IB 교육 목표, 2024). IB 프로그램은 여러 국가와 지역(영역)에서 다양한 국적의 학생들이 함께 학습하고 있습니다. 따라서 IB 학습 및 교수에는 **국제적 맥락**과 **국제적 소양**을 기르는 데 중점을 두며, 이는 평가에서도 반드시 반영되어야 합니다. 이러한 다양한 관점을 실현하기 위해서는 평가 및 교육과정 개발자를 포함하여 서로 다른 문화적 배경을 지닌 교육 전문가들이 참여하는 것이 중요합니다. IB의 교육 목표에서 중요한 것은 서로의 다름을 외면하지 않고 학생들이 그 다름을 생산적이고 긍정적으로 탐구할 수 있도록 하는 것입니다.

일부 교과군에서는 다양한 문화적 관점을 인지하고 문화적으로 폭넓은 교육과정 내용을 강조함으로써 문화적 다양성을 장려합니다. 이러한 접근 방법의 예시는 생물학, 화학, 심리학 및 시각 예술과 같은 여러 DP 과목에서 확인할 수 있습니다. 이 가운데 생물학, 화학, 심리학의 경우 과목 내 선택 구조를 통해 학교가 해당 과목을 교수하는 데 있어 특정 문화적 전통에 적합한 학습 내용을 선택할 수 있습니다.

반면 예술, 언어와 문학과 같은 다른 교과군에서는 권장 및 지정된 교육과정 내용을 통해 국제적 소양이 장려되며, 또한 내부 평가 과제의 범위를 넓혀 강조될 수도 있습니다.

국제적 소양은 단순히 다른 문화에 대한 지식을 습득하고 이해하는 것에만 국한되지 않습니다. **태도와 행동** 역시 국제적 소양을 구성하는 중요한 자질입니다. 그러나 태도는 일반적인 학교 평가를 통해 평가하기 어렵습니다.

이러한 문제는 MYP의 공동체 프로젝트, DP의 창의, 활동, 봉사(CAS)와 같이 평가되지 않는 IB 프로그램의 핵심 요소를 통해 보완됩니다. 학생이 이러한 프로그램 핵심 요소를 완료해야만 IB 디플로마가 수여되기 때문에 비평가 요소 또한 IB 평가의 전반적인 결과에 중대한 영향을 미칩니다.

[그림 3] 원칙 중심적인 행동



학생이 답변할 질문을 선택하게 하는 것은 평가에서 다양한 국제적 요구 사항을 충족할 수 있는 최선의 방법처럼 보일 수 있습니다. 그러나 이 접근 방법은 여러 선택지 간의 **비교 가능성(comparability)**을 유지하는 데 평가상 어려움을 초래합니다. 이 문제는 특히 선택형 문항이 있거나 평가 과제가 매우 개방형일 때 주로 발생하며, 학생들의 교육적 배경이 크게 다를 경우 ‘동등한 요구 수준’이 무엇을 의미하는지 정의하기가 더욱 어려워집니다. 일반적으로 학생이 각자의 경험을 답안에 반영할 수 있는 공통 과제를 설정하는 것이 비교 가능성을 유지하는 데 더 용이합니다. 이러한 경우에는 공통 기준을 유지하는 책임이 채점관에게 주어집니다.

비교 가능성에 대한 정보는 학생 성과 분석을 통해 확인할 수 있으며, 이에 대한 분석은 ‘**성적 부여 및 종합**’ 섹션에서 더 상세히 다룹니다.

국제적 맥락에서 이루어지는 평가는 국가 시스템에서 일반적으로 발생하는 문제 외에도 형평성 측면에서 추가적인 어려움이 생기기도 합니다. 어느 한 국가에서는 매우 적절한 질문이 다른 국가에선 부적절한 질문이 될 수 있습니다. 예를 들어 스포츠, 여행, 오락, 역사적 사건, 심지어 날씨와 관련된 질문조차도 매우 신중하게 준비하고 취급해야 합니다. 이러한 문제를 피하는 유일한 방법은 모든 사회문화적 맥락을 배제한 시험 문제를 준비하는 것처럼 보일 수도 있습니다. 그러나 이렇게 하면 시험 문제의 범위를 제한할 뿐 아니라, 맥락에 기반한 과제를 통해 타당성을 확보하고자 하는 IB의 평가 철학과

일반적으로 좋은 평가 관행에도 어긋나게 됩니다. 맥락을 고려한 과제와 평가는 효과적인 학습을 위해 필수적입니다.

이러한 딜레마를 해결할 수 있는 접근 방법이 두 가지 있습니다. 첫 번째 방법은 과목 교수요목에 적절한 맥락 정보를 명시하거나, 문제의 기반이 되는 사례 연구를 제공하거나, 시험 내용 자체에 적절한 맥락 정보를 학생에게 제공하는 것입니다.

두 번째 방법은 학생이 자신의 맥락을 활용하여 답안을 구성할 수 있도록 하는 개방형 평가 질문과 과제를 더 많이 사용하는 것입니다. 특히 이 접근 방법을 사용할 때는 학생들이 활용하는 내용의 공통 기반이 없을 수 있으므로, 채점의 초점은 과목 내용에 대한 단순한 지식보다는 ‘심화 수준의 이해’에 두어야 합니다. 이는 IB 평가 철학에 매우 부합합니다.

[그림 4] 문화적 맥락의 범위



이 두 가지 방법을 모두 적용하더라도 학생들은 여전히 익숙하지 않은 사회문화적 맥락이 포함된 평가 과제를 마주할 수 있습니다. 이는 학생들이 **더 열린 마음**을 기르고, **세계적 인식**을 높이며, 익숙하지 않은 문화적 맥락에서도 보다 **편안**하고 **능숙**하게 다룰 수 있도록 장려하는 IB의 평가 철학에도 부합합니다.

상당 수의 IB 학생들이 자신의 모국어나 가장 능숙한 언어가 아닌 언어로 시험에 임합니다. 이러한 경우는 대부분 영어에 해당하는데, IB 평가가 자주 실시되는 또 다른 두 언어인 프랑스어와 스페인어로 시험을 치르는 학생들은 대체로 해당 언어에 능숙하기 때문입니다. 따라서 평가 문제를 출제하고 수정, 검토하는 과정에서 문항의 표현이 추가 언어로 학습하는 학생들에게 불리하지 않도록 각별한 주의를 기울입니다. 이는 시험 문항 검토 과정에서 우선적으로 살펴보는 중요한 요소입니다.

IB 평가에서 테크놀로지의 역할

- 학생들은 일상 생활, 학습, 에세이 작성 등 다양한 상황에서 테크놀로지를 폭넓게 활용하고 있습니다. 이들은 계속해서 향후 교육 과정과 직장에서도 광범위하게 테크놀로지를 활용하게 될 것입니다. 따라서 시험에서 테크놀로지를 사용하는 것은 학생들이 실제로 경험하는 세상을 진정성 있게 반영합니다.
- IB는 보다 양질의 유의미한 평가를 제공하기 위해 테크놀로지를 활용합니다. 다만 평가 방식의 선택은 테크놀로지의 사용 여부에 따라 결정되는 것은 아닙니다.
- 숙련된 채점관을 지원하기 위한 테크놀로지의 활용(e-채점)과, 학생을 위해 의미 있는 평가를 설계하기 위한 테크놀로지의 활용은 명확히 구분할 필요가 있습니다.
- 학생이 평가에 사용하는 테크놀로지는 반드시 교실에서 사용해 온 익숙한 것이어야 합니다.

디지털 평가

디지털 평가는 컴퓨터 기기를 사용하여 수행하는 평가를 의미합니다. 이는 평가 목적에 따라 다양한 형식의 문제와 학습 자극 자료를 포함할 수 있으며, 미디어 기반의 상호작용형 평가부터 전통적인 에세이형 시험까지 폭넓게 구성됩니다.

실질적으로 디지털 평가는 각 학생이 시험에서 개별 기기를 사용하여 수행합니다. 현재 MYP에서 사용하는 모델에서는 시험이 사전에 컴퓨터 기기에 설치되므로 인터넷 연결이 권장되지만 필수 사항은 아닙니다. 시험 소프트웨어는 실행 중 기기를 '잠금' 상태로 유지하며, 어떠한 이유로든 소프트웨어가 중단될 경우 이를 기록합니다(정당한 기술적 문제로 인한 중단의 경우에는 감독관이 소프트웨어를 재시작할 수 있음). MYP 디지털 시험의 경우 학생 답안을 일정 간격으로 온라인에 저장할 수 있고 학교의 행정적 부담을 줄일 수 있다는 점에서 온라인 응시를 강력히 권장합니다. 다만 대부분의 학교와 학생들이 MYP 온라인 기능을 활용하고 있음에도 온라인 방식은 시험 응시의 필수 조건은 아니며 IB는 오프라인 방식도 지원합니다. DP/CP 디지털 시험에서는 학생이 시험에만 접근할 수 있도록 제한하는 '잠금 모드'를 사용하므로 인터넷 연결이 필수입니다.

디지털 평가는 필기 시험의 물리적 한계를 넘어 IB가 진정으로 평가하고자 하는 내용을 평가한다는 점에서 큰 장점이 있습니다. 이러한 장점은 '디지털 평가가 타당성에 미치는 이점' 섹션에서 자세히 다루어지며 동영상 자료 제공, 에세이 작성 시 일반적인 워드 프로세싱 도구 사용 허용, 색상이나 글꼴 크기를 학생이 조절할 수 있도록 하는 접근성 지원 등이 포함됩니다.

디지털 평가는 앞으로도 계속 발전해 나갈 것입니다. 테크놀로지의 발전에 발맞춰 IB 평가도 함께 발전할 수 있을 것입니다. 현재 IB는 학생들이 키보드, 마우스 또는 터치 패드를 사용하는 것만 고려하고 있지만, 향후 몇 년 내에 터치스크린 기술이 일반화되면 더 많은 가능성이 열릴 것입니다. 더 나아가 먼 미래에는 가상현실이 평가에서 어떤 역할을 하게 될지도 상상해 볼 수 있습니다. 예를 들어 외국어 평가는 가상현실 환경에서 더욱 실제적인 방식으로 구현될 수 있으며 공상과학 영화에서처럼 가상 실험실을 통해 기술적 과제(과학 문제) 해결도 가능해질 수 있습니다.

IB가 프로그램 전반에서 디지털 전환을 지속적으로 추진함에 따라, 평가 전반에 영향을 미치는 새로운 용어들도 필연적으로 등장하게 될 것입니다. 디지털 평가 관련 용어를 포함한 IB 평가 용어에 관한 해설은 본 간행물의 마지막 부분에서 확인할 수 있습니다.

테크놀로지를 활용한 교수

교실에서의 대화형 화이트보드 활용부터 대규모 온라인 공개 강좌를 통해 더 많은 대상에게 교육을 제공하는 등 테크놀로지는 학생들의 학습 방식을 변화시키고 있습니다. 대부분의 학생은 에세이를 수기로 작성하기보다는 컴퓨터 기기를 사용해 작성합니다.

이처럼 필기 중심의 학습에서 기기 활용으로의 전환은 평가와 교수법 간의 괴리가 생길 수 있습니다. 많은 학생들에게 2~3시간 동안 펜으로 글을 쓰는 경험은 시험을 보는 경우에만 일어나며, 이러한 특징이 평가의 진정성을 크게 약화시킵니다.

디지털 평가의 활용이 확대됨에 따라 인터페이스 사용이 과목에 대한 이해를 보여주는 데 장애물이 되지 않도록 학생들이 평가 형식에 익숙해지는 것이 중요합니다. 이를 위해 숙지용 도구나 모의 시험을 활용할 수 있으나, 가장 이상적인 해결책은 수업 중에도 평가에서 사용하는 것과 동일한 도구를 활용하는 것입니다.

좋은 디지털 평가는 어떤 모습인가?

디지털 평가는 전통적인 서면 기반 시험에서는 구현할 수 없었던 질문을 제시할 수 있는 수단으로 테크놀로지를 활용함으로써 평가의 수준을 향상시킵니다. IB에서 좋은 디지털 평가는 다음과 같은 특징을 갖습니다.

- 테크놀로지를 활용하여 더 타당하고, 실질적이며 적합한 과제를 제시합니다.
- 행정적 부담을 줄이고 운영 효율성을 높입니다.
- 평가 과정에서 새로운 장애 요인을 만들어내지 않습니다.

e-채점

‘e-채점’이란 IB가 평가를 위해 제출된 과제를 채점할 때 컴퓨터 기기를 사용하여 답안지를 화면으로 확인하고 점수를 기록하는 방식을 의미합니다. 접근성 및 포용성 조치와 관련된 예외적인 상황을 제외하고, IB는 더 이상 시험지나 내부 평가 과제의 사본 용지를 채점관에게 발송하지 않습니다. 서면으로 작성된 과제는 전자 형식으로 스캔되며, 채점관은 전용 채점 소프트웨어를 통해 인터넷으로 해당 과제에 접근합니다.

e-채점은 채점의 속도와 품질 측면에서 여러 가지 중요한 이점을 제공합니다.

- IB는 평가를 위해 제출된 모든 과제를 항상 보관할 수 있으며, 필요 시 두 번째 채점관이 즉시 접근할 수 있어 물리적 발송이 불필요합니다.
- IB는 과제를 익명화하여 채점 편향 가능성을 줄일 수 있습니다.
- 채점이 진행되는 동안 채점 기준을 점검할 수 있어, 보다 엄정한 품질 관리가 가능합니다.
- 한 학교의 학생들이 제출한 과제가 한 명의 채점관에게 집중되지 않고, 여러 채점관에게 무작위로 배분됩니다. 이를 통해 IB는 보다 강화된 품질 검증 절차를 수행할 수 있습니다.

IB는 2010년부터 e-채점을 사용해 왔으며, 이는 학생에게 디지털 형식의 과제 제출을 요구하는 것을 의미하지는 않습니다. 현재 대부분의 시험은 여전히 수기로 작성되며, 이후 디지털 형식으로 스캔되어 채점이 이루어집니다.

디지털 평가와 관련된 위험 요소

디지털 평가를 개발하는 과정에서 IB는 학생과 학교의 공정성을 보장하기 위해 반드시 극복해야 할 여러 난제를 인지하고 있습니다. 다음 섹션에서는 IB가 해결해 나가고 있는 주요 위험 요소 가운데 일부만을 중점적으로 다룹니다. 보다 상세한 논의는 본 간행물의 개별 섹션 및 디지털 평가에 초점을 둔 다른 IB 간행물에서 확인할 수 있습니다.

학교의 부담

IB는 오늘날 테크놀로지의 사용이 학생들의 일상에 자연스럽게 통합되어 있으며, 이는 학생들의 평가 경험에도 반영해야 합니다. 그러나 학교가 기존의 절차를 전환하는 데에는 시간이 필요하며, 디지털 시험은 서면 기반 시험과는 다른 형태의 준비가 필요하다는 점 또한 인지하고 있습니다. 이에 따라 IB는 전 세계 어느 지역의 학교이든 과도한 부담을 지우지 않으면서도 기술 발전의 속도에 맞춰 나갈 수 있도록, DP/CP 디지털 시험의 도입을 신중하게 관리할 것입니다.

디지털 평가로의 점진적인 전환 지원하기

IB는 IB 공동체와 협력하여 테크놀로지를 활용해 평가의 질을 향상시키고자 하는 모든 학교가 이를 실현할 수 있도록 충분한 전환 기간을 제공할 것입니다. IB는 테크놀로지의 잠재력이 학습과 교수의 질을 높이는 데 현명하게 활용될 수 있으며, 그 결과 평가 역시 개선될 수 있다고 믿습니다.

테크놀로지 결함

시험 중 어떠한 기술적 결함도 발생해서는 안 됩니다. 모든 학생은 원활하고 방해받지 않는 시험 환경에서 시험을 치를 수 있어야 합니다. 이를 위해 IB는 시험 이전에 학교가 활용할 수 있는 충분한 사전 준비, 시험 및 규정 준수 절차를 마련해 두고 있습니다. IB의 경험에 따르면 대부분의 문제는 학교의 시설에서 발생하며, 이러한 문제는 사전 숙지, 시험, 규정 준수 과정을 통해 시험 이전에 충분히 파악하고 해결할 수 있습니다.

보안

디지털 시험은 최소한 서면 기반 시험과 동등한 수준의 보안을 갖추어야 합니다. IB 평가는 시험이 진행되는 동안 관련 기기의 모든 기능이 ‘차단’되도록 설계되었습니다. MYP 컴퓨터 기반 평가와 마찬가지로 시험 전에 평가 자료가 학교로 발송될 경우, 해당 파일은 암호화되고 비밀번호가 설정됩니다.

IB는 산업 전반에서 더욱 정교한 보안 기술이 발전함에 따라 보안 체계를 지속적으로 강화해 나갈 예정이며, 동시에 서면 기반 시험 역시 그 자체의 보안 위험에 노출되어 있음을 인지하고 있습니다.

테크놀로지를 위한 테크놀로지

IB는 단순히 테크놀로지를 사용하기 위해서가 아니라 평가할 수 있는 내용의 질을 향상시키기 위해 테크놀로지를 활용하는 데 전념하고 있습니다. IB는 테크놀로지의 활용을 통해 평가의 타당성이 어떻게 향상되어 왔는지를 지속적으로 알려 나갈 것이며, 이러한 관점은 평가 개발 과정에 반영될 것입니다.

편향

IB는 일부 학생이 컴퓨터 기기 사용으로 인해 더 나은 성과를 낼 수 있다는 점을 인지하고 있습니다. 그러나 서면 기반 시험 역시 편향에서 자유롭지 않으며, 장시간 필기가 어렵거나 필기 속도가 느린 학생들은 이미 불리한 조건을 가지고 시험을 치르게 됩니다.

IB는 원칙적으로 학생의 필기 능력이나 타이핑 속도를 측정하려는 것이 아니며, 타이핑이나 필기가 빠른 학생이 유리해지지 않도록 평가 과제를 설계할 것입니다.

IB는 이러한 원칙을 실현하기 위해 ‘기기 효과’에 관한 연구 결과를 면밀히 검토하고 있습니다.

기준 변경하기—상향 또는 하향

IB 성적은 고유한 의미를 지니며, IB 선임 채점관의 전문적 판단과 결과 데이터를 통해 이러한 의미가 퇴색되지 않도록 합니다. 예를 들어 학생이 단락을 잘라내거나 재구성할 수 있는 환경에서 에세이를 작성할 경우, 보다 구조화된 글을 상대적으로 수월하게 작성할 수 있습니다. IB는 이러한 점을 성적 급간 설정 시 반드시 고려하여, 서면 기반 시험과 디지털 시험 모두 동일하게 익숙한 학생이 과목에 대한 이해도와 분석 능력을 똑같이 반영하는 결과를 얻게 할 것입니다.

디지털 평가를 통해 IB가 항상 중시했지만 이전에는 평가할 수 없었던 새로운 자질들을 테스트할 수 있게 됨에 따라, 평가에서 학생에게 기대하는 수준을 조정해야 할 수 있습니다(자세한 내용은 ‘기준’ 섹션 참조). 그러나 이러한 조정은 IB의 장기적 목표를 유지하고 현재의 성적 기준표에 명시된 자질을 반영하는 방식으로만 이를 수행할 것입니다.

추가 문헌

IB의 디지털 평가에 관한 더 자세한 내용은 다음 자료를 참고하시기 바랍니다.

- 프로그램 자료실의 “[Transition to Digital Examinations](#)”(영문 제공) 페이지
- “[MYP eAssessment playbook](#)”(영문 제공)
- “[MYP eAssessment: Introduction for school leaders and teachers](#)”(영문 제공, 동영상)
- “[MYP eAssessment Ready! Introduction](#)”(영문 제공)

교육에서의 평가

- 교육에서 평가는 주로 형성 평가(학습을 위한 평가) 또는 총괄 평가(학습의 평가)로 분류됩니다. 오늘날에는, 학습으로서의 평가로 나아가려는 활발한 움직임이 있습니다.
- 어떤 맥락에서든 사용되는 평가 유형은 결과 활용 방식이나 평가 목적에 따라 달라져야 합니다.
- 평가는 교수법에 영향을 줄 수 있으며, 평가가 학습 및 교수에 미치는 영향, 즉 역류 효과가 긍정적이 되도록 설계되어야 합니다.

평가 정의하기

「IB 교육이란 무엇인가?」(2019)에 명시되어 있듯이 IB는 학생에게 적극적인 역할을 부여하고, 효과적인 학습에서 맥락의 중요성을 인식하는 구성주의적 학습법을 따릅니다(Murphy, 1999). 평가가 효과적인 학습 및 교수를 지원하려면, 구성주의적 학습 이론을 기반으로 평가를 설계해야 합니다. 이 개념과 관련된 더 많은 연구 자료는 Shepard(1992), Wood(1998), Black(1999), Lambert and Lines(2000)를 참고하십시오.

‘평가’란 학생의 학습 성과를 수집하고 판단하는 모든 방법을 의미할 수 있습니다. 일반적인 평가 유형으로는 테스트, 시험, 수행평가, 프로젝트, 포트폴리오, 구술 평가 등이 있습니다. 평가는 때로는 장기간에 걸쳐 진행되기도 하고, 때로는 몇 시간 내에 완료되기도 합니다. 평가는 유형과 목적에 따라 교사, 동료 학생, 학생 본인 또는 채점관에 의해 판단될 수 있습니다.

[그림 5] 여러 가지 유형의 평가 예시



학생이 얼마나 우수한지 또는 능숙한지를 평가하는 과제에서는 **보상 모델**이나 **숙달 모델**이 사용됩니다 ([그림 6] 참고).

보상 모델은 대부분 외부 시험에서 사용되며, 한 영역의 강점이 다른 영역의 약점을 상쇄할 수 있게 합니다. 반면, **숙달 모델**은 평가의 각 부분에서 최소한의 성취(숙달)를 요구합니다.

예를 들어, 시험이 세 개의 문항으로 구성되어 있고 각 문항이 10점 만점이라고 가정해 보겠습니다. 합격 점수가 15점이라면, 학생은 각 문항에서 5점씩 받거나, 10점, 3점, 2점을 받아도 합격할 수 있습니다. 후자의 경우 첫 번째 문항에서의 만점이 나머지 문항에서의 낮은 점수를 상쇄합니다.

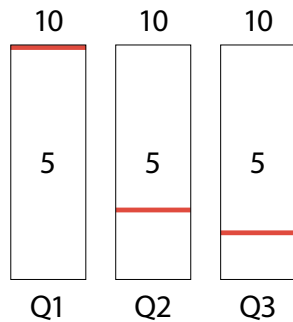
그러나 동일한 시험(각 문항당 10점 만점인 세 개의 문항으로 구성된 시험)에서 각 문항마다 합격 점수 5점이 요구된다면, 첫 번째 학생은 합격하지만 두 번째 학생은 합격하지 못합니다. 또한 10점, 10점, 4점을 받은 학생도 시험을 합격하지 못합니다.

[그림 6] 보상 모델과 숙달 모델의 비교

보상 모델

문제 수: 3
 문제별 점수: 10
 최소 합격 점수(총점): 15/30

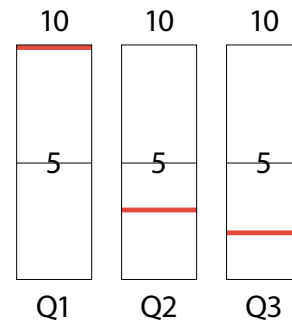
보상 모델을 적용하는 경우:
 모든 문제의 총점이 최소 합격 점수와 같거나 그 이상이면 평가를 **합격**으로 처리합니다.



숙달 모델

문제 수: 3
 문제별 점수: 10
 최소 합격 점수(총점): 5/10

숙달 모델을 적용하는 경우:
 문제 2와 3에서 점수가 부족하여 **불합격**으로 처리합니다.



숙달 모델 평가는 한 요소에서는 매우 뛰어나지만 다른 요소에서는 부족한 수행이 허용되기 어려운 직업 또는 직장 환경에서 주로 활용됩니다. 예를 들어 옷을 만들 때 기본적인 바느질을 할 줄 모른다면 아무리 뛰어난 디자인 전문성을 갖추고 있더라도 이를 상쇄할 수 없습니다.

IB는 다양한 평가 도구를 활용합니다. 여기에는 프로그램 종료 시 시행되는 시험뿐만 아니라, 다양한 과목에 걸쳐 학생이 과정 중 여러 시기와 조건에서 수행하는 다양한 기타 평가 과제(연구 에세이, 서면 과제, 구술 인터뷰, 과학적 및 수학적 조사, 현장 연구, 예술 공연)가 모두 포함됩니다.

평가 접근 방법

평가는 다양한 목적을 위해 활용될 수 있습니다. 특정 평가의 의도된 목적은 평가를 설계하는 방식에 상당한 영향을 줍니다. 일반적으로 평가는 형성 평가와 총괄 평가라는 두 가지 유형으로 구분됩니다.

형성 평가는 학생의 학습 방식과 가장 직접적으로 연계되어 있으며 ‘학습을 위한 평가’로 불립니다. 반면 총괄 평가는 ‘학습의 평가’로 지칭됩니다. 이러한 구분은 총괄 평가가 실제 교실 학습에 얼마나 큰 영향을 미치는지를 간과하게 됩니다. 모든 평가는 적절한 학습을 지원해야 합니다. 총괄 평가는 학습이 끝난 뒤에만 수행되는 활동이 아니라, 학습 및 교수 전반에 통합적으로 기여하도록 설계되어야 합니다.

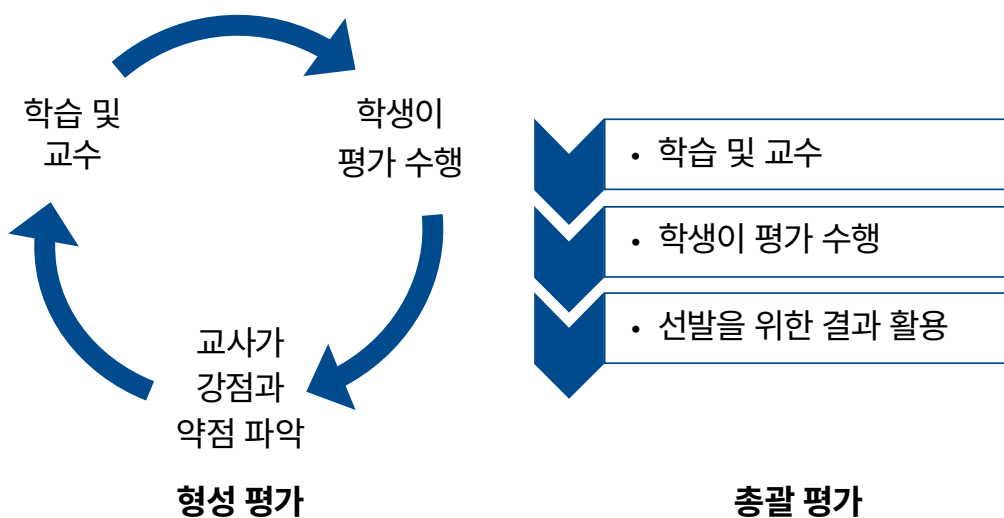
형성 평가의 목적은 학생의 강점과 약점의 특성에 대해 교사와 학생에게 상세한 피드백을 제공하고, 학생의 역량 발달을 지원하는 데 있습니다. 이 과정에서는 교사와 학생 간의 토론과 같은 직접적인 상호작용 형태의 평가 유형이 특히 효과적입니다.

Vygotsky(1962)는 교사를 학습의 지도자가 아닌 지원자로 보았으며, 그가 말한 ‘근접 발달 영역(zone of proximal development)’에서 학생이 활동하는 데 도움이 되는 평가 과제와 도구를 사용해야 한다고 주장했습니다. 근접 발달 영역은 학생이 혼자 할 수 있는 수준과 교사의 도움을 받아 할 수 있는 수준 사이의 성취 범위를 나타냅니다.

‘비계 설정(scaffolding)’이라는 개념은 우드(Wood)와 그 외 연구자들에 의해 제시되었습니다(1976). 이 프레임워크에서 교사는 학습 형성을 위해 비계를 제공하지만, 실제 학습을 구성하는 주체는 학생입니다. 따라서 교사는 학생에게 적절한 수준의 도전 과제인 형성 평가를 설정하고, 학생의 성취도 수준에 따라 이를 지속적으로 조정해야 합니다.

반면 총괄 평가는 학생이 무엇을 할 수 있는지를 측정하는 데 초점을 두며, 일반적으로 교육 프로그램의 이수, 또는 다음 교육 단계로 진행할 준비가 되었는지를 확인하는 데 사용됩니다. 형성 평가가 학생이 어떤 행동을 왜 하는지에 관심을 둔다면, 총괄 평가는 학생이 그것을 올바르게 수행했는지를 확인하고자 합니다. 총괄 평가가 학습 원인을 파악하는 형성 평가보다 덜 유용해 보일 수 있지만, 총괄 평가의 목적은 향후 교수 방향을 제시하는 것이 아니라 학생에 대해 판단을 내리는 것입니다. 총괄 평가의 필요성에 대해서는 ‘총괄 평가에서 학생의 성과 설명하기’ 섹션에서 더 자세히 알아볼 수 있습니다.

[그림 7] 형성 평가 및 총괄 평가의 활용 가능성



형성 평가에서는 각 학생의 성취 수준을 정확히 측정하는 것보다 학생이 아직 충분히 발달시키지 못한 지식, 기능 및 이해를 올바르게 파악하는 것이 더 중요합니다. 이러한 우선순위 간의 균형을 ‘타당성’이라 하며, 이후 섹션에서 더욱 상세하게 다룹니다. 총괄 평가에서는 학생의 성취와 피드백 수준 간의 균형이 서로 정반대가 되어 평가 결과는 학생에 대한 판단을 내리는 데 활용됩니다. 주로 취업이나 교육 기회에서의 경쟁적 선발 과정에 사용되며 향후 교수 활동을 지원하는 데도 활용될 수 있습니다.

서로 다른 국가의 평가 시스템을 분석해 보면 다양한 평가 기법과 접근 방법을 빠르게 확인할 수 있습니다. 이러한 평가 시스템은 모두 기술적 여건, 자원과 시간적 제약, 그리고 국가의 교육 시스템에 미치는 영향과 관련하여 각기 다른 장단점이 있습니다. 어느 맥락에서 완전히 새로운 평가 시스템을 고안할 수 있다 해도, 보편적으로 적용 가능한 최선의 기술적 방안은 존재하지 않습니다. 대신 평가 시스템을 설계하며 내리는 선택들은 사회적 맥락의 가치와 우선순위를 필연적으로 반영합니다. 이 분야에 대한 더 많은 연구는 Cresswell(1996)과 Broadfoot(1996)을 참고하십시오.

또한 총괄 평가가 교수의 질을 측정하는 척도로 점차 활용되고 있다는 점을 인식하는 것이 중요합니다. 이는 평가가 학생이 아닌 교육 시스템에 도움이 되기 위해 수행된다는 또 다른 차원을 드러냅니다.

역류 효과 및 학습

필요한 것은 가능한 한 타당한 평가 과정이다... 그러나 이러한 과정이 역류 효과로 인해 좋은 교수법을 왜곡하거나, 너무 느리거나, 부족한 교육 자원을 지나치게 소모해서는 안 된다.

(Peterson, 1971)

이 인용구에서 IB의 설립자인 알렉 피터슨(Alec Peterson)은 평가가 수행되는 방식이 학습 접근 방법에 영향을 미칠 수 있다는 위험을 인식합니다. Surgenor(2010)가 언급했듯이 이러한 영향은 학생의 학습 및 교수 경험에 부정적인 영향과 긍정적인 영향을 모두 미칠 수 있습니다.

1971년 논문에서 스나이더(Snyder)는 학생들이 평가에서 중요하게 여기는 암시적, 명시적 메시지를 바탕으로 자신들만의 교육과정에 대한 이해를 형성한다고 주장하였으며, 이를 ‘감춰진 교육과정(hidden curriculum)’이라고 설명했습니다. 이는 학생들이 과목을 합격하는 방법은 알지만 과목 자체에 대한 이해는 부족한 상황으로 이어질 수 있습니다. 이를 설명하는 또 다른 방식은 다음과 같은 깁스(Gibbs)의 인용문에서 확인할 수 있습니다.

나는 시험에 재응시할 때 시험을 합격하는 것에만 집중했다. 나는 96점을 받았지만, 내가 처음 본 시험에서 왜 떨어졌는지 그는 이해하지 못했다. 나는 그에게 이번에는 과목을 이해하기보다는 시험을 통과하는 데 집중했다고 말했다. 여전히 나는 해당 과목을 이해하지 못했기 때문에, 어떤 면에서는 목표를 달성하지 못한 셈이다.

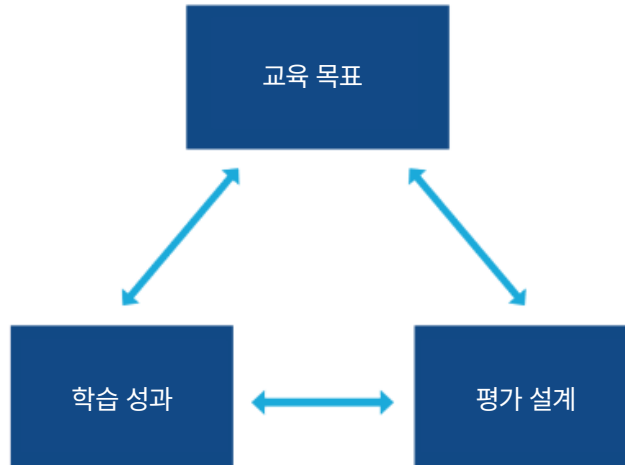
(Gibbs, 1992, p. 101)

역류 효과의 개념은 ‘시험에 나오지 않는 것은 가르치지 않는다’는 격언에서도 잘 드러나며, 이는 평가와 교수가 서로 독립적으로 존재할 수 없음을 강조합니다.

학생이 학습을 입증하도록 장려하려는 교육 목표를 평가가 제대로 뒷받침하지 못하는 데에는 여러 가지 이유가 있습니다. 가장 흔한 이유는 평가의 기반이 되는 학습 성과가 교육의 본래 목적을 효과적으로 반영하지 못하는 경우입니다. 예를 들어, 엘리트 운동선수를 훈련하는 직업 과정에서 학습 성과가 경기 규칙의 이해에만 초점을 맞추는 경우를 들 수 있습니다. 또 다른 문제는 평가가 모든 성취 목표를 다루고 있음에도 불구하고, 실제 평가의 상당 부분이 측정이 용이하다는 이유로 교육적 중요도가 낮은 과제들에 집중되는 경우입니다.

이러한 평가와 교육 목표 간의 상호 의존성은 [그림 8]에 나타나 있으며, 이는(Furst's paradigm, 1958; Frith & Macintosh, 1984)에서 발췌하였습니다. [그림 8]에서 제시된 세 가지의 요소 중 어느 두 요소라도 연결이 끊어지면 평가의 질은 필연적으로 저하됩니다.

[그림 8] 목표, 결과, 설계의 관계



타당성

- ‘타당성(validity)’이란 평가가 **목적에 부합함**을 의미하며, 다양한 측면을 가진 복합적인 개념입니다.
- 평가는 다양한 목적을 위해 활용됩니다.
- 하나의 목적에서 타당한 평가가 다른 목적에는 타당하지 않을 수도 있습니다. 예를 들어 철자 시험은 언어의 유창성을 측정하지 않습니다.
- 이러한 점에서 볼 때, 평가 결과의 타당성은 평가 자체가 아니라 결과 활용 목적에 달려 있다고 주장할 수 있습니다.
- IB는 가장 먼저 프로그램 전체가 타당한지, 프로그램 요소(개별 과정 등)가 타당한지, 개별 평가가 타당한지 순으로 고려합니다.

타당성 정의하기

시험을 치르는 목적은 무엇입니까? 시험은 무엇을 위한 것입니까? 이러한 질문들은 평가가 ‘목적에 부합하는지’ 또는 ‘올바른 것을 측정하는지’를 판단하는 핵심 요소이며, 이러한 질문들에 대한 답변은 종종 서로 상충합니다. 또한 교육의 모든 목적이 하나의 시험으로 평가할 수는 없으며, 일부 목적은 평가 자체가 불가능할 수도 있습니다.

예를 들어 수학 평가의 목적이 다음 네 가지라고 가정해 보겠습니다.

1. 학습 과정이 종료된 후 학생의 이해 수준 파악
2. 추가 학습 또는 직업 선택을 위한 수단으로 활용
3. 학생의 향후 성공 가능성 예측
4. 프로그램의 교육 목표에 대한 교수 강화

이처럼 간단한 목표조차도 서로 조화시키기 어렵습니다. 만약 수학에서의 향후 성공이 기하학보다 미적분에 더 크게 좌우된다면, 평가는 미적분에 더 초점을 두어야 합니까? 그렇다면 이는 기하학에 대한 학생의 이해를 파악하려는 첫 번째 목적과 어떻게 연결되니까? 또한 이러한 목표 중 두 가지 이상이 평가 설계 방식에서 상충할 경우 어떻게 됩니까?

여기 제시된 네 가지 목적은 결코 확정적이거나 완전하지 않습니다. Newton(2007)은 평가 결과의 가능한 활용 방식에 대한 여러 예시를 제시합니다.

<ul style="list-style-type: none"> • 사회적 평가를 위한 활용 • 형성적 활용 • 학생을 관찰하기 위한 활용 • 이전을 위한 활용 • 배치를 위한 활용 • 진단을 위한 활용 • 지도를 위한 활용 • 자격 부여를 위한 활용 • 선별을 위한 활용 	<ul style="list-style-type: none"> • 인가를 위한 활용 • 학교 선택을 위한 활용 • 기관의 관찰을 위한 활용 • 자원 분배를 위한 활용 • 조직의 개입을 위한 활용 • 프로그램 평가를 위한 활용 • 시스템 모니터링을 위한 활용 • 비교 가능성을 위한 활용
--	---

이처럼 평가 결과의 활용 목적이 서로 상충하는 경우가 존재하지만, 그럼에도 불구하고 개인과 기관은 다양한 맥락에서 사람에 대해 의사 결정을 내려야 합니다. 만약 평가 결과가 제공되지 않는다면, 이러한 결정은 제대로 설계되지 않았거나 편향될 가능성이 더 큰 다른 기준에 의존하게 될 가능성이 높습니다. 이 점을 강조하며 Cresswell(1986)은 이 문제를 다음과 같이 설명했습니다.

분명하게도 다른 기준들이 시험보다 신뢰도가 낮다면, 그러한 기준에 더 크게 의존할수록 선발 결정의 신뢰성은 더욱 낮아질 것이다.

(Cresswell, 1986, p. 42)

이처럼 ‘의도한 목적에 부합하는’ 자격(또는 개별 평가와 같이 자격의 일부)이란 개념은 일반적으로 타당성을 의미합니다.

타당성은 평가의 질과 관련하여 필수적인 용어로 자주 사용되지만, 그 의미를 정확히 정의하기 위한 학문적 논의는 오랜 기간 지속되어 왔습니다(Newton, 2012). 더 복잡하게 만드는 것은, **신뢰성**과 함께 주로 사용되는 보다 좁은 의미의 타당성 개념이 존재한다는 점입니다. 이는 평가가 측정하기로 명시한 것을 실제로 얼마나 정확히 측정하는지의 정도를 의미합니다. 본 간행물에서는 혼동을 피하기 위해 두 번째 개념을 ‘구성개념 적합성’이라 부릅니다. 자세한 내용은 ‘구성개념 적합성 및 진위성’ 섹션을 참고하십시오.

평가의 타당성을 결정하는 것은 궁극적으로 판단의 문제입니다. 이러한 판단은 시험 점수를 어떻게 해석하고 어떤 목적으로 활용하는지를 뒷받침하기 위해 체계적으로 수집된 근거에 기반합니다. 본 섹션 앞부분에서 확인된 평가의 상충하는 요구들은 이후 섹션에서 보다 자세히 다루어지겠지만 이러한 요구들 사이에는 언제나 절충이 필요합니다. 중요한 질문은 제공된 증거들이 평가가 실제로 유용할 만큼 목적에 충분히 부합하는지를 뒷받침하느냐입니다. 따라서 타당성에 대해 논할 때는 평가가 타당하다 또는 타당하지 않다는 단순한 선언보다는, 근거나 ‘타당성 주장’이 얼마나 설득력 있는지를 중심으로 논하는 것이 바람직합니다. 결국 어떤 평가가 타당한지는 근거에 기반한 판단의 결과입니다. 타당성은 이분법적 개념과는 다르게 맥락적이고 목적 지향적입니다. 뉴턴(Newton)과 쇼(Shaw)는 우리가 타당성을 주장하기 전에 평가의 목적에 얼마나 부합하는지 고려해야 하며, ‘타당성 확인은 시험 점수의 해석과 활용에 관한 주장의 강도를 수립하고 평가하는 과정’이라고 설명합니다(Newton and Shaw, 2014, p. 3). 따라서 타당성은 평가 자체의 속성이 아니라 평가 결과로부터 도출되는 추론의 속성입니다.

마지막으로 타당성은 평가 설계 과정에서 한 번에 달성되는 것이 아니며, 오히려 평가 주기 동안 지속적으로 발전합니다. 이와 마찬가지로, 타당성 주장은 평가 주기 초기 단계에서 만들어지는 것이 아니라, 설계 과정에서 계속해서 보강되고 정교해집니다.

타당성 주장 만들기

타당성은 단순한 객관적 개념이 아닌 서로 상충하는 요인들 간의 균형을 의미하며, 어떤 방식으로든 ‘증명’할 수 있는 것이 아닙니다. 대신 특정 과정이나 평가에 대해 내려진 결정이 왜 의미 있고 의도된 목적에 부합하는지에 대한 설득력 있는 주장을 구성할 수 있을 뿐입니다. 결과적으로 타당한 것은 평가 (또는 결과) 그 자체가 아니라, 평가가 활용되는 목적입니다. Newton(2012)은 이를 ‘특정 목적(즉, 특정 결정을 내리기 위한 목적)을 위해 특정 평가 절차를 사용하는 것이 타당하려면, 그에 대한 해석적 주장이 충분히 강력해야 한다’고 성찰했습니다.

이와 비슷한 이유로 타당성 주장을 뒷받침하는 근거는 프로그램을 개발, 모니터링, 전달하는 과정에서 자연스럽게 도출되어야 하며, 그 과정에서 이루어진 논의와 결정 사항을 반영해야 합니다. 평가에서의 한 예로, 특정 시험에서 어떤 문제를 출제할지에 대한 판단은 해당 교육과정이 적절히 반영되었는지를 보여주는 근거가 됩니다.

모든 타당성 주장에서 필수 요소는 중요하다고 생각하는 모든 측면을 포괄하는 적절한 구조를 갖추는 것입니다. IB에서 이는 증거 수집을 위한 일련의 질문 형태로 나타납니다.

타당성 유지하기

타당성은 평가 모델이 설계될 때 결정되거나 ‘증명’되는 것이 아닙니다. 오히려 평가 주기 동안, 그리고 실제로 평가 결과에 기반하여 결정이 내려지는 한 계속해서 발전합니다.

IB에서는 특정 평가가 타당함을 입증하기 위한 다양한 근거(즉, 서로 다른 타당성 주장)가 특정 과정의 전반에 걸쳐 여러 차례 수집됩니다.

과정이 개발되거나 검토될 때에는 과정의 목적과 그 목적을 달성하기 위해 어떻게 평가되어야 할지에 대한 논의가 타당성 주장의 핵심을 이룹니다. 특히 구성개념 적합성과 타당성의 다른 측면들 간의 균형이 고려됩니다. 매 기수마다 새로운 시험이나 평가 과제가 출제되는 과정에서 특히 신뢰성, 공정성, 비교 가능성, 관리 가능성, 구성개념 적합성과 관련된 추가 근거들이 생성될 것입니다. 이와 같은 방식으로 교사의 피드백과 학생의 성취도 결과는 해당 평가가 목표를 얼마나 성공적으로 달성했는지에 대한 지표이자, 채점의 신뢰성과 성적의 비교 가능성을 입증하는 근거로도 활용됩니다. 이러한 정보들은 다시 다음 평가를 개발하는 데 활용됩니다.

마지막으로 특정 과정에 포함된 모든 평가와 특정 프로그램에 속한 모든 학습 과정에서 수집된 근거들은 다음 과정과 프로그램을 검토할 때 의사결정의 기반이 됩니다.

시험의 타당성을 평가하는 것은 한 번으로 끝나는 정적인 활동이 아니라, 지속적으로 이루어지는 과정이다.

(Sireci, 2007, p. 477)

타당성 사슬의 요소

- 타당성을 이루는 요소들은 매우 다양합니다. 이 요소들은 하나의 사슬과 같아서, 한 고리라도 끊어지면 전체 평가가 무너지게 됩니다.
- IB는 일반적으로 타당성의 다섯 가지 요소인 신뢰성, 구성개념 적합성, 진정성, 관리 가능성, 공정성 및 편향, 비교 가능성에 초점을 맞춥니다.
- 이 다섯 가지 요소들은 종종 서로 긴장 관계에 있습니다. 평가의 타당성을 판단할 때는 평가의 주된 목적을 고려하여 각 요소의 상대적 중요성을 결정해야 합니다.
- IB는 구성개념 적합성을 반영하는 프로그램, 과정 및 평가를 만드는 것을 가장 중요하게 생각합니다.
- 이는 평가에서 IB가 신뢰성 있게 쉽게 채점할 수 있는 시험보다는 고차원적 사고 기능을 테스트할 수 있는 의미 있는 과제와 질문을 제공하는 데 초점을 두는 것을 의미합니다.

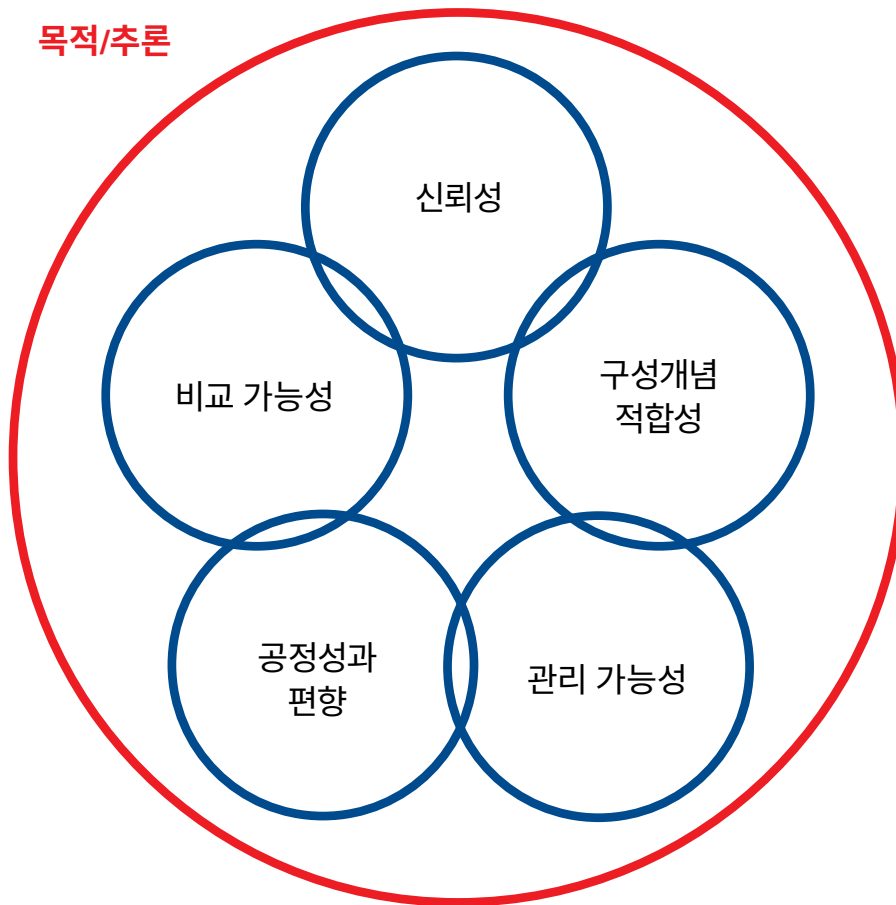
타당성과 신뢰성은 모든 평가 체계에서 본질적인 특징으로 간주되며, 특히 결과가 학생이나 교사에게 매우 중요하고 높은 부담감을 수반하는 평가에서 더욱 그렇습니다. 이러한 특징들은 실제로 여러 유형의 타당성과 신뢰성으로 구성된 다면적 개념입니다.

타당성의 복합적인 특징은 **타당성 사슬**이라는 아이디어를 통해 표현될 수 있습니다(Crooks et al., 1996). 타당성 사슬의 다섯 가지 요소는 개별적으로도 모두 중요하지만 그 자체로는 평가를 타당하게(즉, 목적에 적합하게) 만들기에 충분하지 않습니다. 예를 들어, 어떤 평가는 매우 높은 신뢰성을 가질 수 있으나, 특정 집단에게 구조적인 불이익을 줄 수도 있습니다. 또는 과제가 의도한 것에 정확히 초점을 맞추고 있다 하더라도, 평가의 소요 시간이나 요구 조건 때문에 학생의 체력과 학교의 자원 활용 능력까지 함께 측정하게 되는 상황이 발생할 수 있습니다. 다음 섹션에서는 각 요소들을 더욱 상세히 다룰 예정입니다. 이 요소들은 타당성을 확보하는 데 모두 필수적이지만, 동시에 각 요소 간에는 상충하는 지점도 존재합니다. 예를 들어, 시험이 교육과정의 모든 측면을 다루는 것을 목표로 할 수 있지만, 이는 학생들에게 너무 긴 시험으로 이어질 수 있습니다. 마찬가지로, 시험이 치러지는 특정 문화나 국가에 맞게 맥락화(공정성)를 추구하면, 실제로 동일한 시험이 각 지역에서 시행되었는지에 대한 의문(비교 가능성)이 생길 수 있습니다.

평가가 타당하려면 사슬의 모든 고리가 반드시 존재해야 합니다.

[그림 9] 타당성 사슬

목적/추론



타당성의 여러 측면 간의 균형 맞추기

평가를 설계할 때는 평가의 목적이라는 맥락에서 서로 상충하는 타당성 측면 간의 균형을 맞추는 것이 중요합니다. 하나의 평가가 이러한 모든 요소를 최고 높은 수준으로 달성하는 것은 불가능하기 때문에 절충안을 찾아야 합니다. 마찬가지로 이러한 요소 중 일부는 평가 설계 과정에서 이미 결정되지만 특히 신뢰성과 공정성과 같은 다른 요소들은 평가 시행, 채점, 성적 부여 과정에서 계속 발전합니다.

[그림 10] 타당성 측면 간 상충하는 우선순위의 균형: 공정성과 관리 가능성

		
<p>균형이 적절하지 않음 응시생에게 불합리한 부담을 줌 (관리 가능성 결여)</p>	<p>균형이 적절함 응시생이 합리적인 방식으로 평가됨</p>	<p>균형이 적절하지 않음 출제 주제에 운이 과도하게 작용함(공정성 결여)</p>
<p>평가에는 과정의 모든 측면을 다루는 50개의 확장형 과제 (각 20분)가 포함됨. 이는 16시간 이상 평가를 수행해야 함을 의미함.</p>	<p>평가에는 20개의 단답형 문제(각 2분), 5개의 비교적 긴 과제(각 10분), 2개의 심층 과제(각 30분)가 포함됨. 각 문제는 과정의 서로 다른 측면을 다룸. 평가 전체는 2시간 반 동안 진행되며 과정 주제의 절반을 약간 넘게 다룸.</p>	<p>평가에는 과정의 한 측면에 대한 지식과 이해를 평가하는 1개의 30분짜리 심층 과제가 포함됨 (총 50개 중). 응시생의 최종 성적은 단 하나의 과제에 전적으로 기반함.</p>

이 다섯 가지 요소들은 서로 다른 정의를 갖고 있지만 작용하고 관리되는 방식에서 상당 부분 중첩됩니다. 이 요소들을 종합하면 IB 평가의 타당성이라는 더 넓은 개념을 형성합니다.

마지막으로 IB는 **구성 타당성**에 가장 높은 우선순위를 둡니다. 즉, 평가가 의도한 특성과 역량을 실제로 평가하는 것을 의미합니다. 그럼에도 불구하고, 이러한 우선순위 부여가 타당성 사슬의 다른 측면들을 완전히 무시해서는 안 된다는 점을 염두에 두어야 합니다.

신뢰성

평가에서 신뢰성(reliability)이란 시험 절차가 반복될 경우 동일한 시험 결과를 얻을 가능성으로 정의할 수 있습니다. 이는 반드시 학생이 '올바른 결과'를 얻는다는 의미와 동일하지는 않습니다.

Winkley and Cresswell(2011)은 신뢰성 개념에 대한 서론에서 학생 성과 채점과 같이 신뢰성이 요구되는 영역들과 함께 신뢰성에 위협이 될 수 있는 다양한 요인들의 사례를 제시합니다.

- 1. 채점자 간 신뢰성(Inter-marker reliability):** 채점관마다 특정 문제에 대해 관대하거나 엄격한 정도가 다를 수 있으며, 동일한 채점관이라 하더라도 날에 따라 채점 기준이 달라질 수 있습니다.
- 2. 학생 성과의 가변성:** 학생의 시험 성과는 날마다 달라질 수 있으며, 특히 시험 조건이 변경되었을 때 더욱 그렇습니다. 이러한 조건에는 시험이 오전 또는 오후에 시행되는지, 누가 시험을 감독하는지, 전날 밤의 수면 상태, 또는 시험장 외부의 소음(예: 잔디를 깎는 소리) 등이 포함될 수 있습니다.
- 3. 서로 다른 시험:** 시험마다 출제되는 문제는 달라지며, 이는 학생의 이해도를 다양한 측면에서 평가할 수 있습니다(시험은 일반적으로 교육과정의 일부만 다루는데, 이는 모든 내용을 평가할 시간이 부족하고 학생들이 특정 주제는 복습하되 다른 주제는 그렇지 않을 수 있기 때문입니다).
- 4. 연도별 결과 간의 비교 가능성:** 이 비교 가능성은 학생 결과를 해마다 유의미하게 비교할 수 있도록 합니다.

5. **시험 상세 내용 간의 차이:** 시험 상세 내용이나 교수요목이 변경되었을 때 시간 경과에 따른 비교 가능성을 유지하는 것은 도전적일 수 있습니다.
6. **서로 다른 유형의 평가 활동:** 많은 자격 요건은 다양한 유형의 평가 활동으로 구성되며 이러한 서로 다른 평가 방식은 각각 상이한 신뢰성 문제를 동반합니다.
7. **서로 다른 유형의 질문:** 학생은 질문의 유형에 따라 서로 다른 성과를 보일 수 있습니다.

실제로 IB는 보통 학생이 평가를 한 번 치른다고 예상하며 평가 과정에서 결과의 **일관성**을 추구합니다. 따라서 IB는 이 목록의 1, 3, 6, 7번 항목에 특히 중점을 둡니다.

IB는 점수를 부여하는 기관으로서 평가의 신뢰성을 높이기 위한 다양한 조치를 취하고 있습니다. 채점의 신뢰성을 보장하는 것은 핵심 요소이며, 표준화, 품질 모델, 조정 과정의 주된 목적은 채점관들이 일관된 기준으로 채점하도록 보장하는 것입니다.

채점의 신뢰성

다음 간단한 활동은 채점 신뢰성의 개념과 실제 적용 방식을 보여줍니다.

[그림 11]은 IB 언어습득 평가의 실제 답변에서 발췌한 다섯 가지 예시를 제공합니다.

채점 기준표를 사용하여 언어의 질을 8점 만점으로 평가하고, 각 답안에 부여할 점수를 기록합니다.

모든 답안에 점수를 부여한 후 자신이 수행한 채점의 질에 대한 피드백을 확인합니다. 피드백은 학생 답안지 뒤에 확인할 수 있습니다.

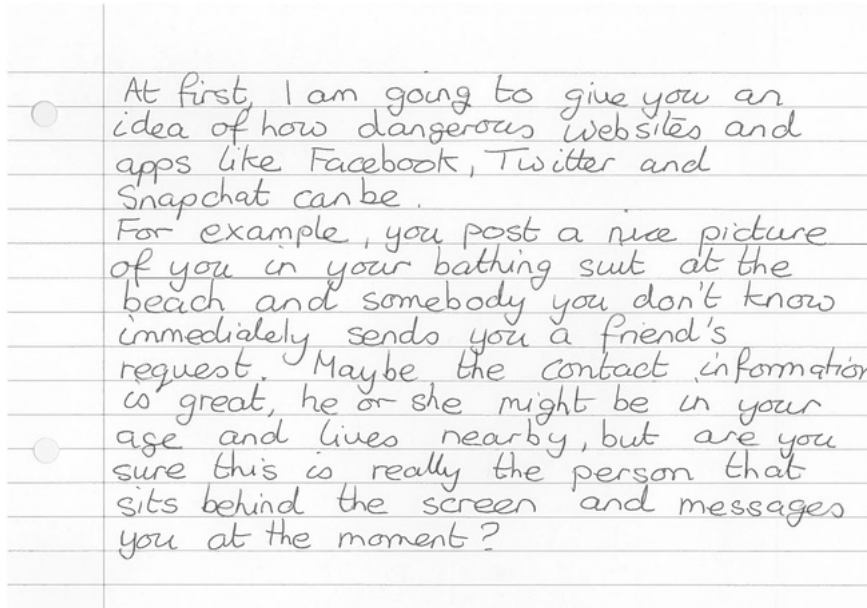
채점 기준표

학생이 성찰 기능을 얼마나 효과적으로 정확하게 적용했습니까?

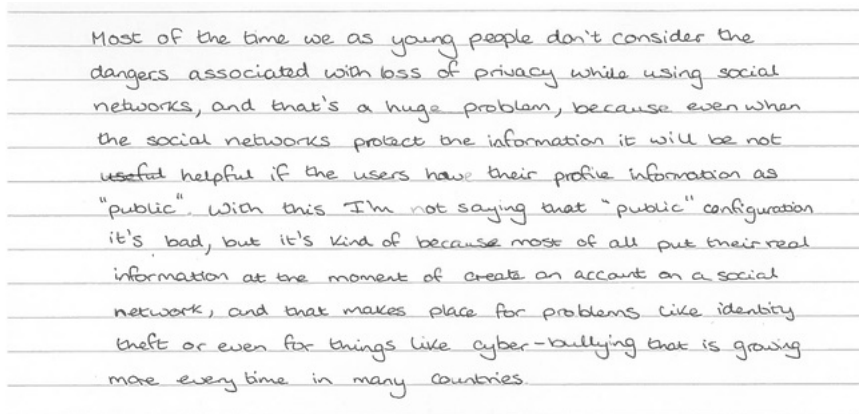
점수	단계 기술어
0	학생은 아래 단계 기술어에서 명시한 어느 기준도 충족하지 못합니다.
1~2	학생은 프로젝트가 자신의 학습에 미친 영향과 목표 달성 여부를 명시합니다.
3~4	학생은 프로젝트가 자신의 학습에 미친 영향과 목표 달성 여부를 요약하고, 이는 부분적인 근거로 뒷받침됩니다.
5~6	학생은 프로젝트가 자신의 학습에 미친 영향을 설명하고 성취 기준에 기반해 결과를 평가하며, 이는 근거로 뒷받침됩니다.
7~8	학생은 프로젝트가 자신의 학습에 미친 영향을 설명하고 성취 기준에 기반해 결과를 평가하며, 이것이 구체적인 근거 및/또는 상세한 예시로 충분히 뒷받침됩니다.

채점 기준표를 가이드로 삼아 답안지를 차례로 검토하여 각 답안에 점수를 부여합니다.

[그림 11A] 채점 신뢰성 활동



[그림 11B] 채점 신뢰성 활동



[그림 11C] 채점 신뢰성 활동

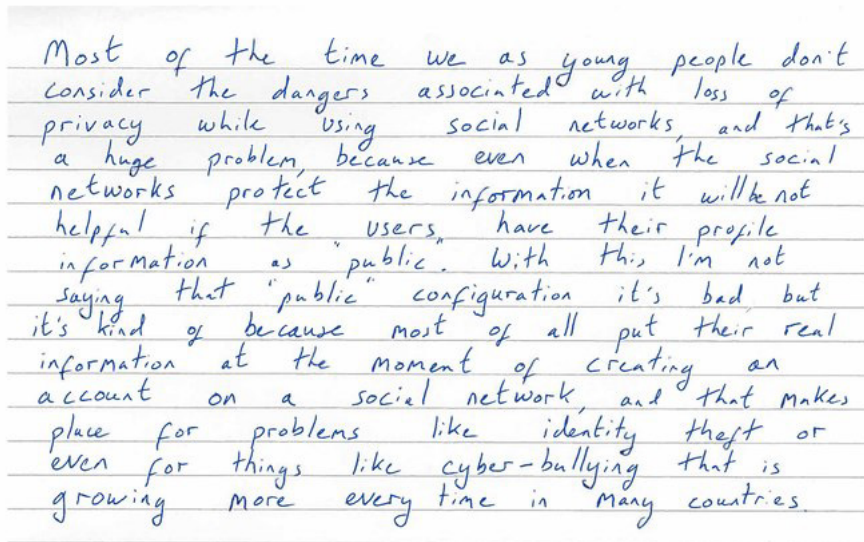
When people think in social networks, the first thing that comes in mind are "facebook", "Twitter" and "Instagram", am I wrong? On Facebook you can meet tons of friends from all around the world just by sending them a "friend request" and if they accept you, you are allowed to see every photo or information about this person.

So, how private do you think "FB" is? Well, actually it is not that private if you don't know to use it well. When you upload photos, you have to make sure that only your friends can see it.

[그림 11D] 채점 신뢰성 활동

We all use Facebook. By posting our personal information online, we could reunite our "long-lost friends", but simultaneously, other people having all sorts of intentions might be reading those information as well. Moreover, statistics revealed that social networks including Facebook make trillions of dollars every year through selling our information to companies, to individuals and even to crime groups. And the tragedy lies not only in overwhelmed advertisements, but also in potential exposure to cyber bullies and crimes.

[그림 11E] 채점 신뢰성 활동



각 답안지의 채점을 마치고 학생별로 점수를 부여한 뒤, 자신의 채점에 대한 피드백을 확인합니다.

피드백

학생 B와 학생 E의 작품이 서로 다른 필체로 쓰인 동일한 작품이라는 점을 눈치챘습니까? 학생 B와 학생 E에게 같은 점수를 부여했습니까?

특정 점수 구간표에서 두 가지 점수 중 어떤 점수를 부여할지 고민했을 수도 있습니다. 만약 다시 채점한다면 같은 결정을 내릴 것이라고 얼마나 확신할 수 있습니까?

여기서 핵심은 학생 과제의 질이 아니라, 같은 사람이 과제를 채점할 때마다 그 사람의 판단이 조금씩 달라질 수 있다는 점입니다. 대략적으로 과제의 질에 대한 관점은 매번 같지만 정확한 점수는 약간씩 변동될 수 있습니다.

며칠 뒤에 이 활동을 다시 수행하고 부여한 점수가 얼마나 비슷한지 확인해보십시오. (처음에 부여한 점수를 기억하려 하지 마십시오.)

IB가 채점 신뢰성을 어떻게 보장하는지에 대한 자세한 내용은 “[Reliability: How we make sure marking is fair](#)”(영문 제공, 동영상)를 참고하십시오.

일관된 결과와 ‘올바른’ 결과

높은 수준의 신뢰성을 확보하려는 목표는 어떤 채점관이 평가하더라도 학생이 같은(공정한) 점수를 받는 것이지, ‘올바른’ 점수를 받는 것이 아닙니다. 학생의 과제가 얼마나 우수한지는 전문적인 판단에 달려 있으며 두 명의 교사가 어떤 점수를 부여할지에 대해 의견이 다른 경우도 흔합니다. 신뢰성의 핵심은 두 교사 모두가 선임 채점관의 판단과 동일한 판단을 내리는 것입니다.

이는 성적에 대한 문의를 다룰 때 특히 어려운 부분입니다. 이러한 경우, 채점관은 마치 이 답안지를 처음 채점하는 것처럼 동일한 점수를 부여해야 하며, 성적 급간이나 개별 학생에게 미치는 영향과 같은 추가 정보에 의해 판단(긍정적 또는 부정적)이 좌우되어서는 안 됩니다.

교육 분야 외에서의 평가 신뢰성을 대한 이해 부족은 이미 잘 알려져 있지만, 시험 결과에 대한 대중적 논의가 점점 확대되는 현시점에서, 이 주제는 더욱 강조될 필요가 있습니다.

여러 문헌에서 제시된 바와 같이, 참가자들은 평가 신뢰성, 특히 측정의 부정확성이 이해하기 어려운 개념이라는 것을 알게 되었다.

(Chamberlain, 2010, p. 3)

가끔 시험이 개인의 능력을 공정하게 평가하지 못한다는 생각이 들곤 한다. 그렇지 않은가? 시험은 그 사람이 당시 어떤 상태였는지, 얼마나 성실하게 공부했는지, 그리고 그 외 모든 요인들에 더 초점이 맞춰져 있다. 시험을 잘 보지 못했던 사람이라도 이후에 두각을 나타내어 자기 분야에서 최고가 될 수 있다. (의료 분야 종사자, 남성)

(Chamberlain, 2010, p. 27)

교사와 학생의 약 63%가 ‘어떠한 수준의 오류도 용납될 수 없으며, 단 한 명의 학생이라도 잘못된 성적을 받아서는 안 된다’를 선택한 반면, 이들 중 50% 이상은 ‘시험 문제의 오탈자와 같은 회피 가능한 실수와, 두 채점관 간 불일치처럼 피할 수 없는 실수는 차이가 있다’를 선택하여 오류에 대한 관용을 보였다. 이러한 불일치는 신뢰성에 대한 지식과 신뢰성 부족에 대한 태도 사이의 약한 관계를 보여준다.

(He et al., 2010, p. 27)

구성개념 적합성 및 진위성

우리는 측정하려는 것을 얼마나 정확하게 측정하고 있습니까? 측정하고자 하는 기능과 지식을 정확히 측정하는 것은 때때로 ‘구성 타당성’으로 불리기도 합니다. 그러나 전반적인 타당성과의 혼동을 피하기 위해 본 간행물에서는 **구성개념 적합성**이라는 용어를 사용합니다.

진위성 있는 평가는 구성개념 적합성과 밀접하게 연관되어 있습니다. **진위성**이란 실제 가능한 범위 내에서 학생이 현실 세계에서 문제를 접하게 될 상황에 맞는 방식으로 평가가 이루어지는 것을 의미합니다. 과제를 맥락에서 분리하거나, 지나치게 단순화하거나, 명백히 인위적으로 구성된 평가는 진위성이 결여된 평가의 예입니다.

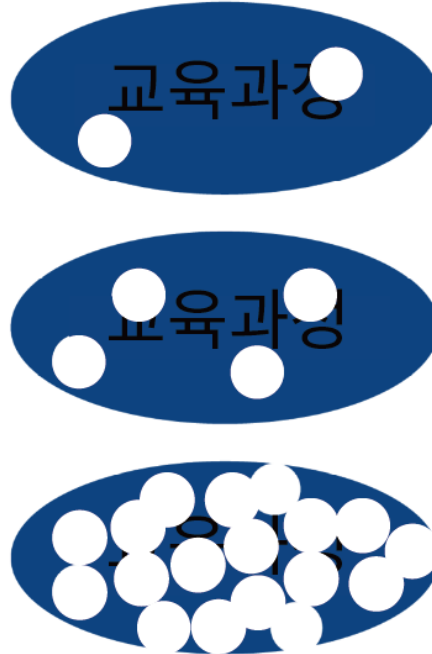
구성개념 적합성이 낮은 부실한 평가 설계의 예는 쉽게 찾아 볼 수 있습니다. 예를 들어 구술 시험에서 편지 작성 능력을 평가하는 경우가 그렇습니다. 그러나 널리 사용되는 과제 설정 접근 방법조차도 실제로 의도한 것을 평가하지 못하는 경우가 많습니다. 예를 들어, 전통적인 문학 시험이 구성되는 방법을 살펴보겠습니다. 학생은 특정 텍스트에서 사용된 문학적 장치에 대해 배우고, 평가는 이러한 장치 중 하나와 관련된 에세이를 작성하도록 요구합니다. 만일 학생이 단순히 교사가 말한 모든 내용을 암기한 것이라면 이것이 문학을 진정으로 이해했다고 할 수 있습니까?

특정 평가 과제에서 학생이 발휘하고자 하는 바를 명확히 이해하고, 실제로 그러한 목적을 얼마나 충실히 달성하는지를 탐구적이고 비판적으로 검토하는 과정이 적절한 수준의 구성개념 적합성을 확보하는 데 있어 최선의 실천이라 할 수 있습니다. 학생이 주어진 과제를 수행하기 위해 필요로 하는 다른 기능들을 고려하는 것은 특히 유용합니다. 에세이 또는 프로젝트와 같은 개방형 과제는 과제가 본래 의도한 연구 및 분석 능력을 보여주기 위해, 학생에게 높은 수준의 글쓰기 실력을 요구하게 된다는 취약점이 있습니다. 이는 평가에서 평가하고자 하는 내용과 무관한 장벽을 제거하고자 하는 보편적 평가 설계(Universal Design of Assessment) 개념에 내포되어 있습니다(Dolan et al, 2013).

구성개념 적합성은 특히 **교육과정**과 긴밀하게 연관되어 있습니다. 많은 경우 평가는 교육과정 자료의 일부만을 기반으로 질문을 구성합니다. 이러한 교육과정 범위가 충분한지 여부는 구성개념 적합성의

문제이자 판단의 문제입니다. 마찬가지로 어떤 질문 유형을 사용할지를 결정하는 것은 평가 설계에서 매우 중요합니다. 질문 또는 과제 유형마다 평가할 수 있는 구성개념의 유형이 다르기 때문입니다.

[그림 12] 시험 및 교육과정 범위



관리 가능성

관리 가능성의 요소는 아직 포괄적으로 연구된 바가 없으며, 이를 측정하기 위한 단일한 정의나 접근 방법도 존재하지 않습니다. 일반적인 용어로 ‘관리 가능성’이란 평가를 수행하는 데 학생, 학교, IB가 참여야 하는 노력을 의미합니다.

학생 관리 가능성은 대체로 평가를 완수하는 데 필요한 노력과 연관됩니다. 예를 들어 만 18세 학생에게 8시간 동안 진행되는 시험을 치르게 하는 것은 과도한 요구로 간주됩니다. 이와 마찬가지로 평가가 수행되는 시간대와 특정 학생이 하루 동안 시험을 치르는 총 시간 역시 학생의 관리 가능성을 평가할 때 고려되어야 합니다. IB는 학생이 교육과정의 일부로 다른 IB 평가를 함께 수행함으로써 발생하는 누적 부담을 고려해야 합니다. 이는 단순히 개별 **과목**이 아니라 **프로그램** 전체의 타당성을 고려한다는 IB의 원칙을 다시 생각해보게 됩니다.

학교는 평가를 위해 필요한 자료를 제공해야 할 수 있으며, 이는 해당 평가의 관리 가능성에 영향을 미칩니다. 예를 들어 공학 분야의 직업 과정에서는 각 학생이 조립할 엔진을 하나씩 보유해야 할 수도 있습니다. 이는 학생 수에 따라 필요한 엔진 수가 현실적으로 감당하기 어려운 수준이 될 수 있습니다.

학교 관점에서 관리 가능성의 또 다른 측면은 학생 과제가 채점을 위해 IB에 어떻게 전달되는가입니다. 발표를 녹음하거나 촬영해야 하는 경우는 단순히 서면 과제를 제출하는 것보다 학교에 더 많은 부담을 줍니다.

마지막으로 **IB**는 검토해야 할 평가 제출 과제의 분량 측면에서 관리 가능성을 고려합니다. 예를 들어, 세 시간 분량의 드라마 제작 과정에 참여하는 것이 학생의 능력을 가장 잘 보여줄 수 있는 증거일 수

있으나, 이를 외부 평가 과제로 삼기에는 실질적으로 실행이 어렵습니다. 다양한 선택 문항을 제공하는 것도 관리 가능성에 영향을 미칩니다. 이는 선택지 간 공통 기준을 정립하는 데 신중하고 시간 소모적인 작업이 필요하기 때문입니다.

관리 가능성, 신뢰성, 구성개념 적합성 사이에는 종종 긴장 관계가 존재합니다. 예를 들어, 평가 횟수나 시험 시간을 늘리면 학생이 전체 교육과정을 얼마나 이해했는지 보여주는 근거를 더 많이 확보할 수 있습니다. 이렇게 하면 관대한 채점과 엄격한 채점이 서로 상쇄될 가능성이 높아져 신뢰성이 향상됩니다. 그러나 이는 평가의 본래 목적보다는, 장시간 시험에서 학생의 수행을 유지할 수 있는 능력을 측정하는 평가로 이어져 결과적으로 구성개념 적합성과 학생의 관리 가능성을 오히려 저하시킵니다.

이러한 이유로 IB는 특히 각 과정의 총 평가량을 중심으로 평가의 관리 가능성을 엄격하게 관리합니다.

공정성과 편향

시험은 학생에게 의도하지 않은 이익이나 불이익을 줄 때 편향되었다고 합니다. 편향된 시험의 예시는 다음과 같습니다.

- 라틴어로만 쓰여진 역사 시험
- 모든 문제가 크리켓 경기의 득점 상황을 바탕으로 구성된 수학 시험
- 바닥에서 2m 높이에 이젤이 설치된 상태에서 치르는 미술 시험

이러한 사례들은 일부 학생들(예: 세 번째 예시에서 키가 작은 학생들)에게 상당히 **불리함**을 초래할 수 있습니다. 실제로는 대부분의 편향 사례가 위 예시들보다 훨씬 미묘하지만, 주의를 기울이지 않으면 학생의 결과에 상당한 차이를 만들 수 있습니다. 특히 편향을 유발하지 않으면서 문항에 맥락을 부여하는 일은 매우 어려운 과제이며, IB의 국제적 특성을 고려할 때 어떤 학생에게는 익숙한 상황이 다른 학생에게는 매우 낯설 수 있기 때문입니다.

‘편향’이란 측정하고자 하는 소질이나 성과의 실제 차이와 무관하게 평가 결과에 차이가 발생하는 경우를 의미합니다. 편향은 평가가 **진행되는 방식**이나 **채점될 때**(채점 신뢰성의 문제), 또는 평가 **과제** 그 자체에서 발생할 수 있습니다.

평가 진행 과정에서 발생하는 편향

본 섹션 앞부분에서 제시한 예시들은 평가 진행 과정에서 발생하는 편향을 보여줍니다. 이 경우 일부 학생에게 이젤이 너무 높은 위치에 설치된 비현실적인 상황을 예로 든 것입니다. 평가가 진행될 때, 특히 시험 도중 편향이 발생하는 양상은 다양합니다. 가장 흔히 발생하는 편향은 시험 시기와 관계가 있습니다. 일부 지역에서 극심한 기온이나 대기 오염이 발생하는 시기, 또는 일부 학생들이 제대로 집중하거나 준비할 수 없는 기간에 시행되는 시험은 편향을 초래합니다. 이러한 문제는 전 세계 여러 국가와 지역(영역)의 서로 다른 요구 사항을 조율해야 하는 IB에게 특히 큰 도전 과제입니다. 모든 학교에 적합한 시험 날짜를 정하는 것은 현실적으로 불가능하며, 어떤 날짜를 선택하더라도 일부 학교에는 부적절할 수밖에 없습니다.

흔히 발생하는 또 다른 편향은 시험장 환경의 차이에서 비롯될 수 있습니다. 한 학생은 직사광선 아래에, 다른 학생은 어둡고 그늘진 자리에서 시험을 치르는 상황을 생각해 봅니다. 시험 규정에 대한 일관성 없는 관행도 또 다른 예시가 될 수 있습니다. 어떤 학생에게는 시간을 엄격하게 적용하면서, 다른 학생에게는 시간적 여유를 허용하는 경우가 이에 해당합니다. IB는 이러한 문제를 관리하기 위해 각

프로그램의 「평가 절차」에 명확하고 일관된 규정을 제시하고 있으며, 이를 위반하는 경우 행정 실책으로 간주하여 처리합니다.

채점 과정에서 발생하는 편향

대부분의 채점 편향은 의도적이기보다는 무의식적으로 발생합니다. 인간의 사고는 빠르게 의사 결정을 내리기 위해 단순화된 사고 방식을 사용하도록 설계되어 있으며, 이는 종종 무의식적 편향으로 이어지게 됩니다. IB는 이러한 편향을 완화할 의무가 있지만, 무의식적 편향 자체를 닦할 수는 없습니다.

채점 과정에서 편향은 여러 이유로 발생할 수 있습니다. 학생의 필체에 대한 개인적 선호(Hughes et al, 1983 참고), 학생의 성별(채점관이 알거나 추정하는 경우)에 따른 우대, 그리고 일부 평가 맥락에서는 크게 관련이 없는 양식, 구두점 및 철자와 같은 요소에 지나치게 주의를 기울이는 경우 등이 있습니다. 이러한 편향은 채점관의 훈련과 채점 결과 검토를 통해 해결할 수 있습니다.

성별, 국적, 학교와 같은 요인에 기반한 무의식적 편향은 문서에도 잘 나타나 있습니다. 이를 최소화하기 위해 IB는 평가를 위해 제출된 모든 과제를 채점 전에 익명화합니다.

또 다른 대표적인 편향 유형으로는 ‘후광 효과(halo effect)’가 있습니다. 이 경우 채점관은 학생의 초기 답변이 매우 우수할 경우, 해당 학생에 대해 전반적으로 긍정적인 인상을 형성하게 되고, 그 결과 이후 문항에서도 필요 이상으로 관대하게 점수를 부여할 수 있습니다.

채점관이 특정 주제의 답안을 공정하고 편향 없이 채점할 수 없다고 판단하는 경우, IB는 제출된 과제의 학문적 가치에 기반하여 편향 없이 채점할 수 있는 사람이 평가할 수 있도록 하는 절차를 마련해 두고 있습니다.

평가 과제와 관련한 편향

편향은 평가 결과의 차이가 측정하고자 하는 대상과 관련이 없을 때 발생합니다. 예를 들어, 특정 문화적 사례를 수학 문제의 배경으로 삼는 경우, 해당 문화 전통에 대한 지식 부족(평가 요구 사항이 아님)이 문제에 대한 이해도와 수행 능력(평가 요구 사항)을 저해할 수 있습니다.

심리측정학적 기반 시험을 구성할 때, 사전 시험에서 예외적인 응답 특성을 보이거나, 학생 집단의 서로 다른 하위 그룹 간에 응답 특성이 현저히 다르게 나타나는 문항(‘차별 기능 문항 기법’)은 편향된 것으로 간주되어 시험에서 제외됩니다. 학생 하위 그룹은 성별, 민족성, 사회 계층, 언어 능숙도 등으로 구분될 수 있으며, 평가하고자 하는 구성개념과 무관하다고 볼 수 있는 특성으로 정의될 수 있습니다.

특정 평가 또는 평가 요소가 편향되었는지 판단할 때, 과제가 기본 구성개념과 어떻게 명확히 연계되는지, 편향을 유발할 수 있는 요소는 무엇인지 주의 깊게 고려해야 합니다. 결정을 순수하게 통계적 근거만으로 내릴 경우, 서로 다른 집단을 구분하도록 설계된 평가와, 편향된 평가를 혼동할 위험이 있습니다. 골드스타인(Goldstein, 1996)과 험프리스(Humphreys, 1986)는 객관적으로 확인된 사실인 ‘차이’와 그 차이의 적합성에 대한 판단인 ‘편향’을 구별하는 것이 유용하다고 제안했습니다.

블랙(Black, 1999)은 평가 문제가 서로 다른 학생들에게 불공정한 영향을 미칠 수 있는 가장 일반적인 여섯 가지 가능성을 제시합니다.

1. 문제가 설정된 맥락은 해당 맥락에 익숙한 학생에게 유리할 수 있습니다. 예를 들어 미국 문화와 관련된 내용을 언급하면 다른 국가의 학생들에 비해 미국에 거주하는 학생에게 유리하게 작용할 수 있습니다.

2. 인간 관계에 관한 에세이 질문은 감정 표현이 자유로운 문화나 가족 배경을 가진 학생에게 유리할 수 있습니다.
3. 선다형 문제는 남학생에게 유리할 수 있습니다.
4. 과제물이나 프로젝트 과제는 여학생에게 유리할 수 있습니다.
5. 특정 사회경제적 배경과 연관된 언어나 관습을 사용하는 문제는 다른 배경을 가진 학생에게 불리할 수 있습니다.
6. 어떤 문제들은 특정 문화적 맥락에서만 이해될 수 있습니다. 예를 들어 독거노인에 관한 문제는 일부 문화적 맥락에서는 낯설지만, 다른 문화에서는 익숙할 수 있습니다.

또한, 성별 편향에 대한 증거가 특히 미국에서 존재하지만, 질문 형식의 어떤 측면이 이러한 결과에 기여하는지는 여전히 명확하지 않습니다.

평가를 설계할 때 특정 평가에 응시하는 학생들 간의 차이에 어떻게 대응해야 할지는 분명하지 않습니다. 평가는 다양한 과제와 문제 유형을 통해 편향의 전반적 영향을 줄이는 방향으로 설계되어야 합니다. 명시적이든 암묵적이든 모든 형태의 문화적 또는 성별 고정관념은 반드시 지양되어야 합니다. 각 문항 내용은 불공정함을 유발하는 것으로 알려진 명백한 유형을 피하기 위해 면밀히 검토되어야 하며, 학생 집단의 여러 그룹 표본을 대상으로 한 사전 테스트를 통해 숨겨진 편향 사례를 발견할 수 있습니다. 그러나 편향 가능성이 있는 모든 문항 유형과 시나리오를 배제할 경우, 평가 설계자와 문항 구성자가 선택할 수 있는 범위는 매우 제한되며, 이로 인한 제약은 오히려 평가의 타당성에 부정적인 영향을 미칠 것입니다. 명백하고 불필요한 위험 요소를 피하는 것을 전제로, 다양한 유형의 평가 과제와 형식을 활용하여 균형 잡힌 방식으로 평가를 설계하는 것이 가장 합리적인 해결책으로 보입니다.

또한 ‘인구 집단 내에서 서로 다르게 정의되는 집단을 어디까지 고려해야 하는가’라는 문제도 제기됩니다. 예를 들어 학습 유형이 서로 다른 학생들을 고려해야 합니까? 히에로니무스(Hieronymus)와 후버(Hoover)(1986)가 언급한 것처럼, 관심사와 동기의 차이를 편향 요인으로 여긴다면 모든 과제 또는 평가 방법이 어느 정도 편향을 지닌다고 볼 수 있습니다. 예를 들어 언어 시험에서 사용되는 텍스트 구절은 필연적으로 일부 학생에게는 더 흥미롭고 다른 학생에게는 덜 흥미로울 수밖에 없습니다. 편향을 배제한 평가에서의 형평성은 중요한 문제이며, 특히 평가 도구에 명백한 편향이 존재할 경우 소송으로 이어질 수 있는 특정 국가나 지역(영역)에서는 더욱 그렇습니다. 그러나 성과 차이와 달리 편향의 증명은 대개 섬세한 판단을 요구하며 이는 평가가 수행되는 특정 사회적 맥락과 밀접하게 연관되어 있습니다.

깁스(Gipps)와 머피(Murphy)(1994)는 “A Fair Test? Assessment, Achievement and Equity”(공정한 시험? 평가, 성취, 공정성)에서 “공정한 시험이란 것은 존재할 수 없고 있을 수도 없다. 실제 상황은 너무 복잡하며 관념은 너무 단순하다.”라고 결론지었습니다. 그러나 이것이 평가 설계자와 문제 출제자가 편향과 불공정의 영향을 줄이기 위해 최선을 다하지 않아도 된다는 의미는 아닙니다. 깁스와 머피는 또한 평가 설계자가 개별 시험 문제의 답안 통계에 따라 조작함으로써 만들어진 결과의 평등보다는, 평가에 대한 기회와 접근의 평등을 목표로 삼아야 한다는 관점을 유지했습니다. 예를 들어 남학생의 상대적 성과를 높이기 위해 영어 시험에 선다형 문제를 도입하는 것이 어디까지 정당화될 수 있는지에 대해 그들은 의문을 제기합니다. 이는 널리 받아들여지는 그 용어의 정의에 따라 평가의 타당성을 왜곡할 수 있기 때문입니다.

공정성이 결여된 평가 과정은 교육의 불평등을 초래하는 요인 중 하나일 뿐이며, 다른 요인들보다 덜 중요한 것일 수 있다는 점은 널리 인식되고 있습니다. 그 밖에도 교육에는 학생의 성과에 큰 영향을 미치는 다음과 같은 불평등 요소가 존재합니다.

- 학교 내 교수의 질 차이

- 학교 간, 그리고 지역 간 교육 자원 수준의 차이
- 개별 학생에게 주어지는 사회 및 가족의 지원 수준의 차이

이러한 요소들 중 어느 하나라도 개별 학생의 성공에 상당한 영향을 미칠 수 있으며, 아무리 공정한 평가 과정이라도 이를 보상할 수 없습니다. 예를 들어 스미스(Smith)와 톰린슨(Tomlinson)(1989)은 학교 효과성(School effectiveness)이 학생의 민족성보다 시험 결과의 차이를 결정짓는 더 큰 요인임을 밝혔으며 이는 서로 다른 집단의 성과 차이를 좁히기 위해 평가 도구를 조정하려는 시도가 때때로 부적절할 수 있음을 시사합니다.

이러한 고려는 성과보다는 소질을 평가하려는 근거를 형성했지만, 사회적 배경과 교육 경험으로부터 완전히 분리된 순수한 소질, 능력 또는 잠재력을 평가하는 것은 불가능하다는 점이 분명해졌습니다. 학업 성취를 사회적 및 문화적 맥락과 완전히 분리하여 바라보는 것 또한 불가능합니다. 교육적 성공의 개념은 특정 사회의 제한된 부분의 기준에 따라 정의되고 측정됩니다.

평가에서 편향 및 장벽 제거하기

앞서 언급한 편향 외에도 반드시 다루어야 할 또 다른 측면은 평가 과제가 어떤 학생에게든 불공정하게 차별할 가능성이 있다는 점입니다. 이는 평가 조건이 평가 장벽을 제거하기 위한 필요한 조치를 시행할 수 있도록 보장함으로써 이루어집니다. 이를 통해 모든 학생이 다른 학생과 동등한 조건에서 자신의 학업 성취 수준을 입증할 수 있게 됩니다.

본 간행물의 ‘모든 학생을 위한 공정성’ 섹션에서는 이러한 개념을 보다 자세히 다루고 있습니다.

편향 인지하기

편향이 긍정적인 동시에 부정적일 수 있다는 점을 항상 염두에 두는 것이 중요합니다. 어떤 과제가 특정 학생 집단에게 특히 익숙하거나 수행하기 쉬운 경우, 그 역시 편향에 해당합니다. 공정한 평가의 목적은 모든 학생에게 동등한 기회를 제공하는 것입니다.

본 간행물에서 평가 설계 과정 또는 채점 과정에서 편향이 어떻게 발생할 수 있는지를 설명해 왔습니다. 이러한 평가 주기의 단계에서 편향을 사전에 고려하는 것은 매우 중요하지만 잠재적 편향에 대해 생각하는 것만으로는 충분하지 않습니다. 오히려 학생 결과에서 편향의 증거를 찾아야 하며, 전체 시험과 비교하여 특정 문제에서 학생이 어떻게 수행했는지 분석해야 합니다.

또한 편향과 관련된 판단은 고정관념이 아닌 증거에 기반하는 것이 중요합니다.

비교 가능성

비교 가능성은 타당성에서 가장 복합적인 측면 중 하나입니다. 평가 결과는 선별 목적으로 학생들을 비교하는 데 자주 활용됩니다. 두 명의 학생이 동시에 같은 시험을 본 경우 7점을 받은 학생이 4점을 받은 학생보다 해당 시험에서 더 높은 성과를 보였다고 합리적으로 확신할 수 있습니다. 그러나 현실에서는 이보다 훨씬 복잡한 비교가 이루어집니다. 이러한 비교의 예시는 다음과 같습니다.

- 역사에서 6점을 받았지만 서로 다른 문제에 답하거나 다른 선택지를 택한 두 학생
- 한 명은 2025년 5월에, 다른 한 명은 2023년 11월에 스페인 문학 시험에서 5점을 받은 두 학생
- 한 명은 물리학에서 4점을, 다른 한 명은 화학에서 4점을 받은 두 학생

- 한 명은 수학에서 4점을, 다른 한 명은 지리에서 4점을 받은 두 학생
- 한 지리적 위치에 속한 학생이 컴퓨터 과학에서 3점을, 중국 문학에서 6점을 받았고 다른 지리적 위치에 속한 학생이 일본 문학에서 5점을, 생물학에서 4점을 받은 경우
- 15세 학생 두 명 중 한 명은 MYP 수료증을 받았고 다른 한 명은 다른 수여 기관의 자격증을 취득한 경우
- 한 명은 표준 수준(SL) 인도네시아어 B에서 6점을, 다른 한 명은 심화 수준(HL) 인도네시아어 B에서 5점을 받은 두 학생

비교 가능성은 두 평가 결과가 어떤 의미에서 동등하다고 볼 수 있는지를 묻는 개념입니다. 과목 간에는 비교 가능성이 특히 어려운데, IB가 서로 다른 것들을 평가한 후 그것들이 동등한 가치가 있는지 판단해야 하기 때문입니다.

타당성이 평가의 추상적 특성이 아니라 특정 목적을 위한 것이라는 개념이 여기서 특히 중요합니다. 학생의 음악 성적은 시각 예술 성적보다 전문 음악가가 될 준비 상태를 더 잘 나타내지만, 역사를 학습할 준비가 되었는지 예측하는 데 있어서는 두 성적 모두 동등할 수 있습니다.

학생마다 강점과 약점이 달라 어떤 과목은 더 쉽게 성과를 내기 때문에 비교 가능성 문제는 더욱 복잡해집니다. 특히 여러 과정을 선택할 수 있는 IB의 구조상, 과목마다 시험을 치르는 학생 집단이 달라지기 때문에 비교 가능성은 IB에 고유한 도전 과제로 다가옵니다.

IB는 비교 가능성에 대해 다음의 세 가지 원칙을 유지하려 노력합니다.

- 한 연도 내 기수 간, 그리고 연도 간에 교과 또는 학문에서 성적을 부여하는 과제의 기준이 비교 가능합니다.
- 교과 또는 학문 간의 성적이 일관된 의미를 가지고 있어 프로그램 수료를 위한 다양한 경로(예: IB 디플로마, MYP 수료증)가 비교 가능합니다.
- IB는 고차원적 사고 기능의 평가에 전념하면서도 다른 자격증에서 취득한 성적이 IB 성적과 여전히 비교 가능하도록 주의를 기울입니다.

비교 가능성 측정하기

두 가지 평가가 비교 가능한지 측정하는 방법은 다양하며, 이 주제에 대한 많은 학술 논문들이 존재합니다. 코에(Coe)와 그 외 연구자들(2008)은 과목 간 비교 가능성에 관한 문헌을 검토하면서, 서로 다른 과목의 시험 난이도를 비교하는 방법을 통계적 방법과 판단적 방법의 두 가지 넓은 범주로 분류했습니다.

통계적 방법은 학생의 평가 성과를 비교하고 추세를 살펴보는 데 초점을 맞춥니다. 이는 두 평가가 비교 가능하다면, 두 평가에 모두 응시한 학생들의 충분한 무작위 표본을 고려할 때 두 평가의 결과가 일반적으로 동일해야 한다는 개념에 기반합니다.

반대로 판단적 방법은 과목 전문가들이 여러 평가를 검토하고 상대적 난이도에 대한 숙고된 의견을 제시하는 방식입니다. 비교 대상이 동일한 수준인지 확인하기 위해 다양한 연구 도구와 기법이 사용됩니다.

두 가지 접근 방법 모두 심각한 개념적 단점을 가지고 있는 것으로 인식되며 코에(Coe)와 그 외 연구자들(2008)은 각 기법에 대해 여섯 가지 광범위한 비판을 제시했습니다.

<표 4> 평가 비교를 위한 통계적 및 판단적 방법에 대한 비판

통계적 방법에 대한 비판	판단적 방법에 대한 비판
<ul style="list-style-type: none"> 연구에서 난이도 외 요인(교수 또는 동기 등)을 측정할 수 있음. 다차원성-과목들이 공통된 특징을 갖지 않을 수 있음. 비전형성-통계가 본질적으로 편향된 학생 집단에 기반하는가? 하위 집단 차이-서로 다른 학생 집단이 서로 다른 난이도를 경험한다면, 상대적 비교 가능성 결론에 문제가 생기는가? 방법 간의 불일치-어떤 방법이 '올바르다'고 할 수 있는가? 평등 강요의 문제-자격을 얻은 학생에게 어떤 영향을 미치는가? 	<ul style="list-style-type: none"> 기준의 범주-서로 다른 과목에 걸쳐 적용하려면 기준이 매우 광범위해야 하므로 부정확해질 수 있음. 요구 수준이 다른 문제에 대한 점수 부여-채점관은 어려운 문제의 덜 우수한 답변에 비해 쉬운 문제의 우수한 답변에 더 많은 점수를 주는 경향이 있음. 서로 다른 과제 유형에 점수 부여-채점관은 단답형과 에세이와 같이 서로 다른 유형의 과제를 비교하는 데 어려움을 겪음. '판단' 방법 역시 통계적 비교에 기반함-일반 학생의 예상 성과 경험에 기반함. 해석과 맥락-학생의 능력이나 공정성에 대한 판단적 평가는 단일 기말 시험과 일련의 모듈식 평가 간의 구조적 차이를 고려해야 함. 판단 종합-대부분의 평가는 여러 기준을 측정하므로 균형을 맞춰야 함.

각 방법론의 지지자들은 대체 접근 방법에 매우 비판적이며, 일부는 현재 존재하는 모든 접근 방법에 근본적인 결함이 있다고 주장합니다. 이 문제를 분석하기 위해 코에와 그 외 연구자들(2008)은 과목 간 비교 가능성을 측정하는 다섯 가지 방법을 영국의 GCSE와 A 레벨 과정에 적용했습니다. 이들은 통계적 방법과 판단적 방법이 과목 간 비교 가능성을 측정하는 데 상당히 높은 수준으로 일치했으며, 두 방법 간의 차이는 과목 간 차이에 비해 훨씬 적었다고 결론지었습니다. 또한 매년 과목별 상대적 난이도가 안정적이라는 점도 확인했습니다.

IB는 채점관의 판단, 통계적 분석, 교사의 피드백을 삼각 검증하여 연도 및 옵션 간의 비교 가능성을 유지합니다('성적 부여 및 종합' 섹션 참고). 또한 교과 또는 학문 수준에서도 통계적 및 전문가 판단 접근 방법을 모두 활용하여 비교 가능성을 검토합니다.

타당성에 대한 IB의 접근 방법

- IB는 구성개념 적합성과 진위성 있는 평가가 신뢰성을 극대화하는 것보다 더욱 중요하다고 믿습니다.
- IB는 통합적인 전인교육을 지지합니다. 프로그램 수준에서 타당성을 지원하는 것을 최우선 과제로 삼습니다. 이는 개별 과정이나 과정 내 선택 사항의 타당성보다 더 중요하게 생각합니다.

타당성이란 서로 상충하는 여러 중요 요구사항들 사이의 균형을 맞추는 복합적이고 다면적인 활동입니다. 서로 상충하는 타당성 요소들의 균형을 맞추는 단 하나의 올바른 방법은 없습니다. 균형을 어디에 두느냐는 궁극적으로 평가를 개발하는 기관의 가치에 기반한 판단입니다.

IB는 실제 세계를 반영하는 방식으로 중요한 것을 평가하는 데 높은 가치를 둡니다. 이 중 첫 번째 요점은 ‘구성개념 적합성’이라는 용어로 설명됩니다. 즉, IB 평가는 채점하기 쉬운 것이 아니라 과목에서 실제로 중요한 것이 무엇인지를 묻습니다. 두 번째 요점은 진위성으로, 이는 IB 평가가 인위적이고 부자연스러운 활동이 아니라 학생들이 실제 세계에서 경험할 수 있는 방식을 반영한 유의미한 과제를 포함한다는 것을 의미합니다.

이러한 목표들은 타당성의 다른 측면, 특히 평가의 신뢰성에 대가를 수반합니다. 의미 있고 진위성 있는 과제는 대체로 채점에 높은 수준의 주관성을 요구하며, 이는 명확한 정답이 있어 객관적으로 채점 가능한 선다형 시험과 달리 채점관 간 더 큰 변동성을 수용해야 함을 의미합니다. 게다가 이러한 과제는 평가의 관리 가능성에도 영향을 미칩니다. IB가 선호하는 평가 유형은 제작, 관리, 채점 모두에서 더 어렵고 시간이 많이 소요됩니다. 또한 학생 역시 더 많은 노력을 기울여야 합니다. 이렇게 하면 시험에서의 성과에만 집중하기보다 의미 있는 연구 과제를 수행해야 하는 등 학생의 과제량이 늘어나게 됩니다.

IB는 다른 기관이 평가 타당성의 균형을 유지하기 위해 다른 우선순위를 선택할 수 있음을 인정하면서도, 외부 검증된 IB 평가에 대해 IB의 입장이 적절하고 정당하다고 확신합니다.

IB는 성공을 위한 동기를 갖고 지식이 풍부하며 탐구심과 배려심이 많은 청소년을 양성하기 위해 다른 교육과정보다 더 많은 노력을 기울이는 것을 목적으로 합니다. IB 교육 목표에 요약된 바와 같이 IB는 교육 프로그램을 통해 학생들이 문화 간의 이해와 존중을 바탕으로 더 나은 세상을 만드는 데 기여하기를 희망합니다. 또한 IB의 모든 프로그램은 IB가 가치 있게 여기고 10가지 자질로 구성된 IB 학습자상을 바탕으로 학생을 양성하기 위해 최선을 다합니다.

평가의 관점에서 중요한 점은 IB의 목적이 개별 과목이나 학문 수준이 아닌 프로그램 수준에서 정의되었다는 점입니다. 따라서 평가 타당성에 대한 질문은 프로그램 수준에서 제기되어야 하며, 개별 성적뿐 아니라 전체 프로그램 수료증 수여를 규정하는 규칙을 포함해야 합니다.

이는 각 과정이나 개별 평가 과제를 고려하는 것이 중요하지 않다는 의미는 아닙니다. 실제로 일부 타당성 요소는 이러한 세부적인 수준에서만 의미를 갖기도 합니다. 그러나 전체적인 타당성 주장을 세울 때 IB는 학생이 이수한 전체 프로그램을 고려합니다.

이 자료에 제시된 대부분의 평가 원칙은 모든 IB 교육 전반에 적용되며 ‘섹션 C: IB 프로그램별 과정’에서는 각 프로그램을 순차적으로 살펴보고 고유한 특징을 논의합니다.

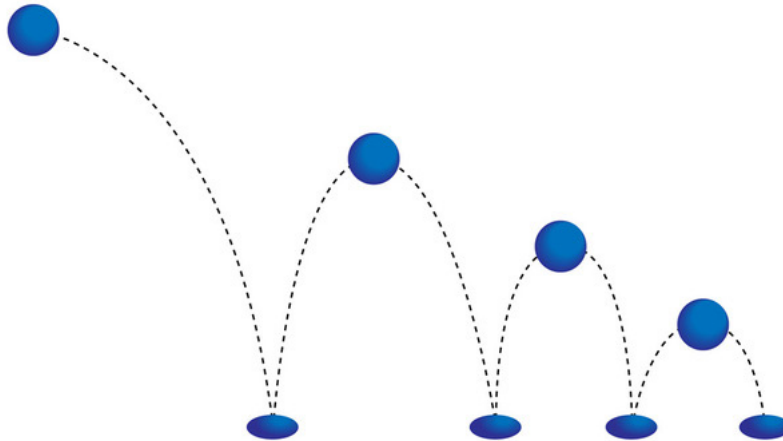
디지털 평가가 타당성에 미치는 이점

- 오늘날의 세상은 서면에 기반하지 않으며, 컴퓨터 기기는 삶의 모든 영역에 스며들어 있습니다.
- 서면 기반 시험으로는 움직이는 이미지를 제공하거나 학생과 의미 있는 상호작용을 구현할 수 없습니다.
- 기기의 다양한 활용성 덕분에 학생은 시험 몇 달 전에 특수 변형 시험지를 요청할 필요 없이, 시험 문제에 접근하기 위해 필요한 시각적, 청각적 설정을 스스로 조정할 수 있습니다.
- 여러 우려에도 불구하고 디지털 평가는 학생의 학업적 부정행위와 학교의 행정 실책으로부터 더 높은 수준의 보호를 제공합니다.

이메일, 문자, 소셜 미디어는 편지를 대체하는 주요 의사소통 수단이 되었으며, 컴퓨터 기기를 사용하는 것은 많은 업무에서 일상적인 일이 되었습니다. IB가 진정성 있는 평가를 목표로 한다면, 평가에서 테크놀로지와 기기 사용을 포함시키는 것이 타당합니다.

역사적으로 평가는 전통적인 시험 방식(예: 종이와 연필)으로 평가할 수 있는 범위가 제한적이었습니다. 서면 기반 시험에서는 학생이 평가를 조작하거나 상호작용할 수 있는 기회 없이 단순한 학습 자극 자료나 질문에만 답해야 했습니다. 가장 기본적인 수준에서 디지털 평가는 오디오와 동영상을 자극 자료로 활용할 수 있으며, 이는 감독관이 교실 전체에 자료를 재생하는 방식이 아니라 학생 각자가 직접 접근할 수 있게 합니다. 중등교육 프로그램(MYP)에서는 필요한 경우 물체의 움직임을 긴 설명과 정적인 그림으로 제시하는 대신, 움직이는 도표와 시뮬레이션을 디지털 평가에 포함할 수 있습니다.

[그림 13] 동적인 이미지 활용을 통한 컴퓨터 기반 평가 문제의 명확성 향상



평가 과제가 점점 정교해지면서, 테크놀로지를 활용해 학생이 문제를 어떻게 탐구하고, 새로운 정보에 어떻게 반응하며, 시뮬레이션과 어떻게 상호작용하는지를 평가할 수 있게 되었습니다. 지능형 학습 지도 시스템을 통해 디지털 평가는 학생의 사고 유도 질문에 대응할 수 있으며, 구술 시험과 같은 교사 주도 평가에서 제공되는 기회를 재현할 수 있습니다. 동시에 학생마다 서로 다른 교사가 배정됨으로써 발생할 수 있는 문제를 피할 수 있습니다. 따라서 디지털 평가의 주요 장점은 서면 기반 시험의 한계를 벗어나 IB가 실제로 평가하고자 하는 요소를 측정할 수 있는 새로운 과제를 설계할 수 있다는 점입니다.

타당성의 또 다른 중요한 측면은 편향을 최소화하는 것이며 특히 포용적인 접근성 조치가 필요한 경우 더욱 그렇습니다. IB는 다른 글꼴이나 다른 색상의 종이를 사용한 시험지 제작 요청을 정기적으로 처리하고 있으며, 점자로 제작된 시험지를 요청받는 경우도 있습니다.

디지털 평가가 이러한 장벽을 완전히 제거할 수는 없지만, 테크놀로지를 통해 학생은 자신의 요구나 선호에 맞는 글꼴 크기와 색상을 직접 선택할 수 있습니다. 또한 스크린 리더기를 지원하는 소프트웨어를 사용하는 학생들에게도 접근성 측면에서 큰 이점이 있습니다. 일부 학생은 기기 사용만으로 접근성 요구를 모두 충족하지 못할 수 있지만, 디지털 평가는 추가 지원 없이도 더 많은 학생이 평가에 접근할 수 있게 하며, 필요한 경우 그 지원을 활용할 수 있습니다.

디지털 평가에 관한 가장 큰 우려 중 하나이자 타당성에 대한 주요 위협은 시험 보안입니다. 일부 비평가들은 디지털 평가가 서면 기반 시험보다 보안성이 낮다고 주장하지만, 실제로 디지털 평가는 서면 기반 시험이 직면하는 일부 위험으로부터는 더 안전한 반면, 다른 유형의 위험에는 노출될 수 있습니다.

디지털 시험의 주요 장점 중 하나는 보안된 전송과 IB가 지정한 시간 범위 내에만 접근 가능하며, 학교에서 접근 권한을 부여하기 전까지 사용자가 시험 내용을 열람할 수 없다는 점입니다. 이는 누구나 열람할 수 있고 시험 며칠 전부터 각 학교가 책임지고 보관해야 하는 서면 기반 시험지 발송 방식과 비교할 때 훨씬 안전한 방식입니다.

IB는 시험이 진행되는 동안 학생들이 인터넷에 연결되는 기기를 사용할 수 있기 때문에 디지털 평가가 추가적인 어려움을 초래한다는 점을 인지하고 있습니다. 학생들이 기기를 통해 승인되지 않은 자료에 접근할 위험을 줄이기 위해 IB 디지털 시험은 시험이 진행되는 동안 다른 모든 콘텐츠에 대한 접근을 차단하는 ‘잠금 모드’로 운영되며, 이를 통해 시험 보안과 학문적 진실성을 뒷받칩니다. 또한 학생이 문제에 답하는 과정(예: 로그 파일)과 최종 답안을 모두 기록할 수 있어 부정행위가 의심되는 행동을 파악할 수 있습니다.

기준

- 모든 기준의 정의는 **교육과정, 평가, 성과**의 세 가지 요소로 구성됩니다.
- 형성 평가에서는 주로 교육과정과 평가 기준에 초점을 맞춥니다.
- IB는 **성취 기준 평가**(이전에는 ‘약한 절대 평가’로 알려짐)를 채택하고 있으며 이는 평가 기준의 균형과 이전 연도 학생 성과 비교를 통해 평가 기준을 설정하는 것을 의미합니다.
- 기준을 유지하는 것은 기준을 세우는 것만큼 중요합니다.

기준의 세 가지 측면

평가 맥락에서 ‘기준’이라는 용어는 일반적으로 학생에게 주어진 과제의 난이도를 의미합니다. 기준은 비교 가능성의 핵심 요소입니다. 기준이라는 개념은 다음과 같은 세 가지 관점에서 살펴볼 수 있습니다.

[그림 14] 기준의 세 가지 측면



기준의 세 가지 측면은 서로 조합되어 과목 전체의 난이도를 결정하지만, 각 측면에서의 변화는 영향력과 소요 기간이 서로 다릅니다.

IB의 경우, **교육과정**의 변경은 교육과정 검토 주기 동안 이루어지는 것이 일반적이지만, 필요에 따라 공식 검토 주기 외에도 변경이 이루어질 수 있습니다. 이때 공정성을 위해 IB는 교사가 이러한 변화를

반영하여 교수법을 조정할 수 있도록 충분한 시간을 제공해야 합니다. 디플로마 프로그램(DP)과 직업연계 프로그램(CP)의 경우 조정 기간은 이상적으로 2년(과정의 기간)입니다. 중등교육 프로그램(MYP)은 5년 과정이므로(실제 평가는 통상 마지막 2년을 대상으로 하지만) 이보다 더 긴 기간이 필요할 수도 있습니다.

평가 기준은 보통 12~18개월이 소요되는 시험 개발 주기에 의해 결정됩니다. 교육과정 기준과 달리, 동일한 난이도의 두 세트 문제를 설계하는 것은 불가능하므로 평가 기준은 매년 약간씩 변동합니다. 심지어 이전 시험과 동일한 문제를 반복하더라도, 학생들은 두 번째에는 어느 정도 익숙해지기 때문에 시험의 난이도는 낮아지게 됩니다. 때때로 이전 시험 결과가 기대에 미치지 못할 경우, 시험의 평가 기준을 변경할 수 있으며, 필요한 변화의 정도에 따라 이미 개발 중인 시험도 수정될 수 있습니다. 이 경우에도 공정성을 위해 IB는 교사에게 변경 사항을 안내해야 하지만, 변경 정도에 따라 그 기간은 교육과정 변경에 비해 짧아질 수도 있습니다.

성과 기준(예: 성적 급간)은 평가 기준의 변경 사항과 균형을 맞추기 위해 매년 조정됩니다. 예를 들어 시험의 난이도가 높아져 학생들의 점수가 전반적으로 낮아졌다면, 이전 연도와 동일한 수준의 학생이 동일한 성적을 받을 수 있도록 성적 급간을 하향 조정합니다. 다른 두 기준 유형과 달리, 성과 기준의 경우 교사보다 주로 채점관이 이를 명확히 숙지해야 하며, 이를 통해 시험 기수 전반에서의 성과 기준을 유지할 수 있습니다. 성과 기준은 성적 급간을 설정하는 시점에서 조정될 수 있습니다. IB가 이러한 변경을 계획하고 있는 경우, 학교가 학생들의 예상 성적 수준을 관리하고 조정할 수 있도록 통상 사전에 이를 공지합니다. 이러한 경우 채점관이 새로운 기준이 무엇을 의미하는지 명확히 이해하는 것이 향후 기수에 적용하기 위한 핵심 단계입니다.

평가의 전반적인 기준을 설정함에 있어서는 이 세 가지 서로 다른 정의의 균형을 맞추는 것이 중요합니다. 간단한 자료에 매우 도전적인 문제를 설정하거나 간단한 문제 모음에 매우 높은 성과 기준을 설정하는 것이 가능하지만, 이런 경우 대개 구성개념 적합성 수준이 매우 낮아집니다. 예를 들어 시험에서 만점을 받아야만 7점의 성적을 받을 수 있도록 설정한다면, 이는 학생에게 극도로 높은 정확성을 요구하게 됩니다. 이러한 경우 과목에 대한 이해도가 매우 높지만 약간 부주의하거나 글쓰기 능력이 부족한 학생은 7점을 받기 어렵게 됩니다. 이는 과연 해당 평가가 의도한 목적에 부합하다고 할 수 있습니까?

기준의 정의는 **형성** 및 **총괄** 평가 모두에 적용됩니다. 그러나 형성 평가에서는 교사가 어떤 피드백을 제공할지 결정하는 과정에서 성과 기준을 설정하는 것이 보통 교사의 판단의 일부로 이루어집니다. 형성 평가에서는 특정 시험이 학생에게 적절하고 향후 교수에 도움이 되는 유의미한 정보를 제공할 수 있도록 교육과정 및 평가 기준을 훨씬 더 많이 고려하는 경향이 있습니다.

상대 평가 및 절대 평가

- ‘상대 평가’란 학생의 성과를 순위화하고 각 성적에 고정된 비율의 학생을 배정하는 방식을 의미합니다.
- ‘절대 평가’란 학생의 성과에서 무엇을 확인해야 하는지에 대한 설명에 따라 성과 기준을 설정하는 것을 의미합니다.
- IB는 성취 기준 평가(‘약한 절대 평가’로 알려짐)를 사용하며, 이는 기준의 균형과 이전 연도 학생 성과 비교를 통해 평가 기준을 설정하는 것을 의미합니다.

상대 평가와 절대 평가는 평가의 성과 기준을 설정하고 유지하는 서로 다른 두가지 방식입니다.

상대 평가

상대 평가는 대개 표준화 시험과 연관됩니다. 원리는 전형적인 학생 표본에 테스트를 시험적으로 실시한 후 그 결과(정의상 정규분포 또는 종형 곡선을 이뤄야 함)를 준거 척도로 삼아 이후 동일한 시험을 치르는 모든 학생의 점수를 산출하는 것입니다. 이렇게 초기 시험에서 점수의 표준 분포를 도출하는 과정은 ‘규준화’라고 부릅니다.

상대 평가의 기술적 설명은 반드시 모든 시험 결과 모음에 고정된 분포가 적용됨을 의미하지는 않으며 고정된 분포는 기존의 규준화 단계에서만 사용됩니다. 이후 학생들의 점수 분포는 이러한 정규분포와 다를 수 있습니다.

실제로 상대 평가는 보통 학생의 성과에 따라 순위가 부여되고 각 성적을 받는 학생의 비율을 고정하는 방식을 가리킵니다. 상위 15%의 학생에게 가장 높은 성적을 부여하는 경우를 예로 들 수 있습니다.

절대 평가

절대 평가는 글레이저(Glaser, 1963)에 의해 처음 제시되었습니다. 이는 성과 기준 설정에 중대한 변화를 나타내며, ‘명확히 정의된 행동적 영역에 관하여’ 학생의 성과를 측정하는 것에 강조합니다(Popham, 1978).

절대 평가에서는 학생의 성과를 각 성적에서 기대되는 것에 대한 사전 정의된 설명과 대조하여 비교합니다. 이는 일반적으로 과목이나 평가 전문가의 전문적인 판단에 의하여 수행됩니다.

이 접근 방법의 한계 중 하나는 모호하지 않고 모든 전문가가 동일하게 이해할 수 있는 명확한 설명을 만드는 것이 매우 어렵다는 점입니다. 실제로 ‘아무리 정교하게 표현해도 모호하지 않은 해석을 허용하는 기준은 없다’고 주장되어 왔습니다(Wiliam, 1993). 전통적인 절대 평가 결과는 관련 영역에 대한 숙달이 입증되었는지 여부로 판단되는 경우가 많습니다.

실제로 두 접근 방법 모두 심각한 단점이 존재합니다. 엄격한 상대 평가는 현재 시험이 이전 시험과 동일한 난이도라는 강력한 증거를 요구하는 반면, 절대 평가는 전문가 판단이 문제의 난이도를 정확히 고려하지 않는 곳 앤 크레스웰 효과(Good and Cresswell effect, 1988)의 영향을 받습니다.

성취 기준 평가

IB는 기준에 기반하면서도 곳 앤 크레스웰 효과(Good and Cresswell effect)(1988)의 근거를 인정하는 접근 방법인 ‘성취 기준 평가’(Newton, 2011)를 사용하며, 이는 과거 약한 절대 평가(Baird et al., 2000)로 불렸습니다. IB에서의 실제 적용에서는 전문 채점관이 기준(성적 기준표)에 기반하여 성적 급간이 위치할 수 있는 좁은 범위를 설정하고, 이를 전년도의 성적과 일치하도록 산출된 성적 급간과 비교합니다. 이 두 급간이 일치하는 지점에서 성적이 설정되며, 서로 다를 경우 상충하는 근거를 어떻게 일치시킬 수 있는지 추가 논의가 이루어집니다.

마지막으로 절대 평가와 상대 평가는 출제되는 문제 모음의 유형보다는 학생 답안의 분석과 해석에서 더 많이 다르다는 점을 유념하는 것이 중요합니다.

기준 유지하기

- 적절한 기준이 설정되고 나면 IB는 기준이 매년 일관되게 적용되는지 확인해야 합니다.
- 교육과정 기준은 교육과정 검토에서 다시 고려됩니다.
- 평가 기준과 성과 기준은 전문적인 판단과 통계적인 근거를 종합하여 유지됩니다.

비교 가능성은 타당성의 필수적인 측면입니다. 기준을 유지한다는 맥락에서 이는 IB가 매년 동일한 기준이 적용되도록 보장해야 함을 의미합니다. IB 기준은 점수가 아닌 성적의 의미에 기반합니다. 점수와 성적의 차이에 대한 설명은 본 간행물의 ‘**점수와 성적의 차이점**’ 섹션에서 확인할 수 있습니다.

교육과정 기준은 모든 기수 간 동일하게 적용되기 때문에 가장 유지하기 쉬운 기준입니다. 평가는 전체 교육과정을 충실히 반영해야 하며, 이는 시험 출제 과정에서 모니터링됩니다. 그러나 외부 요인으로 인해 교육과정 기준이 변경될 수도 있습니다. 대표적인 예는 바로 컴퓨터 활용 기능입니다. 기기가 발전하고 일상생활의 일부가 되면서, 한때 전문적이고 어렵다고 여겨졌던 지식과 이해가 이제는 보편화되었습니다. 이로 인해 내용 자체는 동일하더라도, 해당 주제에 대한 교육과정 기준은 변화하게 됩니다.

이러한 문제를 해결하고 내용을 최신 상태로 유지하기 위해 IB는 교육과정 검토 주기를 운영합니다. 검토 결과, 교육과정 기준이 변경될 경우 IB는 학생에게 불이익이 발생하지 않도록 평가 기준과 성과 기준의 균형을 조정해야 합니다.

동일한 수준의 난이도를 가진 시험은 존재할 수 없기 때문에 IB 시험의 **평가 기준**은 기수마다 달라집니다. 일부 교육 시스템에서는 각 문제의 난이도를 확립하기 위해 광범위한 사전 시험을 실시하며, 학생이 경험하는 난이도에 대해 높은 수준의 확신을 확보하도록 시험을 신중하게 구성합니다. 그러나 IB는 시험 문제가 사전에 유출될 위험이 있기 때문에 사전 시험을 시행하지 않습니다. 따라서 IB는 경험 많은 출제자와 검토관의 전문적 판단에 의존하여 일관되고 균형 잡힌 시험을 구성합니다.

또한 IB는 학생의 답안을 바탕으로 **성과 기준**을 조정하여 전체 기준이 유지되도록 합니다. 성과 기준을 유지하는 두 가지 주요 방법은 이전 섹션에서 다루었으며, 성과 기준을 유지하는 것이 IB의 성적 부여 과정의 목적입니다.

총괄 평가에서 학생의 성취 설명하기

- 성적은 학생의 성취를 매우 단순화한 관점으로 나타내지만, 대학, 고용주, 학교와 같은 이해관계자들이 선발 과정에서 합리적인 판단을 내릴 수 있게 합니다.
- 만약 보다 복합적이고 전인적인 정보만 제공된다면 이해관계자들은 의미 있는 선발 결정을 내리기 위해 이러한 증거를 단순화해야 하며, 이 과정에서 IB가 성적 급간을 설정할 때만큼 세심한 주의를 기울이지 않을 수 있습니다.
- IB 평가는 학생에게 유의미한 결과를 제공하기 위해 전문적 판단의 활용이 중요합니다. 동시에 IB는 공정성을 보장하기 위해 이러한 판단이 객관적이고 편향되지 않은 증거로 뒷받침되어야 한다는 점을 인식하고 있습니다.

성적의 영향

훌륭한 요리사가 보유한 지식, 기능, 경험을 생각해봅시다. 이 요리사는 완벽하게 균형 잡힌 요리를 만들기 위한 재료와 맛의 조합에 관한 지식을 가지고 있습니다. 또한 재료를 선택하고 준비하는 기능, 조리 기능, 플레이팅과 연출 기능도 갖추고 있습니다. 더 나아가, 기법을 활용해 완성도를 높이고, 음식이 완벽하게 조리된 때를 판단하며, 여러 재료와 요리의 조합을 통해 식탁에서 최고의 만족을 이끌어낸 경험도 가지고 있습니다.

이제 이 모든 복합성을 7점 만점의 단일 성적으로 압축하여 누가 최고의 요리사인지를 결정해봅시다. 이때 결과는 거의 무의미합니다. 척도를 100점, 심지어 1,000점까지 늘린다 하더라도 상황은 크게 달라지지 않습니다. 구분의 폭은 넓어질 수 있으나, 서로 다른 기능을 비교해야 한다는 근본적인 문제는 여전히 남아 있습니다.

이것이 바로 평가가 직면하고 있는 문제입니다. 학생의 지식, 이해, 능력의 복합성을 단일 결과로 가장 잘 나타낼 수 방법은 무엇입니까? 설령 학생에 관한 모든 정보를 정확히 포착하여 요약할 수 있다 하더라도, IB는 여전히 이들의 역량을 완벽히 반영하는 성적을 부여하는 것은 불가능합니다.

이에 대한 대안으로 종종 제시되는 방식은 아예 성적을 부여하지 않는 것입니다. 대신, IB가 각 학생에게 잘한 점과 부족한 점을 개별적으로 설명해 줄 수도 있습니다. 이러한 접근 방법은 바람직한 학습 및 교수 원칙과는 부합하지만, 학생 간 비교가 매우 어렵다는 심각한 한계를 지닙니다.

IB 평가의 목적은 학생이 향후 학습이나 직업 진로를 위해 자신의 성취와 성장을 증명할 수 있도록 지원하는 데 있습니다. 이는 수여 기관(주로 DP와 CP의 경우 대학)이 어떤 학생들에게 이러한 기회를 제공할 수 있는지 결정하기 위해 일종의 선발 과정이 필요함을 의미합니다.

분명하게도 다른 기준들이 시험보다 신뢰도가 낮다면, 그러한 기준에 더 크게 의존할수록 선발 결정의 신뢰성은 더욱 낮아질 것이다.

(Cresswell, 1986, p. 42)

선별을 해야 한다면 IB는 그 결정을 최대한 공정하고 의미 있게 만들어 학생을 지원해야 할 책임이 있습니다. IB가 수여 기관에 학생에 대한 서술적 설명만 제공한다면 그들은 학생을 비교할 다른 방법을 찾아야 할 것이며, 이는 성적이 부여되는 시험 결과보다 신뢰성과 비교 가능성이 현저히 떨어질 것입니다.

이는 총괄 평가가 완벽하거나 최선의 방법이라는 뜻은 아니지만, 많은 대안들보다는 공정하다는 의미입니다. 또한 총괄 평가는 평균 평점, 교과 외 활동, 개인 에세이, 추천서 등 다른 관련 지표와 함께 사용될 때 의사 결정에 유용한 정보를 제공할 수 있습니다. 아마도 가장 중요한 것은, IB가 이러한 평가를 최대한 공정하고, 의미 있으며, 신뢰할 수 있게 만들기 위해 끊임없이 노력하고 있다는 점입니다.

[그림 15] 각 쌍의 두 학생을 구별해야 하는가?



[그림 15]의 예시에서 첫 번째 쌍의 학생들은 DP 성적 기준표에 제시된 일반적인 5점과 6점 기준으로 구별됩니다. 필요하다면 두 학생 사이에서 결정을 내리는 것이 아마도 공정할 것이며, 시험을 다시 치르더라도 동일한 결과를 얻을 가능성이 높습니다. 그러나 두 번째 쌍의 학생들에게는 그 차이가 의미 있다고 보기 어려우며, 두 학생은 아마도 동일한 능력을 가지고 있을 것입니다.

이 예시에서는 모든 차이가 의미 있는 차이가 아니라는 점, 그리고 성적은 IB가 어느 지점에서 학생 간 구별이 가능하다고 판단하는지를 보여 주는 지표임을 보여 줍니다. 물론 성적 급간의 경계선(기준의 바로 위 또는 아래)에 있는 학생의 경우 어느 성적도 타당하다고 볼 수 있습니다. 그러나 대다수의 학생에게 있어 성적의 차이는 전반적인 성과에서 의미 있는 차이를 나타냅니다.

이는 곧 성적을 몇 개의 구간으로 나누는 것이 적절인가라는 질문으로 이어집니다. 성적이(합격 또는 불합격) 2구간으로 나뉘었다면 대다수의 학생은 공정한 점수를 받게지만 경계에 있는 학생들에게는 그 결과의 영향이 매우 클 수 있습니다. 이와 반대로 성적이 20구간으로 나뉘었다면 훨씬 더 많은 학생이 경계 또는 그 부근에 위치하겠지만 잘못된 범주에 속했을 때 각 학생에게 미치는 영향은 상대적으로 작아집니다. 크레스웰(Cresswell, 1986)은 이러한 개념을 더욱 자세히 탐구했습니다.

IB는 일반적으로 **7구간의 성적**을 선택하여, 평가가 제공할 수 있는 의미 있는 범주의 수를 나타내며, 경계에 있는 학생 수와 잘못된 성적이 미치는 영향 간의 균형을 적절하게 이루고 있습니다.

전문적 판단의 중요성

IB 평가에서 중점적으로 다루는 복합적이고 고차원적인 사고 기능은 단순하고 판단이 배제된 채점 방식에 적합하지 않습니다. 학생의 답안은 매우 다양한 형태로 제시되며, 동등하게 타당한 올바른 답안이 여러 방식으로 존재할 수 있습니다. 연구에 따르면, 복합적인 지식과 기능은 작고 개별적인 구성 요소로 나누어 가르쳐서는 안 되며 이러한 원칙은 학생 답안을 채점할 때도 동일하게 적용됩니다. 따라서 채점 기준표를 개발할 때 IB 과목 전문가들은 모든 학생을 일관되게 채점할 수 있도록 엄격한 지침과 이를 달성하는 방법을 제시해야 합니다.

이는 IB가 채점관의 전문적인 판단, 특히 채점 기준을 설정하고 설명하는 선임 채점관의 전문적인 경험과 판단에 큰 비중을 두어야 함을 의미합니다. 이 모든 요소는 평가 시스템의 신뢰성에 큰 도전을 의미하지만 결과의 정확성이라는 맥락에서 반드시 극복해야 할 도전입니다.

IB 평가 채점하기

- 채점이란 학생이 부여받은 과제를 얼마나 잘 수행했는지를 평가하는 과정입니다.
- 채점은 답안의 개별 요소에 초점을 맞출 수도 있고, 보다 전체적이고 포괄적인 판단을 내릴 수도 있습니다.
- 형성 평가의 경우, 채점 결과가 수치적 점수가 아닌 순수하게 서술형으로 제공될 수도 있습니다.

채점 정의하기

이 섹션에서 ‘채점’ 및 ‘채점 방법’은 MYP, DP, CP 총괄 평가에 대해 IB가 검증하는 채점을 의미합니다. 형성 평가는 모든 프로그램에 걸쳐 사용되지만, 대부분 교실에서 이루어지며 IB는 이를 채점관의 채점을 통해 검증하지 않습니다.

채점은 학생이 얼마나 과제를 잘 수행했는지 설명하지 않습니다. 학생의 성취 수준은 문제의 난이도, 각 수준별 예상 합격 점수 등 여러 요인에 따라 달라집니다. 채점은 학생의 답안을 이상적인 모범 답안과 비교하는 것입니다.

즉, 채점은 과제의 본질과 평가 도구가 적절하다고 판단하는 채점 유형에 따라 다양한 방식으로 수행될 수 있습니다.

채점 방법

IB는 다양한 평가 도구를 사용하며, 각 평가 도구가 요구하는 특성에 따라 적절한 채점 방식이 결정됩니다.

학생의 답이 맞거나 틀리거나 둘 중 하나인 것처럼 때때로 채점은 매우 **객관적**입니다. 이는 몇 가지의 키워드만 요구되거나 여러 선택지 중에서 하나를 고르는 형태의 문제에서 나타납니다. 자동 채점은 객관적으로 정답 또는 오답이 명확한 사전 정의된 채점 기준표에 따라 테크놀로지를 활용하여 학생의 과제를 평가하는 과정입니다. IB는 정답이 명확하고 채점관의 판단 없이도 100% 정답이라 확신할 수 있는 선다형 문제에 자동 채점을 활용합니다. IB는 기대되는 답안의 범위가 좁더라도 판단이 필요한 경우에는 사람을 대신하여 인공지능 기반 채점 방식을 사용하지는 않습니다.

그 밖의 경우 채점은 훨씬 더 **주관적**입니다. 채점관은 학생의 답안이 수용 가능한 수준인지 또는 ‘점수 구간표’의 설명 중 어떤 것이 ‘완벽한’ 답안과 비교했을 때 학생의 답안을 가장 잘 설명하는지 전체적으로 판단해야 합니다. IB에서 사용하는 점수 구간표 유형의 예시를 <표 5>에서 확인할 수 있습니다.

<표 5> IB 점수 구간표 예시

점수	단계 기술어
0	아래 기술어에서 명시한 기준을 충족하지 못합니다.
1~3	<ul style="list-style-type: none"> 이슈에 대한 이해를 거의 보여주지 못합니다. 과목별 어휘를 사용하지 않았거나 지속적으로 부적절하게 사용합니다. 이슈에 대한 설명이 매우 부족합니다. 논점이 피상적이며 자주 불분명합니다. 답변이 설명 위주입니다. 제시한 분석이 피상적이거나 일관되지 않습니다. 다양한 관점에 대한 논의가 거의 없거나 전혀 없습니다. 결론이 포함된 경우, 매우 피상적이거나 답변의 나머지 부분과 일관되지 않습니다.
4~6	<ul style="list-style-type: none"> 이슈에 대한 기본적인 이해를 보여줍니다. 과목별 어휘를 사용하지만 대개 부적절하게 사용합니다. 이슈에 대한 설명은 기본 수준이며 충분히 전개되지 않았습니다. 논점이 대개 부정확하거나 모호하며, 무엇을 전달하려는지 대개 불분명합니다. 제시한 분석이 제한적이며 답변이 전반적으로 분석보다는 설명에 치중되어 있습니다. 다양한 관점에 대한 논의가 제한적입니다. 단순한 결론이 포함되어 있습니다.
7~9	<ul style="list-style-type: none"> 이슈에 대한 이해를 어느 정도 보여줍니다. 과목별 어휘를 때때로 적절하게 사용합니다. 특정 이슈에 대한 설명이 만족스러운 수준이지만, 부분적으로 명확성과 전개가 부족합니다. 적절한 논점을 제시하지만 정확성과 상세 설명이 부족합니다. 답변에 분석이 포함되지만 전개가 부족합니다. 다양한 관점에 대해 어느 정도 논합니다. 결론이 포함되어 있습니다.
10~12	<ul style="list-style-type: none"> 이슈에 대해 우수한 이해를 보여줍니다. 과목별 어휘를 대부분 적절하게 사용합니다. 이슈에 대한 설명이 명확하지만 추가 전개가 필요합니다. 논점이 적절하고 정확하지만 상세 설명이 부족합니다. 답변에 비판적 분석이 포함되어 있지만 전개가 부족합니다. 다양한 관점에 대해 논합니다. 답변은 제시된 주장과 일관된 결론으로 이어집니다.
13~15	<ul style="list-style-type: none"> 이슈에 대해 매우 우수한 이해를 보여줍니다. 과목별 어휘를 정확하고 정밀하게 사용합니다. 이슈에 대한 설명이 명확하고 잘 전개되었습니다. 논점이 적절하고 정확하며 상세합니다. 답변에 잘 전개된 비판적 분석이 포함되어 있습니다. 다양한 관점에 대해 비판적으로 논합니다. 답변은 제시한 주장과 일관되며 명확하게 제시된 합리적 결론으로 이어집니다.

고려해야 할 또 다른 사항은 한 가지 과제에 대해 **기준**이라 불리는 여러 측면을 각각 분리하여 채점하는지 여부입니다. 예를 들어 에세이를 문법의 질, 주요 사실의 정확성, 에세이 구조, 결론의 질이라는 네 가지 개별 기준에 따라 평가될 수 있습니다. 이러한 접근 방법에서는 평가 기준이 서로 독립적으로 작용하는 것이 바람직합니다. 학생이 동일한 답변으로 여러 평가 항목에서 점수를 획득하는 것은 공정하지 않기 때문입니다.

기준 채점의 반대는 **전반적인 인상 판단**입니다. 이 방식에서는 채점관이 이상적인 답변의 여러 측면을 내부적으로 균형을 맞추고, 과제 전체를 종합적으로 반영하는 최종 판단을 내립니다.

신뢰성 있는 채점을 어렵게 하는 주요 요인은 과제의 성격과 학생 답변의 예상 범위입니다. 다양한 답변이 예상되는 광범위한 과제는 제한된 범위의 답변만 나오는 범위보다 신뢰성 있게 채점하기가 더 어렵습니다.

다양한 채점 접근 방법에 대한 자세한 내용은 ‘[채점하기](#)’ 섹션을 참고하십시오.

형성 평가 채점하기

모든 IB 프로그램은 형성 평가를 사용합니다. 다만 초등교육 프로그램(PYP)에서는 형성 평가가 **유일한** 평가 방식입니다. 형성 평가에서의 채점은 반드시 수치적일 필요가 없습니다. 한 과제를 ‘완벽한’ 답안과 비교하고, 두 답안 간의 유사점과 차이점을 중심으로 서술적인 피드백을 제공할 수 있습니다. 이는 학생이 ‘우수함’의 기준을 명확히 이해할 수 있도록 돕는 공통 성취 기준의 형태로 제시될 수도 있습니다. 이 때문에 두 답안의 우수성을 비교하기 어려울 수 있지만, 형성 평가의 목적은 학습을 지원하기 위한 피드백을 제공하는 데 있습니다.

교실에서는 교사가 형성 평가를 활용하여 학생들의 지식, 이해 또는 기능에서의 격차를 파악할 수 있습니다. 이러한 상황에서는 신뢰성의 중요도는 상대적으로 낮습니다. 점수나 성적을 부여할 필요가 전혀 없습니다. 평가 결과는 학생이 학습해야 할 주제나 기능 목록일 수 있습니다. 이러한 상황에서는 학생에게 학습 과정에서 중요한 요소, 혹은 학생이 자신의 배움을 드러낼 수 있는 요소에 대한 피드백이 필요하기 때문에 구성개념 적합성이 중요합니다.

윤리적 사고방식 기르기

윤리적 사고방식 정의하기

윤리적 사고방식이란 **정직성, 존중, 진정성**의 원칙에 따라 **책임감 있고 진실되게** 행동하려는 의지를 말합니다. 교육적 맥락에서 윤리적 사고방식은 학문적 진실성을 함양하며, 정당하고 정직한 결과물을 만들어내는 것의 중요성을 강조합니다. 이 사고방식은 단순히 규정과 정책을 넘어 윤리적 의사결정과 행동을 학습 공동체의 문화 속에 자리 잡게 합니다. 이는 학생이 학습 과정을 가치 있게 여기고, 원칙에 기반한 선택을 하며, 공정한 학습 환경을 조성하는 데 기여하도록 장려합니다.

윤리적 사고방식을 기르는 이유

윤리적 사고방식을 기르는 것은 교육적 성과의 신뢰성과 타당성을 보장하는 데 필수적입니다. 이는 평가의 진실성을 보호함으로써 학생들의 성취가 실제 능력을 반영할 수 있도록 합니다. 더욱이, 윤리적 행동을 함양하는 것은 고등 교육과 직업 환경에서의 성공에 필수적인 가치를 심어주어 학생들이 교실 밖의 삶에 대비하게 합니다. 윤리적 원칙을 무시되면 학교 공동체의 신뢰가 흔들리고 기관의 신뢰성이 위태로워집니다. 윤리적 사고방식을 함양함으로써 학교는 학생이 삶의 모든 영역에서 공정성과 존중을 실천하는 원칙적인 세계 시민으로 성장할 수 있도록 힘을 실어줍니다.

윤리적 사고방식은 어떻게 형성되는가?

윤리적 사고방식을 기르는 데는 전인적인 공동체 차원의 접근이 필요합니다. 학교는 학생의 발달 단계에 맞춘 명확한 기대 사항을 전달하는 것에서 시작해 학문적 진실성을 학교 문화에 깊이 자리 잡게 합니다. 교사들은 윤리적 행동의 본보기를 보이고, 예방 전략을 설계하며, 진실성에 관한 논의를 교육과정에 통합하여 중추적인 역할을 수행합니다. 코디네이터와 학교 지도자는 반드시 정책이 명확히 정의되고 일관되게 적용되어 모든 이해관계자를 대상으로 한 교육을 통해 이를 뒷받침해야 합니다. 학부모 및 법적 보호자와의 열린 소통은 서로 기대하는 바를 일치시키고 가정 내 윤리적 실천의 중요성을 강화하는 데 기여합니다.

정기 워크숍, 윤리적 과제에 대한 분명한 예시, 진실성이 실제 상황에서 어떻게 적용되는지에 대한 대화 촉진은 이러한 원칙을 더욱 굳건하게 만듭니다. 처벌 초점을 두기보다 정직의 이점을 강조함으로써 학교는 학생이 윤리적 의사결정에 필요한 기능과 이해를 기르는 데 지원을 아끼지 않는 환경을 조성하게 됩니다.

윤리적 사고방식은 단순한 목표가 아니라, 공동의 책임, 일관된 메시지, 학교 공동체 모두의 변함없는 헌신을 통해 달성되는 지속적인 실천입니다. 자세한 내용은 인쇄 가능한 자료인 ‘**윤리적 사고방식 기르기**’를 참고하십시오.

평가에서 인공지능의 윤리적 사용

평가와 인공지능(AI)의 상호작용을 논의할 때 고려해야 할 두 가지 주요 요점이 있습니다.

- AI가 평가되는 대상에 어떤 영향을 미치는가?
- AI는 학생에 대한 판단을 내리는 과정에서 어떤 역할을 하는가?

AI는 이미 일상의 일부가 되었으며, 이에 책임 있게 대응하는 것은 필수적입니다. 특히 IB 학생들이 이 테크놀로지를 올바르게 이해하고 윤리적으로 활용할 수 있도록 돕는 것이 중요합니다.

“내가 수행한 과제라는 말은 무엇을 의미하는가?”

이는 AI가 평가 대상에 미치는 영향이라는 첫 번째 도전의 근본적인 핵심 질문입니다. 교육에서 AI에 대한 초기 우려 중 하나는 학생들이 AI에게 자신의 과제를 대신 수행할 것을 요청함으로써 ‘부정행위’를 저지를 수 있다는 점이었습니다. 특정 평가에서 AI가 학생보다 더 나은 성과를 낼 수 있다는 다양한 주장도 제기되었습니다. 이러한 우려는 평가의 가장 근본적인 질문으로 이어집니다. 특정 질문의 목적은 무엇입니까? 학생에 대해 어떤 증거를 수집하고자 합니까? 역사적 사건의 날짜와 같은 사실 정보는 인터넷을 통해 몇 초 만에 확인할 수 있지만, 비행기 착륙처럼 시간이 매우 중요한 상황에서는 인터넷 검색이 적절하지 않을 수 있습니다.

이는 AI의 다른 용도도 마찬가지입니다. 때로는 표현 방식보다 내용 자체가 핵심인 경우가 있습니다. 이러한 경우 의사소통 방식에 대한 점수는 필요하지 않을 수 있으며, AI가 학생의 아이디어를 보다 가독성 있는 형태로 제시하는 데 도움을 줄 수도 있습니다. 이와 반대로 시를 창작할 때 각 단어의 선택이 매우 중요합니다. 학생들은 시의 최종 결과물 중 어떤 부분도 AI로 작성해서는 안 되지만, 예를 들어 유의어 사전처럼 활용할 수는 있습니다.

AI의 보편화로 제기된 핵심 교육 질문은 ‘학생들이 일상에서 AI를 책임감 있고 윤리적으로 사용하려면 무엇을 학습해야 하는가?’입니다. 이 중요한 질문은 시간이 지나며 계속해서 발전해 나갈 것입니다. 한편 이 질문에 대한 답변에는 학생에게 다음과 같은 교육을 포함합니다.

- 정보를 비판적으로 수용하고 잠재적 편향이나 오도된 결론을 인식할 수 있어야 합니다.
- 출처와 참고 문헌을 확인하여 정보가 정확하게 사용되고 있는지 점검할 수 있어야 합니다.
- 무엇이 자신의 작업이며 무엇이 그렇지 않은지를 이해하고, 이를 정직하게 제시할 수 있어야 합니다.

“안녕하세요, 이번 시험의 AI 평가 어시스턴트입니다.”

AI는 학생이 평가받는 방식을 점차 변화시킬 것입니다. 디지털 평가로의 전환은 훨씬 더 상호작용적인 과제를 가능하게 하며, AI의 도입은 이를 한층 더 발전시킬 수 있습니다. 예를 들어 학생은 디지털 아바타에게 자신의 아이디어를 설명하고, 이들과 상호작용할 수도 있을 것입니다. 이러한 테크놀로지는 평가에 대한 다양한 접근 방법을 가능하게 할 수 있습니다. 예를 들어, 현재는 인간 채점관과의 상호작용이 필요하여 관리가 어려운 인터뷰를 더 적극적으로 활용할 수 있습니다.

객관식 답변(예: 선다형 문제 또는 끌어내기 문제)을 자동으로 채점하는 방식은 이미 오랫동안 평가 도구로 사용되어 왔지만, 머신러닝과 AI의 발전은 보다 주관적인 문제와 복잡한 답변에도 자동 채점을 적용할 수 있는 가능성을 넓히고 있습니다. 가까운 미래에도 위험 부담이 큰 최종 점수는 여전히 인간 채점관이 결정해야 하지만, AI는 채점의 질을 향상시키고 두 번째 채점관이 검토해야 할 과제를 식별하는데 큰 잠재력을 지니고 있습니다.

AI 채점은 교실 수업의 일부인 형성 평가를 지원할 수 있으며, 학생의 이전 답변을 바탕으로 각 학생의 능력에 맞게 평가 문제를 조정할 수도 있을 것입니다. 이러한 접근 방법은 과거에는 선다형 문제에 사용되었지만 AI의 발전으로 다양한 문제 유형에 확대되어 학생을 위한 개인 맞춤형 시험을 제공할 수 있게 되었습니다.

새로운 테크놀로지의 영향을 고려할 때 가장 중요한 요인은 평가 목적이 의도한 학습 결과와 계속해서 관련성을 유지하는지, 그리고 평가 자체가 그 목적에 대해 타당성을 유지하도록 하는 것입니다.

학생이 학습을 지원하기 위해 AI를 사용하는 것은 도움이 될 수 있지만, 학습을 대신하여 AI를 사용하거나 자신의 것이 아닌 과제를 제출하는 것은 허용되지 않는다는 점을 기억해야 합니다.

추가 문헌

- 인공지능(AI) 도구 사용 설명을 위한 핵심 요점」(한 페이지 자료 시트)
- 학생 과제물에서의 인공지능(AI) 활용에 관한 13가지 시나리오 평가(전문)

수행 정의하기

- 본 섹션에서 설명하는 수행은 중등교육 프로그램(MYP), 디플로마 프로그램(DP), 직업연계 프로그램(CP)의 총괄 평가에 적용됩니다.
- 국제 바칼로레아(IB)가 총괄 평가를 제공하지 않는 초등교육 프로그램(PYP)이나 학교에서 수행하는 모든 형성 평가에는 적용되지 않습니다.

이 섹션에서는 IB가 외부 채점 또는 조정이 이루어지는 평가에 참여한 학생들의 성과를 산출하기 위해 사용하는 수행 방식을 개략적으로 설명합니다. 이는 PYP의 일부로 이루어질 수 있는 교사가 구성한 평가와 같이 IB가 수행하지 않는 다른 평가에는 적용되지 않습니다.

원칙은 우리가 무엇을 하는지에 대한 **이유**를 설명하며 **수행**은 그 것을 어떻게 하는지 **방법**을 설명합니다. 이 섹션에서는 IB가 평가 결과의 타당성을 확보하기 위해 사용하는 높은 수준의 수행을 설명합니다.

그 다음 단계의 세부 내용은 각 수행을 실행하는 개별 단계를 설명하는 IB의 **절차**입니다. 학교와 관련된 이러한 절차의 상세 정보는 프로그램 관련 「평가 절차」 간행물에서 확인할 수 있습니다.

IB 평가가 측정하는 것과 사전 학습의 역할

총괄 평가의 목적은 학생이 MYP, DP, CP에서 학습을 마칠 때 **지식, 이해, 기능** 측면에서 보이는 성과를 측정하는 데 있습니다. 총괄 평가에는 다음을 포함한 여러 시사점이 존재합니다.

- 총괄 평가 결과는 특정 시점에서의 학생 성과를 반영합니다. 총괄 평가는 학생의 잠재력이나 학습 진척도를 측정하지 않습니다.
- IB 평가는 학생이 특정 문제나 특정 날짜에 저조한 성과를 보임으로써 발생할 수 있는 평가 부정확성을 최소화하도록 설계되었습니다. 이는 일반적으로 학생에게 자신의 실력을 보여줄 여러 기회를 제공하기 위해 다수의 시험을 시행함으로써 달성합니다. 다만 시험 횟수는 관리 가능한 수준이어야 하며, 학생에게 과도한 부담을 주어서 안 됩니다.
- 총괄 평가는 학생 집단이 아닌 개별 학생의 지식, 이해, 기능을 반영합니다.
- IB는 평가 성적을 부여할 때 사전 학습을 고려하지 않습니다. 이는 해당 IB 프로그램을 시작하기 전 학생이 취득한 자격, 등급 또는 성취를 고려하지 않는다는 것을 의미합니다. IB 교과 가이드는 사전 학습이 과목 또는 학습 영역과 어떤 관련이 있는지 설명합니다.

학생 성취도 보고하기

- IB 교육 목표의 초점은 더 나은 세상을 만들어 갈 수 있는 청소년을 양성하는 데 있으며, 성공적인 평가는 이러한 목표를 지원해야 합니다.
- IB 성적에는 의미가 있으며, 성적 급간은 이러한 의미를 고려하여 설정됩니다.
- 성적은 학생의 성취를 매우 단순화한 관점으로 나타내지만, 대학, 고용주, 학교와 같은 이해관계자들이 선발 과정에서 합리적인 판단을 내릴 수 있게 합니다.
- 보다 복잡적이고 전인적인 정보가 제공된다면 의미 있는 선발 결정을 위해 그 근거를 단순화하는 책임은 다른 이들에게 전가되며 이들은 IB가 성적 급간을 설정할 때만큼 신중하게 이 과정을 수행하지 않을 수 있습니다.
- 학생의 성취는 시험 결과만을 의미하는 것이 아니며, 성적을 다룰 때도 누군가에게는 실망스러운 결과가 다른 누군가에게는 큰 성취일 수도 있습니다.

IB의 목표는 서로 다른 문화를 이해하고 존중하며, 더 나은 평화로운 세상을 실현하는 데 기여할 수 있는, 지식이 풍부하고 탐구심과 배려심이 많은 청소년을 기르는 것입니다.

(IB 교육 목표 2024)

IB 성적의 의미

IB 평가의 결과는 성적입니다. 이러한 성적들은 학생이 답안에서 보여준 작업의 수준을 나타냅니다.

IB는 각 성적에 대한 설명을 제공합니다. 이러한 설명들은 MYP, DP, CP에 따라 다르며, 이는 IB가 연령대별 학생에게 기대하는 과제 수행 수준을 반영하기 때문입니다.

[그림 16] DP 성적 기준표 예시

교과군 3 (개인과 사회) 성적 기준표

개인과 사회

7점

비판적 사고 기능에 나타나는 개념적 인식, 통찰, 지식 및 이해력을 보여줍니다. 전개가 훌륭하고, 논리적이며 일관적인 구조를 갖추고, 적절한 예시로 뒷받침된 답변을 제시할 수 있는 높은 수준의 능력을 보여줍니다. 과목에서 사용되는 용어를 정확히 사용합니다. 과목과 관련된 문헌 자료에 익숙합니다. 증거를 분석 및 평가하고 지식과 개념을 종합하는 능력을 갖추었습니다. 대안적인 관점과 주관적이며 이념적 편향을 인식하고, (바뀔 수도 있지만) 합리적인 결론을 내릴 수 있는 능력을 갖추었습니다. 비판적인 성찰적 사고력을 일관적으로 보여줍니다. 데이터를 분석 및 평가하거나 문제를 해결하는 매우 높은 수준의 능력을 보여줍니다.

6점

상세한 지식과 이해력을 보여줍니다. 일관적이며, 논리적인 구조를 갖추고, 우수한 전개를 보여주는 답변을 제시합니다. 지속적으로 적절한 용어를 사용합니다. 지식과 개념을 분석, 평가 및 종합하는 능력을 갖추었습니다. 관련 연구, 이론, 이슈에 관한 지식을 보여주고 다른 관점들과 이러한 관점들이 형성된 맥락을 인식합니다. 비판적 사고력을 지속적으로 보여줍니다. 데이터를 분석 및 평가하거나 문제를 해결하는 높은 수준의 능력을 보여줍니다.

5점

과목 관련 용어를 사용해 과목에 대한 탄탄한 지식과 이해력을 보여줍니다. 답변이 논리적인 구조를 갖추고 있고 일관적이지만, 전개가 잘 되지 않는 않습니다. 지식과 개념을 통합하려는 시도가 일부 보이며 능숙한 답변을 제시할 수 있는 능력을 갖추었습니다. 평가적이기보다는 설명 위주입니다(하지만 상충되는 관점을 제시하고 전개하는 능력이 일부 보입니다). 비판적 사고력을 일부 보여줍니다. 데이터를 분석하고 평가하거나 문제를 해결하는 능력을 보여줍니다.

이 일반적인 성적 기준표는 교과군의 모든 과목에서 동일해야 하지만 IB는 각 기준표가 실제로 무엇을 의미하는지 더 쉽게 이해할 수 있도록 종종 과목별 맥락에 맞춰 설명합니다. 과목 맥락으로 인해 기준 자체가 변경된 것이 아님을 이해하는 것이 중요합니다. 언어에서 4점은 과학에서 4점 또는 예술 과목에서 4점과 동일한 의미를 가져야 합니다. 이는 모든 성적을 동등하게 간주하는 IB의 프로그램 접근 방법에 내재되어 있습니다. 그러나 서로 다른 두 과목의 성과를 비교하는 개념이 타당한지에 대해 교육학자들 사이에서 많은 논의가 있어 왔습니다. 서로 다른 두 과목의 성과를 어떻게 비교할 수 있습니까? 시도하는 것 자체가 의미가 있습니까? 이 개념은 ‘비교 가능성’섹션에서 더 자세히 다루고 있습니다.

[그림 17] 두 과제를 어떻게 비교할 수 있는가?



9. (a) $x = e^{-3y+1}$

양쪽의 자연로그를 가져옴

$$(f^{-1}(x)) = \frac{1}{3} (\ln x - 1)$$

(b) Q의 좌표는 (1,0)임

$$\frac{dy}{dx} = \frac{1}{x}$$

at Q, $\frac{dy}{dx}$

$$y = x - 1$$

(c) 필요 구역을 A로 두고

$$A = \int_1^e 1 dx - \int_1^e \ln x dx$$

부분적분법을 사용하여 다음을 찾음

$$= \left[\frac{x^2}{2} - x \right]_1^e - [x \ln x - x]_1^e$$

$$= \frac{e^2}{2} - e - \frac{1}{2} \left(\frac{e^2 - 2e - 1}{2} \right)$$

타당성에 대한 고려의 일환으로 서로 다른 과목 간 IB 성적 비교 가능성에 대한 논거는 IB 성적의 목적에 기반합니다. IB 성적은 이해관계자가 학생의 성취 수준을 비교할 수 있도록 하기 위해 마련되었습니다. 따라서 성적이 지니는 의미의 동등성을 위해 통계적이고 정성적인 방법을 사용하는 것은 의미가 있습니다.

점수와 성적의 차이점

점수와 성적은 다릅니다.

[그림 18] 점수와 성적을 비교하기 위한 비유



‘우수한’ 응시생이 대부분의 과제를 완수했을 것으로 기대합니다.



‘우수한’ 응시생이 과제의 일부만 완수했을 것으로 기대합니다.

점수와 성적의 차이를 설명하는 데 활용할 수 있는 여러 비유가 있습니다. 예를 들어 [그림 18]에서 걸은 거리는 누군가가 얼마나 멀리 이동했는지를 나타내는 일반적인 측정치로서 점수를 의미할 수 있습니다. 그러나 특정 거리를 걷는 것이 얼마나 큰 성취인지 이해하려면 그 사람이 어떤 환경에서 걷고 있었는지를 고려해야 하며, 이는 성적을 설정할 때 반영됩니다.

- 점수는 학생이 과제를 **얼마나** 완수했는지 나타냅니다.
- 성적은 학생의 점수가 나타내는 성취 수준을 정량화하기 위해 과제가 **얼마나 어려운지**를 고려합니다.

[그림 19]에 제시된 두 가지 예시를 고려해 봅시다. 첫 번째 예시의 경우 16세의 학생이 ‘우수한’ 성적을 받으려면 과제의 모든 부분을 정확하고 올바르게 수행해야 합니다. 두 번째 예시에서는 ‘우수한’ 성적을 받기 위해 요구되는 수준이 훨씬 낮습니다.

[그림 19] 다양한 난이도의 평가 과제



‘우수한’ 응시생이 대부분의 과제를 완수했을 것으로 기대합니다.

‘우수한’ 응시생이 과제의 일부만 완수했을 것으로 기대합니다.

이러한 고려 사항은 양질의 평가를 설정하는 데 있어 직면하는 과제 중 하나로 이어집니다. 학생은 자신의 잠재력을 충분히 발휘할 기회가 필요하며, 과제가 너무 단순할 경우 그러한 기회를 얻을 수 없습니다. 이와 반대로 과제가 너무 어려우면 많은 학생이 과제를 시작조차 할 수 없는 상황이 발생할 수 있습니다. 이렇게 되면 학생을 성적으로 변별하기 어려워집니다.

과제가 지나치게 단순하면 이해도가 아닌 정확도만 측정하게 되고, 이로 인해 주제를 잘 이해했음에도 사소한 실수로 높은 성적을 받지 못할 수 있습니다.

점수와 성적의 차이점에 대한 자세한 내용은 “[The difference between marks and grades: How we take into account the difficulty of an assessment, and not just how much a student got right](#)”(영문 제공, 동영상)를 참고하십시오.

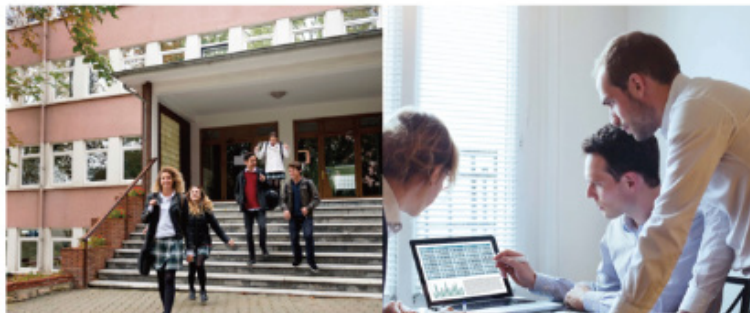
성공적인 시험 기수

[그림 20] 성공적인 시험 기수에 대한 다양한 관점



저는 모든 문제에 대한 답을 알고 있었습니다.

인근 학교보다 우리 학교의 응시생이 더 높은 성과를 냈습니다.



우리의 과정에 알맞은 학생을 선별할 수 있었습니다.

우리의 과정에 알맞은 학생을 선별할 수 있었습니다.

성공의 모습은 관점에 따라 크게 달라집니다. IB에서는 높은 수준의 구성개념 적합성과 실용적인 전달 모두에 초점을 맞춥니다. 시험 기수가 성공적이었다고 평가받으려면 다음과 같은 결과를 도출해야 합니다.

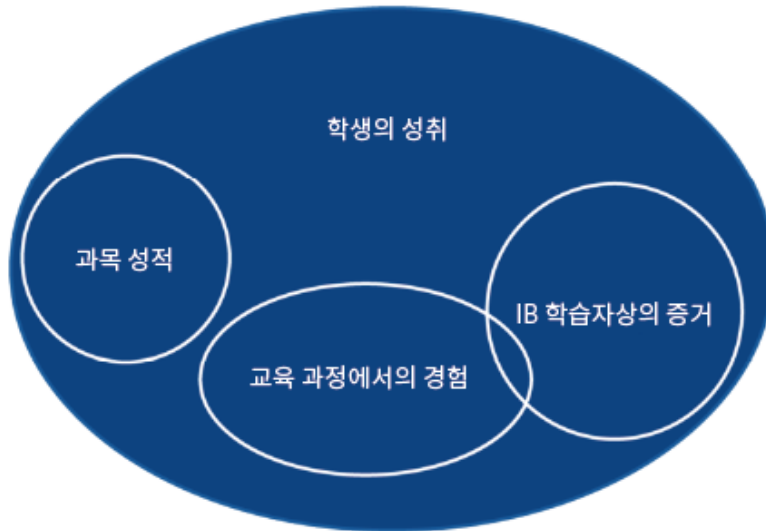
- 평가는 모든 학생에게 자신의 능력을 보여줄 공정한 기회를 제공했습니다.
- 학교와 학생에게 제공된 경험은 최대한 간단명료했습니다.
- 모든 이해관계자는 IB가 발표한 결과(성적)에 대해 신뢰를 유지했습니다.

성적과 성취도

IB 교육 목표에 명시된 바와 같이 IB 교육의 목표는 단순한 일련의 학업 성적 이상의 의미를 지닙니다. 이는 IB 학습자상에 반영되어 있으며 「IB 교육이란 무엇인가?」(2019)에 명시되어 있습니다.

평가 결과는 이러한 교육 목표 가운데 제한된 일부에만 초점을 맞출 수 있습니다. IB 시험에서 사용되는 보상 모델을 고려할 때, 수학 평가에서 배려심이나 도전 정신에 어느 정도의 점수를 배분해야 합니까? 심지어 이러한 중요한 자질에 ‘수치적 가치’를 부여하는 것이 합리적이라고 볼 수 있습니까? IB 평가 원칙 중 하나는 긍정적인 역류 효과를 가져오는 것이며, 이는 이러한 목표를 지원합니다.

[그림 21] 학생의 성취를 전체적으로 나타낸 그림



IB 프로그램 결과를 보고할 때, 학생의 성취의 전체 범위를 반영하기 위해 평가 성적 외의 요소도 고려하는 것이 중요합니다.

IB는 평가에서 학생들의 최종 성취도만을 기록하며, 그 결과를 달성하는 과정이 얼마나 어려웠는지 또는 학생들의 잠재력이 어느 정도인지는 전혀 명시하지 않습니다. IB는 이러한 두 요소 모두가 중요하다는 점을 인지하지만 총괄 평가만으로 이를 의미 있게 측정하는 것은 불가능하다고 판단합니다. 이러한 평가는 학생에 대한 전인적인 관점을 가진 학교의 책임입니다.

학생의 성취에 대한 ‘부가 가치’ 측정이나 ‘사전 성취에 기반한 예측 성적’을 산출하는 방법이 존재하지만, 학생 성과의 지표로 사용할 때는 주의가 필요합니다. 이러한 측정법은 평균적인 학생을 기준으로 하며, 각 학생은 고유한 특성과 자질의 조합으로 프로그램 전반에서 함께 학습할 기회를 가진 사람들에 의해 적절히 인정받아야 합니다.

예측 성적

「프로그램 기준 및 운영방침」(2020)에 명시되어 있듯이 ‘학교는 IB 검증 평가의 일환으로 학문적 진실성을 존중하고 최대한 정확하게 IB 예측 성적을 IB에 전달합니다’.

학습 과정 전반에서 학생의 성과가 다양할 수 있기 때문에 최종 성적을 예측하는 것은 어려울 수 있습니다. 예측 성적은 여러 중요한 기능을 수행합니다. 예측 성적은 성적 협의 위원회에서 근거 자료로 활용되며, 이를 통해 IB는 전체 학생 집단의 전반적인 강점과 각 과목의 성적 분포를 이해할 수 있습니다. 또한 예측 성적과 실제 성적 간에 큰 차이가 있는 경우를 식별하는 데 도움이 되며, 필요 시 추가적인 품질 점검을 실시할 수 있게 합니다. 학교 입장에서는 예측 성적이 학생, 학부모, 법적 보호자의 기대 수준을 설정하고 관리하는 데 도움이 됩니다.

교사가 모든 학생의 성적을 완벽히 정확하게 예측할 수는 없겠지만 가능한 한 정밀하게 예측하는 것을 목표로 합니다. 예측 성적이 정확할수록, 해당 과정 전반에서 더욱 유용하고 건설적인 자료가 됩니다.

예측 성적은 평가를 위한 중요한 데이터입니다. 과목 전체 수준에서는 과목별 성적 급간을 설정하는 데 고려되며, 정확성에 대한 근거가 있을 경우 학교 수준에서도 고려됩니다. 따라서 예측 성적을 신뢰할 수 있는 데이터로 사용하려면 정확성이 중요합니다.

또한 예측 성적은 다음과 같습니다.

- 학교에서 사용할 수 있는 데이터 중 하나이며, 학교는 이를 학생과 공유할지 선택할 수 있습니다.
- 프로그램 이수 이후 진로 선택 시 학생의 역량 지표로 작용하며, 학생의 성공 기반을 마련합니다.
- 점수가 아닌 성적(1~7, A~E)입니다.

성적 예측 시 예측 성적이 아닌 것을 이해하는 것이 중요합니다. 예측 성적이 아닌 것은 다음과 같습니다.

- 불리한 상황을 보완하기 위한 수단
- 징계 도구
- 이전 시험 기수의 성적 급간에 기반한 공식
- 대학 입학과 같은 예측 성적의 다른 사용법과 혼동해선 안 됨
- 학생, 학부모, 법적 보호자를 안심시키기 위한 수단
- 학생의 역량을 과대평가하거나 과소평가한 성적

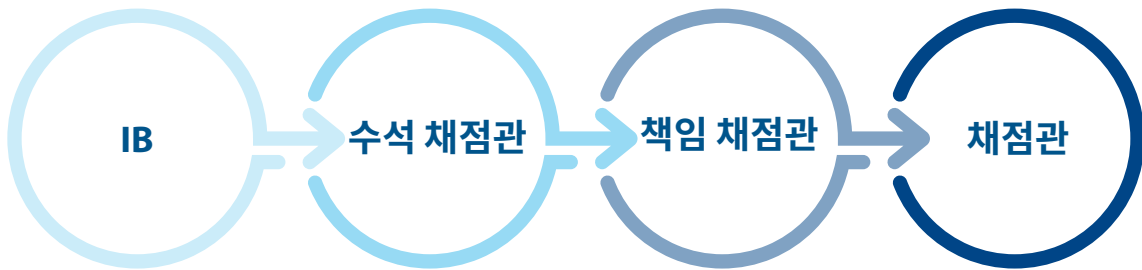
예측 성적 및 정확성을 향상시키기 위한 전략에 관한 자세한 내용은 인쇄 가능한 자료인 ‘[예측 성적: 교사를 위한 가이드](#)’(PDF) 및 프로그램 자료실의 다음 추가 지침을 참고하십시오.

- **MYP resources(MYP 자료) > MYP eAssessment(MYP 컴퓨터 기반 평가) > Cross-session resources(모든 세션 자료) > “Predicting IB Grades”**(영문 제공)
- **DP resources(DP 자료) > Assessment(평가) > Cross-session resources(모든 세션 자료) > “Predicting IB Grades”**(영문 제공)
- **CP resources(CP 자료) > Assessment(평가) > Cross-session resources(모든 세션 자료) > “Predicting IB Grades”**(영문 제공)

평가 과정: 역할과 책임

- 평가 과정에서 각 역할은 고유한 책임과 기능이 요구됩니다.
- 경우에 따라 한 사람이 평가 주기의 서로 다른 단계에서 여러 역할을 수행할 수 있습니다.
- 이러한 역할 중 일부는 IB가 직접 수행하며, 다른 일부는 IB 공동체의 전문가들이 담당합니다. 후자의 경우, 최종 승인과 평가의 측면에 대한 품질 유지는 IB가 책임을 집니다.

[그림 22] 평가 주기 내 주요 역할의 책임과 의무



IB

- 평가의 모든 측면에 대해 책임을 짐
- 채점관 채용, 채점관 수준, 결과 발표 등 평가 과정을 담당함
- 학생의 학업적 부정행위, 학교의 행정 실책, 포용적인 접근성 조치 및 불리한 상황과 관련된 이슈에 대해 결정을 내림
- 수석 채점관이 제안한 성적 급간을 수용하거나 이의를 제기함

책임 채점관

- 하나의 평가 항목을 담당함
- 해당 항목에서 학생의 답안에 부여될 점수를 최종적으로 결정함
- 모든 채점관이 해당 항목의 채점 기준을 이해했는지 확인함
- 해당 항목의 성적 급간을 설정할 수 있도록 수석 채점관을 지도함

수석 채점관

- 자신의 과목(교과군)의 모든 평가 항목을 개괄적으로 담당함
- 시험 출제를 포함하여 모든 항목 간의 기준이 일관되게 유지되는지 확인함
- 평가와 관련한 모든 학업적 이슈에 대해 중재하고 결정을 내림
- IB에 최종 성적 급간을 제안함

채점관

- 책임 채점관이 설정한 기준에 따라 학생의 과제를 채점함

책임 채점관 및 수석 채점관

책임 채점관의 책임과 의무는 다음과 같습니다.

- 내부 평가(IA)와 같은 특정 항목을 관리합니다.
- 시험 기수 내 및 기수 간 채점 기준을 설정하고, 표준화 논의와 품질 모델을 통해 해당 기준을 채점관들에게 설명합니다.
- 자신이 담당하는 항목에 대한 성적 급간 권고 사항과 관련하여 수석 채점관을 지도합니다
- 이해가 상충하지 않는 한 보통 평가 내용 개발자로 활동합니다.
- 각 시험 기수마다 외부 교육 전문가로 IB와 계약하며, 대체로 여러 기수에 걸쳐 기여합니다.

수석 채점관은 여러 관련 항목의 품질 유지를 담당합니다. DP와 CP에서는 수석 채점관이 전체 과목을 담당하며, MYP(보통 학문당 하나의 항목만 있음)에서는 과학이나 언어습득과 같은 교과군을 담당합니다. 이들은 해당 분야에서 IB의 학업 전문가로 활동합니다.

수석 채점관의 책임과 의무는 다음과 같습니다.

- 전체 과목(DP/CP) 또는 교과군(MYP)의 관련 항목 전반에 걸쳐 품질을 보장합니다.
- IB의 학업 전문가로서 기준을 유지하고 일관성을 보장합니다.
- 이의 제기 및 분쟁 해결을 지원합니다.
- 성적 부여 과정을 주도하고 IB에 성적 급간을 권고합니다.
- 특정 평가 항목에서는 책임 채점관으로도 활동하는 경우가 많습니다.
- 특정 기간 동안 IB와 계약합니다.

IB는 응시 인원이 적은 과목의 경우 담당 채점관을 임명합니다. 담당 채점관은 자신이 맡은 과목에서 수석 채점관과 동일한 역할과 책임을 가지며, 대개 더 적은 수의 책임 채점관과 채점관으로 구성됩니다. 수석 채점관은 본 간행물에서는 다루지 않는 교육과정 검토 활동을 통해 IB 교육과정 및 평가 방안을 논의하고 개선하는 데 IB와 협력합니다.

기타 채점관 역할

채점관은 책임 채점관이 설정한 기준에 따라 평가를 위해 제출된 과제를 채점할 책임이 있습니다. 이들은 품질 모델을 통해 기준을 이해하고 이를 적용하고 있음을 입증해야 합니다.

채점관은 특정 시험 기수에 채점하기 위해 IB에 지원합니다. 채점관은 채점하고자 하는 과목의 전문가여야 하며, 일반적으로 해당 과목을 가르치는 교사입니다. 또한 적절한 연령대의 학생들을 교수한 경험을 갖추어야 합니다. IB는 추천인 정보 제공을 포함한 엄정한 지원 절차를 통해 채점관의 자격을 검증합니다. 일반적으로 채점관은 한 기수당 과제로 항목 하나와 시험 항목 하나만 채점하도록 배정받습니다. 두 항목의 채점 기간은 서로 겹치지 않습니다. 채점관의 보수는 '실제' 학생 답안지당 지급됩니다. 실제 답안지에는 채점관이 채점 기준을 이해하고 있음을 IB에 입증하기 위한 채점 자격 취득용 답안지나 시드 답안지('품질 모델' 섹션 참조)가 포함되지 않습니다.

팀 리더는 경험이 풍부한 채점관으로, 다른 채점관이 올바른 채점 기준을 이해하도록 지원하는 역할을 IB로부터 부여받습니다. 또한 이들은 자격 확인 과정을 통해 채점관을 지원하고, 시드 답안지 채점 결과가

확정 기준에서 크게 벗어난 경우 피드백을 제공합니다. 과목별 팀 리더 수는 전체 채점관 수에 따라 달라지며, 이는 전체 학생 집단의 규모에 따라 결정됩니다. 응시하는 학생의 수가 적은 과목의 경우, 책임 채점관이 모든 채점관을 지원할 수도 있습니다.

표준화 팀 구성원은 대체로 품질 모델 개발로 이어지는 정성적 논의에 참여하는 경험이 풍부한 채점관이며, 책임 채점관의 기준이 실제로 적용 가능하고 준수될 수 있도록 그 적용 방식에 대해 논의합니다.

채점관의 역할에 관한 자세한 내용은 “[Introducing: The examiners: Who are they and what do they do?](#)”(영문 제공, 동영상)를 확인하십시오.

IB 직원의 책임

과목 관리자 및 평가 운영 분석자는 IB 관점에서 신뢰할 수 있는 평가를 보장하는 데 관여하는 주요 역할입니다.

IB 내 별도 팀들은 특수 변형 시험지 및 기타 포용적인 접근성 조치의 요청, 불리한 상황 관련 요청, 학문적 진실성 관련 사안, 채점관 수준 및 학교 조정 요인의 모니터링 등을 관리합니다.

평가 개발 및 시행 책임자 및 평가 원칙 및 수행 책임자의 지원을 받는 평가 담당자는 과목별 성적 협의 위원회의 협의 결과로 수석 채점관이 제시한 성적 급간을 고려하고 이를 승인하거나 수석 채점관에게 권고 사항을 재고하도록 요청할 책임이 있습니다.

궁극적으로 평가 주기에서 이루어지는 모든 결정에 대한 책임은 IB에 있습니다. IB는 외부 전문가를 활용하여 공정하고 양질의 평가를 만들지만, 최종 책임은 IB에 있습니다.

시험 출제자의 역할

평가 내용 개발자는 시험 출제 과정에서 지정된 역할을 수행하며, 여기에는 출제, 수정, 검토 업무가 포함될 수 있습니다. 평가 내용 개발자에게 요구되는 주요 기능은 다음과 같습니다.

- 창의성: 흥미를 유발하고 교육과정에 부합하는 질문 구성
- 문화적 인식: 문제가 편향되지 않도록 보장
- 의사소통 기능: 충실히 번역되고 의도한 기능을 평가하여 신뢰할 수 있는 채점으로 이어지도록 명확하고 모호하지 않은 문제 작성

그 밖의 평가 제작 역할에는 다음이 포함되지만 이에 국한되지는 않습니다.

- 기술적 디자인 편집자: 시험 문제의 형식 구성
- 번역 담당자: 평가 내용을 요구되는 언어로 번역
- 외부 검토자: 학생의 관점에서 시험을 수행해 보고 문제의 길이와 명확성에 대한 피드백 제공

평가 출제 방식에 관한 자세한 내용은 ‘[평가 제작](#)’ 섹션을 참고하십시오.

채점관 위계 체계

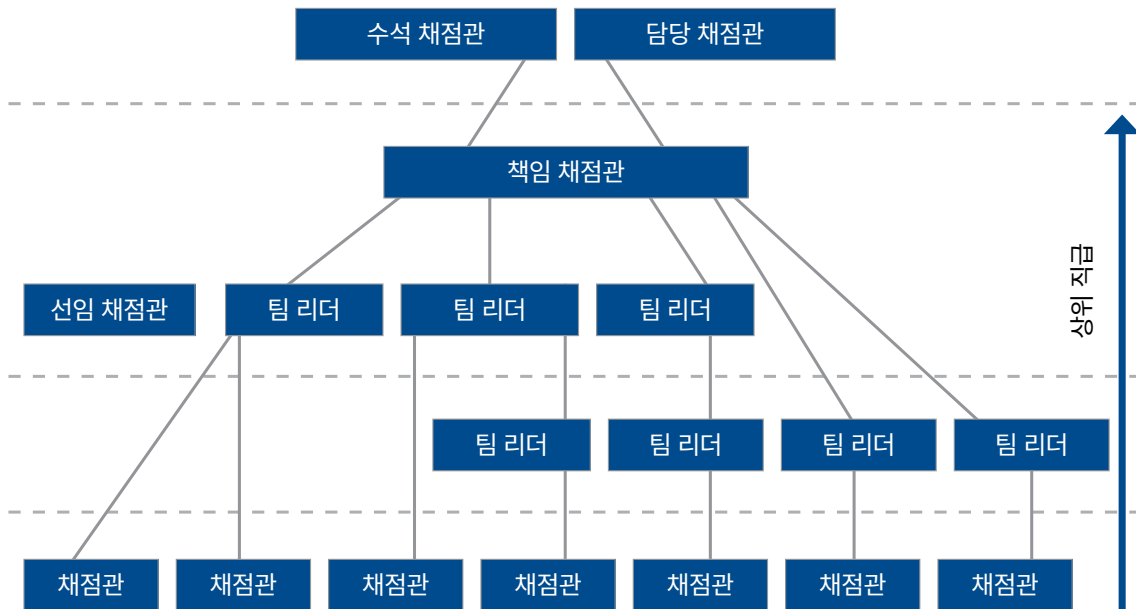
- 책임 채점관은 특정 항목에 부여할 점수를 최종적으로 결정하는 사람이며, 책임 채점관이 설정한 기준은 확정 기준입니다. 다른 모든 채점관은 반드시 이 확정 기준을 따라야 합니다.
- IB는 기준 설정 과정에 참여한 채점관이 품질 모델이나 팀 리더를 통해 기준을 학습한 채점관보다 책임 채점관의 기준을 더 잘 이해한다고 봅니다.
- ‘선임’이란 채점관 위계 체계에서 책임 채점관과의 구조적 근접성을 의미합니다. 이는 특정 개인이 채점관으로 활동한 기간이나 교수 경력을 의미하지 않습니다.

IB 채점의 근본 원칙은 학생 과제에 대한 ‘올바른 점수’를 최종적으로 결정하는 사람이 책임 채점관이라는 것입니다. IB는 ‘올바른’ 점수에 대해 여러 가지 합리적인 견해가 존재할 수 있음을 인정하지만, 어떤 채점관이 채점했는지에 따라 학생의 점수가 달라지는 것은 공정하지 않다고 봅니다. 따라서 책임 채점관이 기준을 설정하고, 다른 모든 채점관은 이 기준을 따라야 합니다.

책임 채점관은 기준을 설정할 때 선임 채점관으로 구성된 팀의 지원을 받습니다. IB는 선임 채점관들이 책임 채점관과 기준에 대해 직접 논의할 기회를 가지므로, 이들이 기준을 가장 잘 이해할 것이라고 간주합니다. 다만, 책임 채점관의 기준에 맞춘 모든 채점관의 채점 신뢰도가 매 시험 기수마다 도표화되어 채점관 신뢰도에 대한 지표를 제공합니다.

채점관 사이에는 위계 체계가 존재합니다. 특히 학생 수가 많은 과목의 경우 [그림 23]에 제시된 예시와 같은 구조가 있을 수 있습니다.

[그림 23] 학생의 수가 상당히 많은 과목에서 채점관 역할의 위계



위계 체계는 상황에 따라 유연하게 조정될 수 있습니다. 채점 자격 취득용 답안지 또는 시드 답안지에서 도출된 데이터 분석 결과, 특정 채점관이 위계상 더 선임인 채점관보다 책임 채점관의 기준을 더 잘 이해하고 있음이 확인될 경우, 이러한 점은 위계 운영에 반영됩니다. 팀 리더는 기준을 단순히 적용하는데 그치지 않고 이를 명확히 설명할 수 있어야 하므로, 팀 리더를 선발할 때에는 채점 신뢰성만이 유일한 고려 요소는 아닙니다.

이상적으로는 모든 채점관이 동일한 기준에 따라 채점하므로 위계 체계를 기반으로 판단을 내려야 할 상황이 발생하지 않아야 합니다. 그러나 현실적으로 위계 체계는 IB가 어떤 점수를 채택해야 하는지 결정하는 데 필수적입니다. 특히 까다롭거나 논란의 여지가 있는 답안지의 경우, IB는 해당 답안지를 책임 채점관에게 전달하여 적절한 점수를 최종적으로 결정합니다.

평가 공정성

- IB 평가는 모든 학생에게 동등한 성취의 기회가 보장될 때에만 공정하다고 할 수 있습니다.
- 학교의 행정 실책이나 학문적 진실성 정책을 위반하는 학생의 행위는 규칙을 성실히 준수한 학생들에게 불이익을 초래합니다. IB는 이러한 행위를 방지하기 위해 가능한 모든 노력을 기울입니다.
- 각 프로그램의 「평가 절차」의 ‘일반 규정’ 섹션과 「학문적 진실성 정책」에 명시된 IB 규칙은 학생의 학업적 부정행위나 학교 행정 실책이 발생할 가능성을 최소화하도록 설계되었습니다. 궁극적으로는 학업적 부정행위가 용납되지 않고 적절히 보고되는 학습 문화를 조성할 수 있는 주체는 학교뿐입니다.
- 일부 평가 방식은 다른 형식에 비해 학업적 부정행위에 덜 취약할 수 있으나, 평가를 설계할 때 구성개념 적합성(즉 실제로 평가하고자 하는 요소를 평가하는 것)을 최우선으로 고려해야 한다는 IB의 원칙은 변함이 없습니다.
- 학생의 학업적 부정행위나 학교의 행정 실책이 의심되는 모든 경우는 반드시 IB에 보고되어야 합니다.

학문적 진실성 정의하기

학문적 진실성은 교육의 기본 원칙이며 우리 자신이 한 개인으로서 신뢰받을 수 있도록 책임 있게 행동하려는 선택입니다. 이는 학습, 교수, 평가 과정에서, 그리고 규정에 맞는 정직한 학술적 과제를 수행하는 과정에서 윤리적인 의사결정 및 행동의 토대가 됩니다. 학문적 진실성은 엄격한 규칙으로 강요되어서는 안 됩니다. 대신 학교와 학부모 및 법적 보호자를 포함한 더 넓은 공동체 안에서 긍정적인 문화로 자리 잡아, 모두에게 더 공정한 평가 결과가 유기적으로 도출되도록 해야 합니다.

평가가 타당성을 갖추기 위해서는 동일한 평가를 치른 다른 학생들과 비교하여 해당 학생의 성취에 대한 정확한 분석이 제공되어야 합니다. 이러한 이유로 IB는 채점, 성적 부여, 시험의 편향 제거 전반에 걸쳐 일관된 접근 방법을 유지하는 데 각별한 주의를 기울입니다. IB가 수립한 규칙과 규정 또한 이러한 기회의 형평성을 보장하고 공정성을 확보하기 위한 또 하나의 중요한 요소입니다.

「학문적 진실성 정책」에서는 학업적 부정행위를 학생 본인 또는 제3자가 하나 이상의 평가 항목에서 불공정한 이익을 얻을 가능성이 있는 의도적 또는 비의도적 행위로 정의합니다. 이러한 행위는 해당 학생뿐 아니라 같은 평가를 치른 모든 학생에게 영향을 미치며, 학위 자격의 타당성을 떨어뜨립니다. 따라서 IB는 학업적 부정행위를 매우 심각하게 받아들이며, 부정행위의 방지와 결과에 관한 구체적인 내용은 「학문적 진실성 정책」에서 확인할 수 있습니다.

평가 설계는 학업적 부정행위를 방지하는 데 있어 중요한 수단입니다. 일부 평가 유형은 다른 유형에 비해 모니터링이 더 용이하기 때문입니다. 예를 들어, 시험의 경우 내부 평가 과제에 비해 다른 사람의 도움을 받는 것이 훨씬 어렵습니다. IB는 평가 접근 방법을 고안할 때 이러한 요소를 고려하지만, 원칙은 학업적 부정행위의 가능성을 방지하기 위해 구성개념 적합성을 저버려서는 안 된다는 것입니다.

관리 가능성 역시 타당성의 중요한 요소이며, 학업적 부정행위의 발생 가능성을 줄이는 과정에서 함께 고려되어야 합니다. 특히 디지털 평가에서 시험장을 설정하는 것은 학교에 실질적인 도전 과제가

될 수 있으며 반드시 이를 고려해야 합니다. IB는 규칙과 규정을 수립할 때 학교의 입장을 반영하기 위해 노력하며 코디네이터와 주임 교사들의 다양한 경험(좋은 사례와 어려움 모두)에 귀 기울이고자 합니다.

학문적 진실성과 윤리적 행동의 의미에 대한 자세한 내용은 “[Academic Integrity in the IB: Making the Right Choices](#)”(영문 제공, 동영상)를 참고하십시오.

이해의 상충

IB에서는 시험 내용에 대한 접근은 엄격히 통제되며, IB는 이해 상충을 구성할 수 있는 시험 응시 학생과의 모든 연관성을 적극적으로 관리합니다. 이러한 이해의 상충이 발생할 것으로 예상되면 교직원의 역할과 책임은 재조정됩니다.

IB의 원칙은 채점관이 자신의 학생의 과제 또는 이해 상충 관계에 있는 학교(예: 해당 채점관이 다른 과목을 지도하는 학교, 최근까지 재직했던 학교)의 학생 과제를 채점할 수 없습니다. IB가 통제할 수 없는 불가피한 상황(예: 요구된 기준에 맞추어 채점을 할 자격을 갖춘 사람이 아무도 없는 경우)으로 인해 이러한 원칙을 적용할 수 없는 극히 예외적인 상황에서는, 별도의 채점관이 해당 채점을 검토합니다. 이때 IB는 해당 채점관이 자신의 학생에게 다른 기준을 적용했다는 어떠한 정황이 없음을 입증해야 할 것입니다.

학교의 행정 실책 및 학생의 부정행위 관리하기

각 프로그램별로 관련된 「평가 절차」 간행물의 ‘일반 규정’ 섹션 및 「IB 시험 시행 지침서」 또는 “[The conduct of IB Middle Years Programme on-screen examinations](#)”(영문 제공)에는 행정 실책 발생 가능성을 최소화하기 위한 규칙과 지침이 명시되어 있습니다. 서면 기반 시험의 경우 「IB 기밀 시험 자료의 안전한 보관을 위한 지침서」에서 학교 내 IB 기밀 자료의 보관 방식과 이를 시행하기 위해 학교가 취해야 할 조치를 안내합니다.

시험은 DP 및 CP의 경우 「IB 시험 시행 지침서」, MYP의 경우 “[The conduct of IB Middle Years Programme on-screen examinations](#)”(영문 제공)에 명시된 지침에 따라 반드시 감독되어야 합니다. 시험 감독관은 시험이 진행되는 동안 학업적 부정행위가 발생하지 않도록 주의를 기울여야 합니다.

IB는 평가 기수 동안 예고 없이 시험 현장을 방문하여 모든 필수 시험 절차가 준수되고 있는지 확인합니다. 이러한 점검은 학교의 IB 절차 이행 현황을 확인하기 위한 표본 조사입니다. 그러나 IB는 일상적으로 높은 기준을 유지할 수 있는 학교장과 프로그램 코디네이터에게 행정 실책 방지에 대한 주요 책임을 부여합니다. 아울러, 교사와 학생 모두가 모범적인 행동과 높은 수준의 진실성을 실천하도록 장려하는 학교 문화가 정착되었는지 확인해야 합니다.

시험 기간에 학문적 진실성 기준을 지키는 것뿐 아니라, 교실과 가정에서 수행되는 과제가 학생 자신의 작업이며, 그리고 표절이 발생하지 않도록 확인하는 것이 중요합니다.

현재 학생의 과제 수행을 ‘지원’ 가능한 다양한 웹사이트와 인공지능(AI) 도구가 존재합니다. 이러한 부정행위를 방지하는 가장 효과적인 방법은 바로 교사입니다. 교사는 학생과 지속적으로 학습 과정을 함께하며, 제출된 과제가 학생의 평소 수행 수준과 일치하는지 판단할 수 있기 때문입니다. 이러한 이유로 학생과 교사 모두 해당 과제가 학생 본인의 것임을 확인해야 합니다. 단순히 표절 검사 소프트웨어에 의존하는 것만으로는 충분하지 않습니다. 교사는 학생이 내부 평가를 수행할 때 함께 작업하여 과제의 진위성을 파악해야 합니다. 과제가 한 번 제출된 이후에는, 표절이 확인되더라도 IB는 대체 과제물의 제출을 수용할 의무가 없습니다.

어떠한 수준의 표절도 허용되지 않으며 외부의 저작물을 인용하거나 AI 소프트웨어를 활용할 경우에는 반드시 「효과적인 인용 및 출처 표기」 및 「학문적 진실성 정책」에 명시된 기준에 따라 정확하게 출처를 표기해야 합니다. IB는 학생의 학업적 부정행위가 의심되는 경우를 식별하기 위해 다양한 소프트웨어를 사용합니다. 부정행위의 증거가 발견되면 공식 조사가 시작됩니다.

학문적 진실성이 정착된 문화를 조성하고 관리하는 것은 모든 IB 월드 스쿨에 요구되는 사항이며 이를 반복적으로 위반할 경우 학교 인증에 영향을 미칠 수 있습니다.

국제 시험 및 표준시간대와 관련한 어려움

IB는 학교가 전 세계에 분포해 있다는 특성 때문에, 대부분의 국가 교육 체계와 달리 모든 학생이 동시에 시험을 치를 수 없다는 고유한 어려움에 직면합니다. 이로 인해 어떤 학생은 다른 지역의 학생들이 시험을 시작하기도 전에 이미 시험을 마치게 되며, IB는 이러한 상황이 악용되지 않도록 학생과 교사에게 높은 수준의 학문적 진실성을 요구합니다. 학생 수가 충분하여 실행 가능한 경우, 한 표준시간대에서 시험을 마친 학생과 다른 표준시간대에서 시험을 시작하는 학생 사이의 시간차를 줄이기 위해 별도의 ‘시험 구역’이 마련되지만, 시험 시작 시간에는 상당한 차이가 있을 수 있습니다.

IB는 이러한 위험을 완화하기 위해 시험 내용이 공유될 수 있는 웹사이트와 소셜 미디어 플랫폼을 면밀히 모니터링하는 등 다양한 조치를 시행하고 있습니다. 시험 내용을 시험장 밖으로 반출할 수 없게 한 규칙 역시 시험 문제가 온라인에 유출될 가능성을 줄여 줍니다. 오전 또는 오후 시험이 2시간보다 짧고 해당 시간대의 유일한 시험일 경우, 시험 시작 시각 후 최소 2시간 동안 학생을 계속 감독해야 한다는 규정은 시험 직후 공유가 금지된 정보 유출 가능성을 줄이는 데 도움이 됩니다.

또한 IB는 시험 문제가 단순한 지식 암기가 아니라 이해를 평가하도록 설계함으로써, 설령 문제를 사전에 알게 되더라도 얻을 수 있는 실질적인 이점이 제한적일 것입니다. 마지막 조치는 시험 종료 후 24 시간 동안 학생들에게 시험에 대해 논의하지 않도록 요구함으로써 ‘무심코 이루어지는 논의(innocent discussion)’의 가능성을 완전히 차단하는 것입니다.

IB는 표준시간대 차이를 악용한 부정행위의 위험을 줄이기 위해 지속적으로 새로운 방안과 개선안을 모색하고 있습니다.

디지털 평가의 장점

디지털 평가는 학업적 부정행위를 관리하는 데 있어 다양한 이점을 제공합니다. 시험 내용에 접근할 수 있는 시간은 엄격히 제한되며, IB는 누가 언제 시험 내용에 접근했는지 모니터링할 수 있습니다.

또한 추가 시간 부여나 일시정지와 같이 학교가 시험에 적용한 모든 수정 사항은 기록되며, 해당하는 경우 이에 대한 정당한 사유를 제출해야 합니다. 이를 통해 IB는 공정한 시험 운영을 보장하며, 합리적인 조정에 대해서는 학교에 더 많은 책임을 위임할 수 있습니다.

디지털 평가는 메타데이터를 통해 학생이 시험을 치른 경험을 보다 세밀하게 파악할 수 있게 하며, 이는 학업적 부정행위가 의심되는 사례를 조사할 때 증거 자료로 활용될 수 있습니다. 또한 디지털 시험 답안의 전자식 특성 덕분에 수기로 작성된 답안에서는 불가능한 의심스러운 유사한 답안들을 대규모로 확인할 수도 있습니다.

IB는 디지털 평가가 특히 해킹 시도와 같은 새로운 형태의 학업적 부정행위에 노출될 수 있음을 인지하고 있으며, IB는 이에 대응하기 위한 다양한 도구와 절차를 마련해 두었습니다.

자료

교사, 학교 및 학생은 학업적 부정행위가 용납되지 않는 문화를 조성하고, 발생 시 이를 해결하기 위해 경계를 늦추지 않음으로써 부정행위 사례에 도전하고 막는 데 가장 적합한 위치에 있습니다. 이 주제는 본 간행물의 '윤리적 사고방식 기르기' 섹션에서 자세히 다뤄집니다. 「학문적 진실성 정책」을 포함, 학교가 학문적 진실성 문화를 조성하는 데 도움이 되는 자료는 IB 웹사이트의 'Academic integrity(학문적 진실성)'에서 확인할 수 있습니다.

학업적 부정행위와 관련해 추가 지원이 필요하거나, 학교에서 다루지 않는 우려 사항이 있거나, 또는 학문적 진실성과 관련된 새롭게 떠오르는 이슈에 대해 의견이 있는 경우 IB 지원팀에 문의하시기 바랍니다.

모든 학생을 위한 공정성

- 타당성을 유지하기 위해 형성 평가와 총괄 평가는 반드시 학습자의 가변성을 고려해야 하며, 어떠한 학생도 차별해서는 안 됩니다. IB 평가는 이러한 학습자의 가변성을 처음부터 염두에 두고 설계됩니다.
- 공정성을 보장하는 가장 좋은 방법은 모두가 접근할 수 있는 평가를 설계하는 것입니다. IB 평가는 보편적 설계 원칙을 따라 평가의 타당성에 영향을 줄 수 있는 장벽을 제거합니다. 학교는 반드시 IB에서 수립한 지침을 준수하여 서면, 구술, 디지털 시험의 환경이 어떤 학생에게도 유리하거나 불리하게 작용하지 않도록 해야 합니다.
- 「**접근성 및 포용성 정책**」은 특수 변형 시험지와 같이 평가에 대한 접근성을 높이기 위해 추가 조치가 필요한 상황을 다룹니다.
- 평가의 접근성이나 타당성에 영향을 미칠 수 있는 예기치 못한 상황이 발생할 경우 이러한 상황의 부정적 영향을 완화하기 위한 일련의 조치는 「**불리한 상황 정책**」에서 확인할 수 있습니다.
- 학습 및 교수 단계는 형성 평가와 함께 총괄 평가 및/또는 IB 평가에 대비할 수 있도록 일관되어야 합니다. 포용적 접근성 조치의 핵심 원칙 중 하나는 이러한 조치가 학생의 평소 작업 방식으로서 학습, 교수, 평가 전반에 걸쳐 사용되어야 한다는 점에서, 이러한 일관성은 특히 중요합니다. 일관성은 디지털 평가를 포함한 모든 평가 방식 전반에서 필수적입니다.
- 이러한 조치의 목적은 모든 학생에게 공정성을 보장하는 데 있습니다. IB는 결정을 내릴 때 모든 학생에게 무엇이 공정한 것이 무엇인지 고려해야 합니다.

모든 학생은 가능한 한 공정한 평가 환경에서 자신의 능력을 발휘할 수 있어야 합니다. 일부 표준 평가 조건은 학생들이 자신의 성취 수준을 입증하지 못하게 함으로써 일부 학생들에게 불리하게 작용할 수 있습니다. 또한 학생이 통제할 수 없는 불리한 상황이나 사건 역시 학생의 성과에 영향을 미칠 수 있습니다. 공정성을 보장하기 위해서는 반드시 이 두 가지 요인을 모두 고려하여 적절한 조치를 마련해야 합니다.

평가에서 공정성을 보장하는 최선의 방법은 모든 학생이 동일한 조건에서 동일한 평가를 수행하는 것입니다. 이를 위해 평가는 학생이 가질 수 있는 다양한 요구 사항을 면밀히 고려하여 개발됩니다. 학습자의 가변성을 고려하고 학생의 요구를 충족하기 위한 가장 효과적인 접근 방법은 보편적 학습설계 및 보편적 평가 설계 원칙에 기반하여 평가를 설계하는 것입니다. 이러한 원칙에 따라 평가가 설계된 이후 다음 단계는 IB 「**접근성 및 포용성 정책**」에 명시된 포용적인 접근성 조치를 통해 접근성을 보장하는 것입니다.

포용적인 접근성 조치는 장벽을 제거하고 모든 학생의 평가 접근을 가능하게 설계됩니다. 이러한 장벽은 평가에서 신경 다양성, 장애, 개인적 어려움 또는 다양한 사회정서적 상황을 충분히 고려하거나 반영하지 못할 때 발생할 수 있습니다.

학생의 고유한 요구와 관련된 모든 합리적인 조정은 검토 대상이 됩니다. 자세한 내용은 「**접근성 및 포용성 정책**」을 참조하십시오.

불리한 상황이란 학생이 통제할 수 없는 상황으로, 학생의 성과에 부정적인 영향을 미칠 수 있는 경우를 의미합니다. 「**불리한 상황 정책**」에는 이러한 영향에 대응하고 이를 완화하기 위한 절차가 요약되어 있습니다. 포용적인 접근성 조치는 장벽을 제거하여 모든 학생이 평가에 접근할 수 있도록 시행됩니다.

이는 장벽을 마주한 학생에게 특혜를 제공하는 것이 아니라, 모든 학습자가 다른 학생과 동일한 조건에서 자신의 기능, 지식, 능력을 보여줄 수 있도록 공정한 접근 기회를 보장하는 데 목적이 있습니다.

포용적인 접근성 조치 원칙

포용적인 접근성 조치 원칙은 「접근성 및 포용성 정책」에 명시되어 있습니다. 평가는 각 프로그램의 목적에 반영된 청소년의 발달 단계를 고려해야 합니다. 포용적인 접근성 조치는 모든 프로그램에서 형성 평가와 총괄 평가 모두에 적용되며 학습 및 교수 경험 전반에 걸쳐 고려되어야 합니다.

IB는 모든 과목에서 학생에게 부여되는 성적이 반드시 해당 학생의 성취 수준을 정확하게 반영하도록 보장해야 합니다. 이를 위해 포용적인 접근성 조치의 필요 여부와 관계없이 모든 학생에게 동일한 평가 기준이 적용됩니다.

합리적인 조정을 포함한 포용적인 접근성 조치는 학생이 평가에 접근할 수 있도록 하기 위한 사전 심사 조치입니다. 이러한 조치는 구술, 서면 또는 디지털 시험에 대해 사후에 소급하여 요청할 수 없습니다.

학생을 위해 요청된 조치는 어떤 평가 항목에서도 해당 학생에게 이점을 제공해서는 안 됩니다.

포용적인 접근성 조치는 모든 평가 요구 사항을 충족할 수 있는 역량을 갖춘 학생을 대상으로 합니다.

학생에게 포용적인 접근성 조치가 필요한 경우 학교는 「접근성 및 포용성 정책」에 명시된 IB의 기준에 따라 해당 조치를 제공할 수 있습니다. 평가를 위해 요청된 조치는 학생의 평소 작업 방식을 반영해야 합니다. 학습 과정 전반에서 학생의 요구를 가장 잘 충족시키는 조치를 선택하여 총괄 평가로 원활하게 이어질 수 있도록 하는 것은 학교의 책임입니다. 학생의 평소 작업 방식과 일치하지 않는 조치에 대한 요청은 매우 예외적이고 특수한 경우에만 승인됩니다.

IB는 국제적 소양에 기반한 교육철학을 따릅니다. 평가 접근성 지원이 필요한 학생들 간의 형평성을 위해 본 정책은 각국 및 지역(영역)별로 수용되고 있는 관행을 고려한 결과를 나타냅니다.

각각의 포용적인 접근성 조치 요청은 개별 사례에 따라 심사됩니다. 과거에 IB 또는 다른 수여 기관에서 승인된 조치가 있더라도, 이번에 요청된 조치의 승인 여부에 영향을 미치지 않습니다.

IB는 학생에 관한 모든 정보를 기밀로 유지합니다. 필요에 따라, 정보를 기밀로 유지하도록 교육받은 IB 직원 및 IB 최종 학위 수여 위원회의 구성원과 공유될 수 있습니다.

학교가 포용적인 접근성 조치를 승인 없이 이행하거나 IB가 명시한 조건을 준수하지 않은 경우, 학생은 해당 과목 및 수준의 성적을 부여받지 못할 수 있습니다.

학생이 시험 언어를 능숙하게 사용하지 못하는 이유가 학습 지원 필수 사항에서 비롯된 것임을 입증할 수 있는 경우 포용적인 접근성 조치가 승인될 수 있습니다.

채점관에게는 학생의 개인적 상황이나 불리한 상황에 대한 정보가 제공되어서는 안 됩니다.

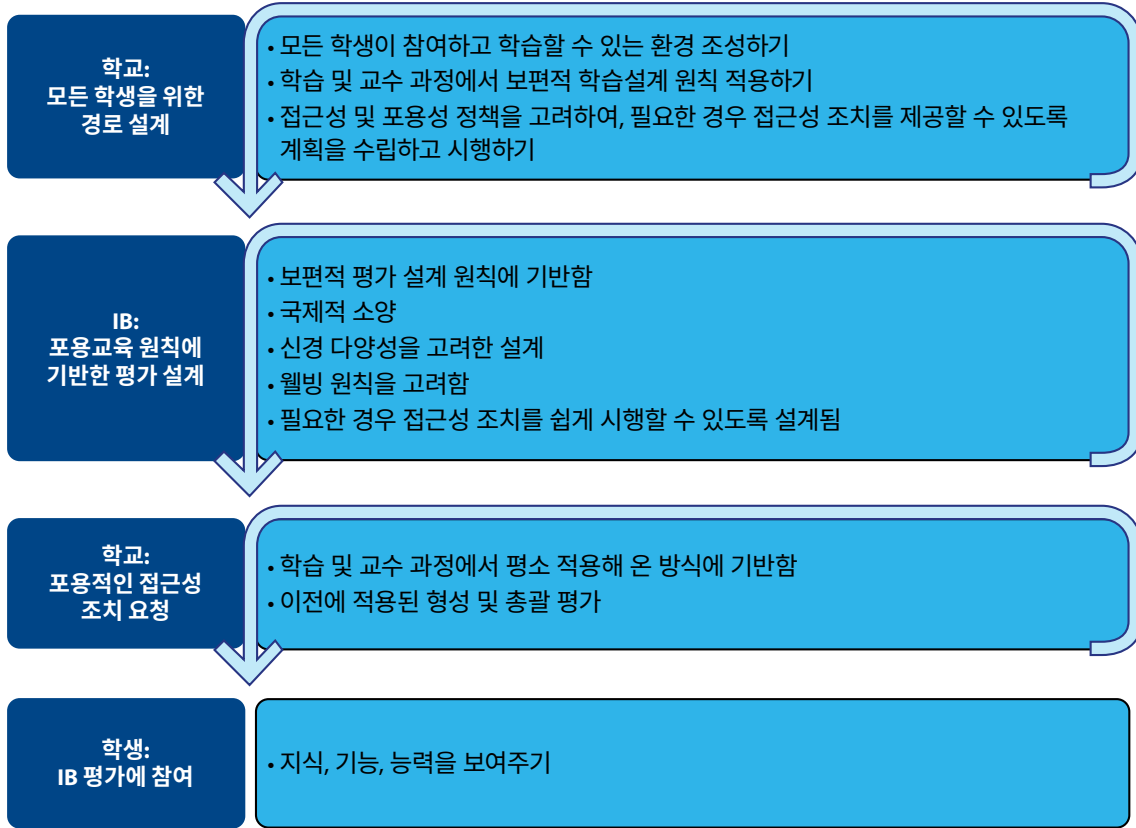
내부 평가 과제의 경우 교사는 학생의 과제를 채점할 때 어떠한 수정도 해서는 안 됩니다.

사용 가능한 포용적인 접근성 조치의 목록은 정기적으로 개정됩니다. 유사한 요구를 지닌 모든 학생에게 동일하게 제공할 수 있는 경우, 학교가 제안한 대안적 조치도 고려될 수 있습니다.

학생 및/또는 승인된 포용적인 접근성 조치로 인해 시험 중 다른 학생에게 방해가 될 수 있는 경우, 해당 학생은 반드시 분리된 공간에서 시험을 치러야 하며, 각 프로그램의 시험 시행 규정에 따라 감독관의 지도하에 있어야 합니다.

포용적인 접근성 조치의 특성으로 인해 발생한 문제나 시험 중 학생에게 예기치 못한 어려움이 발생한 경우 가능한 한 신속하게 IB 지원팀에 보고해야 합니다.

[그림 24] 포용적인 접근성 조치의 시행



평가 면제

일반적으로 모든 평가 항목에 대한 면제는 부여되지 않습니다. 그러나 평가 항목 전체 또는 일부에서 학생이 수행 불가능한 생리학적 기능을 요구하는 경우 면제가 승인될 수 있습니다. 평가 항목에 대한 면제 요청을 제출하기 전에 모든 합리적인 조정 방안을 고려했는지 숙고해야 합니다. 면제는 이에 대한 충분하고 타당한 근거가 있는 경우에만 승인됩니다. 학생이 해당 항목에서 요구되는 기능을 수행할 수 없다는 점을 명확하고 완전하게 문서로 입증되어야 합니다.

평가 면제와 관련한 원칙 및 절차에 대한 자세한 내용은 「접근성 및 포용성 정책」의 포용적인 접근성 조치 목록을 참고하십시오.

디지털 평가의 포용적인 접근성 조치 기회

디지털 평가는 학생에게 평가가 제시되는 방식을 보다 주도적으로 선택할 수 있는 권한을 제공합니다. 컴퓨터 기기는 다양한 글꼴, 텍스트 크기, 색상을 제공하여, 개인의 선호와 요구에 맞게 조정할 수 있게 하고 보편적으로 접근할 수 있도록 합니다. 보편적 설계 원칙을 적용함으로써 접근성이 향상되었지만, 평가에 접근하기 위해 여전히 추가 조치가 필요한 경우가 있습니다. 이를 위해 디지털 기능은 이러한

조치를 구현할 수 있는 대안으로 고려됩니다. 사용 가능한 조치는 「[접근성 및 포용성 정책](#)」에 나열되어 있습니다.

디지털 시험을 위한 포용적인 접근성 조치의 목표는 종이 기반 시험과 마찬가지로 장벽을 제거하는 것입니다. 공정성을 보장하기 위해 승인된 접근성 조치는 디지털 시험과 종이 기반 시험 모두에서 일관되게 적용되어야 합니다.

불리한 상황

불리한 또는 예기치 못한 상황이란 학생이 통제할 수 없는 것으로, 학생의 성과에 부정적인 영향을 미칠 수 있는 상황을 의미합니다. 여기에는 일시적인 질병 또는 부상, 심각한 스트레스, 이례적으로 어려운 가정 상황, 가족 및 친지의 사망, 또는 학생의 건강이나 안전을 위협할 수 있는 사건 등이 포함됩니다. 또한 소요 사태, 자연재해처럼 학교 공동체 전체에 영향을 미치는 사건도 불리한 상황에 해당할 수 있습니다.

학교 측의 과실로 발생한 상황은 불리한 상황에 포함되지 않습니다. 학교는 교직원 관련 문제를 포함하여 모든 학생이 프로그램 및 평가 요구사항을 준수하도록 관리할 책임이 있습니다.

불리한 상황에 범주에 포함되거나 제외되는 사항과 완화 조치에 관한 자세한 내용은 「[불리한 상황 정책](#)」에서 확인할 수 있습니다.

보편적 평가 설계

앞선 섹션에서는 IB 평가에 대한 공정한 접근을 보장하기 위해 포용적인 접근성 조치가 필요한 상황을 어떻게 관리하는지 살펴보았습니다. 그러나 가장 이상적인 해결책은 애초에 참여의 장벽이 존재하지 않도록 평가를 설계하는 것입니다. 보편적 평가 설계는 접근성, 포용성, 공정성, 문화적 다양성 및 감수성을 함께 고려하는 개념입니다. 이 접근 방법을 따름으로써 IB 평가는 설계, 개발, 시행 과정 전반에서 평가 모델의 다양한 장벽을 제거하도록 고안됩니다.

처음부터 포용적이고 구성개념과 무관한 요소를 최소화한 평가를 설계하면, 추가 접근성 조치를 도입하거나 시행해야 할 필요가 크게 줄어들어 더 많은 학생에게 도움이 됩니다.

보편적 평가 설계는 모든 학습자에게 접근 가능한 학습 환경을 조성하는 것을 목표로 하는 보편적 학습 설계라는 더 넓은 프레임워크의 한 측면입니다. 이 접근 방법은 신경 다양성, 선호, 능력, 장애, 관심사, 문화, 언어, 배경을 포함하지만 이에 국한되지 않는 더 넓은 범위의 다양성이 고려됩니다.

보편적 학습설계 원칙은 교육과정 및 평가 설계, 교육과정 및 평가 개발, 학교 운영을 포함한 교육의 다양한 측면 전반에 적용될 수 있습니다. IB에서의 보편적 학습설계에 관한 자세한 내용은 라오(Rao) 외 연구자들의 연구(2016)를 참고하십시오.

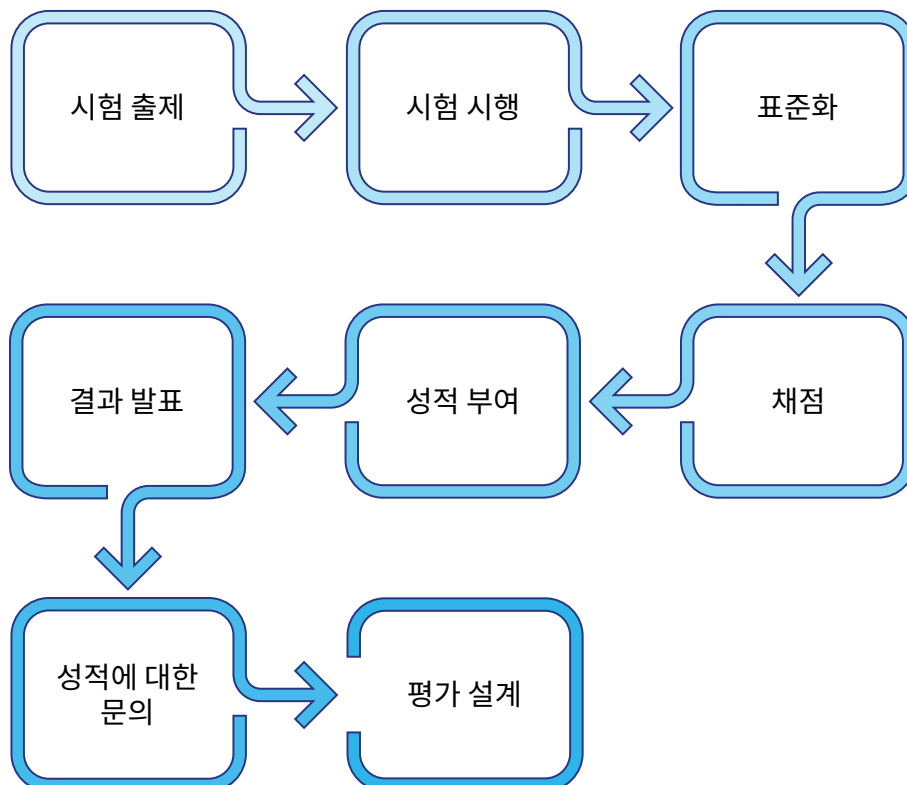
평가 주기

- 평가 제작 과정은 각 단계가 이전 단계의 영향을 받아 다음 단계로 이어지는 하나의 연속적인 주기로 이해해야 합니다.
- 시험(및 이에 해당하는 채점 기준표)을 제작하는 데 평균적으로 18개월이 소요되어 IB는 서로 다른 기수의 시험을 병행하여 제작합니다.

평가의 전 과정은 제작에서 시작하여 학생 응시, 채점관의 채점, 결과 발표까지 하나의 주기로 볼 수 있습니다. 이 과정에서 중요한 측면은 IB가 각 기수에서 얻은 경험과 학습을 바탕으로 다음 기수의 평가 질을 향상시킨다는 것입니다.

[그림 25]는 이러한 과정을 설명하는 여러 방식 중 하나입니다. 여러 단계를 다양한 방식으로 분리하거나 결합할 수 있지만, 이 도식은 평가의 전반적인 주기를 이해하는 데 훌륭한 틀을 제공합니다. <표 6>에는 각 단계가 요약되어 있으며, 해당 단계와 관련된 섹션으로 연결되는 링크가 제공됩니다.

[그림 25] 평가 주기



<표 6> 평가 전 과정의 단계

평가 주기의 단계	설명
시험 출제	각 개별 시험을 출제하는 과정입니다. 이 과정에는 문제에서 다룰 주제 결정부터 구체적인 문제의 작성과 편집, 여러 언어로 번역, 올바른 형식으로 구성, 최종적으로 품질 평가까지 모든 것을 포함합니다.
시험 시행	학생이 학교에서 시험을 치르는 단계입니다.
표준화	선임 채점관이 채점관 팀에게 학생 과제를 어떻게 채점해야 하는지를 설명하고, 채점 품질 보증 과정에 사용될 '확정 채점' 답안지를 선별하는 과정입니다.
채점	채점관이 개별 학생의 과제를 검토하고 부여할 점수를 결정하는 과정입니다. 채점관은 책임 채점관이 설정한 지침을 반드시 따라야 하며, 이 지침이 정확하게 적용되고 있는지 확인하기 위해 정기적인 모니터링이 이루어집니다.
성적 부여	선임 채점관은 점수(시험에 따라 달라짐)를 어떻게 성적(항상 동일한 의미를 지님)으로 환산할지를 결정합니다. 성적 부여 이후에는, 채점관의 성적 분포를 모니터링하고, 특정 문제와 관련된 증거를 바탕으로 잘못된 성적을 받을 위험이 있는 학생을 식별하는 등 후속 활동이 포함됩니다.
결과 발표	학교와 학생에게 결과를 발표하는 단계입니다.
성적에 대한 문의	학교가 시험 과정에서 오류가 있다고 판단하는 경우, 해당 사항을 제기하고 IB에 학생의 과제를 다시 검토해 줄 것을 요청할 수 있는 기회입니다.
평가 설계	학생의 성과와 시험 문제에 대한 해석 방식은 평가 설계에 중요한 근거 자료가 됩니다. IB는 이러한 정보를 활용하여 무엇을 평가할 것인지에 대한 목적과 접근 방식, 그리고 과제의 수와 유형 등을 개선합니다.

평가 주기에 관한 자세한 내용은 “[The life cycle of an exam: The steps that we go through from an exam being created to grades being awarded](#)”(영문 제공, 동영상)를 참고하십시오.

디지털 평가가 평가 주기에 미치는 영향

디지털 평가의 도입은 평가 주기의 원칙 자체를 변화시키지는 않습니다. 다만, 평가 주기의 일부 단계에서는 IB가 보다 신속하게 업무를 수행할 수 있도록 합니다(예: 시험지와 답안지를 스캔 센터에 보내지 않아도 됨). 그러나 평가 주기의 각 단계는 여전히 모두 수행되어야 합니다.

평가 제작

- 학생이 평가를 치르기까지, 평가 제작 과정에서 여러 단계의 최종 검토를 거쳐야 합니다. 여기에는 다음과 같은 과정이 포함됩니다.
 - 평가 출제 및 내용 승인
 - 구성 및 교정
 - 편역성
 - 번역
 - 학교에 평가 전달
- 학생이 요청한 모든 수정 사항(포용적인 접근성 조치) 또한 이 최종 단계에서 함께 고려됩니다.
- 서면 기반 시험이든 디지털 시험이든, 평가 제작 주기의 원칙은 변하지 않습니다.

우수한 평가 항목

‘섹션 A: 평가 원칙’에 명시된 좋은 평가의 자질과 더불어, IB는 교육과정 개발자와 출제 팀이 다음의 핵심 요소들을 염두에 둘 것을 권장합니다.

- 평가는 학생이 할 수 없는 것을 찾기보다, 학생이 할 수 있는 것을 보여줄 기회를 제공해야 합니다.
- 평가는 가능한 한 넓은 범위의 학생이 접근할 수 있어야 하며, 동시에 합리적인 변별이 가능해야 합니다.
- 가장 우수한 평가는 모든 학생이 접근할 수 있는 평가입니다. IB는 평가 제작 전반에 걸쳐 보편적 설계를 적용하여, 가능한 한 포용적인 접근성 조치의 필요성을 줄이는 것을 목표로 합니다.
- 평가는 반드시 교과 가이드에 명시된 교육과정만을 평가해야 합니다.
- 평가 개발 중 채점 방식을 고려함으로써 평가하려는 내용이 실제 점수로 인정될 수 있도록 해야 합니다. 이는 시험과 더불어 채점 기준표도 함께 개발되고, 채점관을 위한 예상 허용 오차에 관한 지침이 포함됨을 의미합니다.

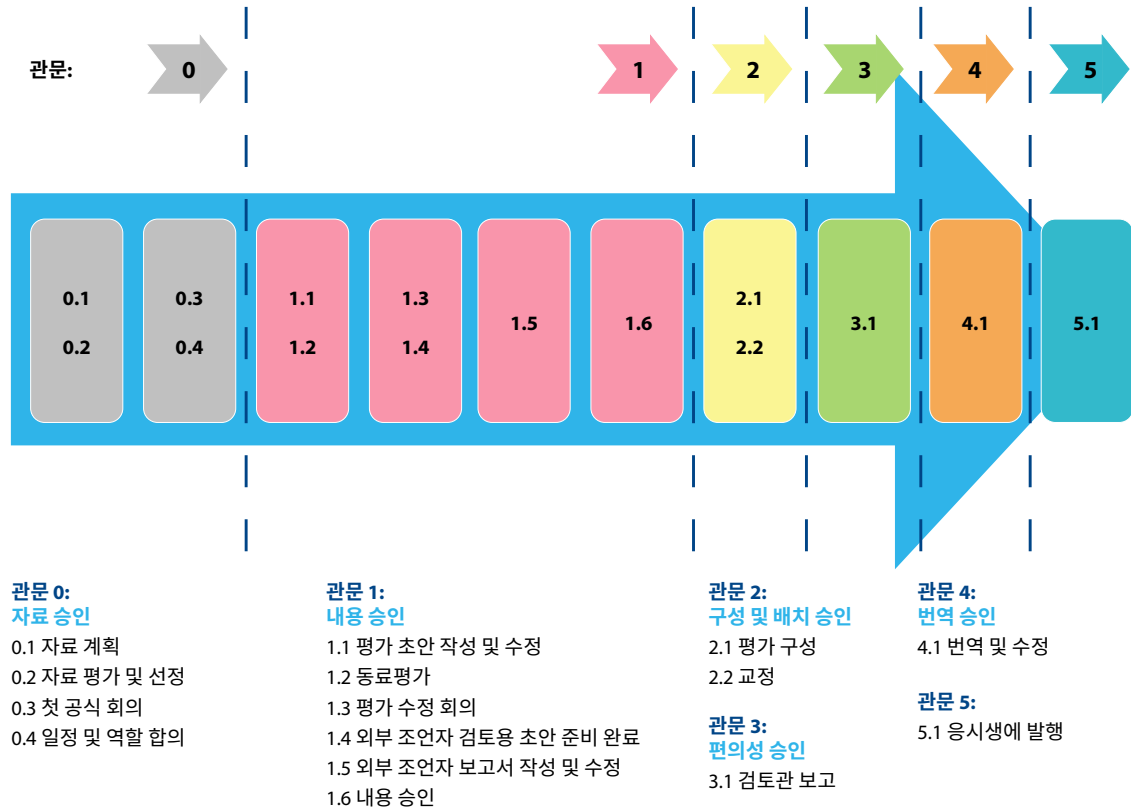
지시어

지시어는 교수요목 내용과 시험 문제에서 사용되는 핵심 용어 및 구문으로, 특정 지시에 대한 답변에서 무엇을 요구하는지를 나타냅니다. 이는 학생이 작성해야 하는 답안의 유형과 수준도 시사합니다.

지시어가 다른 언어로 번역될 경우 의미의 미묘한 차이가 사라질 수 있어 특별한 주의를 요합니다. 이를 방지하기 위해 IB는 각 시험 언어별로 교과 가이드에 지시어를 제시하고, 각각의 의미를 설명합니다. 번역된 시험을 검토할 때의 주요 책임 가운데 하나는 알맞은 지시어가 사용되었는지 확인하는 것입니다.

시험 출제 과정 개요

[그림 26] 시험 출제 과정



[그림 26]은 시험 출제 과정의 여러 단계를 보여줍니다. 평가를 설계하는 일은 약 18개월이 소요되는 장기적인 과정으로, 시험 출제자, 과목 담당자, 교과 전담 교사, 편집자, 설계자, 외부 검토관, 언어 학자 등 광범위한 외부 전문가와 IB 직원이 참여합니다.

시험 출제와 동시에 채점 기준표를 개발하는 것은 매우 중요합니다. 이를 통해 시험 문제와 채점 기준이 평가 의도를 완전히 일치시킬 수 있습니다.

내용 승인까지의 과정

시험 출제의 첫 번째 단계는 시험 출제자를 선정하는 것입니다. 이는 보통 책임 채점관에게 요청되지만 IB는 학교와의 연관성이나 연수 워크숍 운영 책임 등 이해의 상충으로 이어질 수 있는 모든 요소를 함께 고려합니다.

시험 출제에 관여하는 역할은 다음과 같습니다.

- 출제자(외부)
- 외부 조연자/표준화 담당자(외부)
- 검토관/언어 능숙자(외부)
- 과목 담당자(IB 직원)
- 제작 편집자(IB 직원)

- 교열 담당자(IB 직원)
- 설계자(IB 직원)
- 번역가(외부)
- 언어 교정자(외부)

IB가 여러 표준시간대를 대상으로 시험을 출제하는 경우, 각 시간대별로 완전히 독립된 출제 과정을 거칩니다.

출제 및 내용 승인

출제자는 시험과 채점 기준표의 초안을 작성해야 합니다. 이 초안은 이후 다른 출제자 및 과목 담당자와 함께 여러 단계의 검토 과정을 거칩니다. 일부 과목에서는 이후 공식 평가 수정 회의를 통해 다음 측면에서 각 시험을 검토합니다.

- 문제의 인지적 요구 수준 및 시험의 전반적인 난이도
- 발생 가능한 편향
- 접근성
- 교육과정 범위
- 발행된 평가 모델과 시험이 일치하는지 여부
- 과거 시험 및 샘플 자료와의 연관성
- 문제들이 공개된 자료의 예시 문제와 동일하지 않은지 여부

이 회의를 통해 또 하나의 완성된 초안이 도출되며, 이는 이후 외부 조연자 또는 표준화 담당자에 의해 검토됩니다.

외부 조연자/표준화 담당자는 시험 출제 및 채점 기준표 작성에 참여한 적이 없는 또 다른 과목 전문가입니다. 이들은 새로운 관점을 제공하며 모든 시험 전반의 난이도 수준, 이전 시험과의 공통점 및 차이점, 교육과정과 일치하는 정도에 대한 의견을 제시합니다. 외부 조연자가 제안한 변경 사항을 출제자와 과목 담당자가 검토하여 최종안 작성에 반영합니다.

최종안은 내부 검토를 거치며, 필요한 경우 과목 담당자 및 출제자와 함께 교열 담당자 및 설계자의 전문적인 검토를 거칩니다.

시험 내용의 최종 승인은 출제자가 내리며, IB 과목 담당자는 시험이 IB의 품질 기준에 따라 적절히 검토되었는지 확인합니다.

채점 기준표는 이 단계에서 확정되지 않고, 표준화 과정에서 실제 학생 답안을 바탕으로 최종 검토됩니다.

구성 및 교정

내용이 승인되면 문제를 구성하는 텍스트가 확정됩니다. 이후 설계자는 시험이 시행될 형식(서면 기반 또는 디지털)에 맞추어 문제를 구성합니다.

전통적인 서면 기반 시험의 경우, 지침(또는 평가 기준표)이 포함된 표지를 추가하고, 정해진 형식 규칙을 적용하며, 여러 바코드와 빈 페이지를 포함해 시험 책자를 제작합니다. 이 과정에는 품질 수준에 맞게 도표를 다시 그리거나 적절한 이미지를 온라인에서 확보하는 작업도 포함됩니다.

디지털 평가의 경우 설계자가 적절한 개발 환경에서 시험을 제작해야 하며, 여기에는 지침/평가 기준표 및 적절한 편집 형식이 포함됩니다.

모든 경우에 제작 편집자가 승인된 내용과 대조하여 완성본을 교정합니다.

편의성 승인

검토관은 이 시점까지 아직 평가 제작에 참여하지 않은 또 다른 과목 전문가입니다. 검토관은 문제에 답하는 데 소요되는 시간까지 고려하여 학생의 입장에서 실제로 시험을 치르듯이 평가를 수행합니다. 이 과정의 목적은 오류를 발견하고 지침이나 문제에 모호한 표현이 있는지 확인하는 것입니다.

검토관은 평가 수행 후 채점 기준표도 함께 검토합니다.

응시하는 학생 수가 적은 언어 과목에서는 문학 전문가보다 제2 언어를 유창하게 구사하는 또 다른 전문가가 검토관 역할을 수행하기도 합니다.

검토관의 공식 피드백을 반영되면 평가는 최종 확정되며 학교로 안전하게 발송될 준비를 마치게 됩니다.

품질 관리

시험에 오류나 모호한 요소가 있을 경우, 시험을 치르는 학생에게 중대한 영향을 미칠 수 있으므로, 완벽한 품질 확인이 필수입니다. IB는 이러한 문제를 매우 엄중하게 다룹니다.

공식적인 승인 절차 외에도 평가 제작 주기의 각 단계마다 품질 점검 절차가 병행됩니다.

번역

IB는 학생에게 다양한 시험 언어로(비언어) 시험을 볼 수 있는 기회를 제공합니다. 현재 대부분의 과목은 영어, 프랑스어, 스페인어로 제공됩니다. 현지 협정에 따라, 일부 특정 과목은 독일어, 일본어, 한국어로도 제공됩니다.

시험의 번역 필요성은 초안 단계부터 고려되며, IB 출제 팀과 교열 담당자는 번역 과정에서 발생할 수 있는 문제를 충분히 인지하고 있습니다. 공식적인 번역 절차는 영어 시험이 최종 확정된 이후 시작됩니다.

특정 언어로 시험을 치른 학생이 유리하거나 불리해지지 않도록, 번역 과정에서 문제의 의미가 바뀌지 않는 것이 무엇보다 중요합니다. 이를 위해 IB는 교과 전담 교사를 외부 번역가로 고용하여 전문 용어 정확성을 보장합니다. 이후 이중 언어를 구사하는 외부 교정자가 번역본을 원본 시험과 비교하여(단순한 텍스트 비교가 아니라 시험 전체를 기준으로) 최종 품질 점검을 수행합니다.

특수 변형 시험지

가장 이상적인 상황은 모든 학생에게 적합한 평가를 설계하는 것입니다. 그러나 보편적 설계를 반영하는 것만으로는 충족할 수 없는 특정한 요청이나 요구 사항도 존재하며, 이러한 경우에는 특수 변형 시험지를 제공해야 합니다. 이는 「접근성 및 포용성 정책」에 따라 관리되며, 일반 평가와 동일한 절차로 제작되어 학교에 전달됩니다.

시험

- 시험의 목적은 모든 학생에게 일관된 경험을 제공하고 학업적 부정행위의 가능성을 최소화하는 통제된 환경에서 학생이 지식, 이해, 기능의 적용 능력을 입증할 수 있도록 하는 데 있습니다.
- 시험 기수를 설정할 때 IB는 학생의 기대에 부합하기 위해 가능한 빨리 채점을 완료하면서, 모든 학생의 요구와 학교의 관리 가능성을 균형 있게 고려해야 합니다.
- IB는 학업적 부정행위의 가능성을 최소화하기 위해 시험장 행동 규칙을 명확히 공표합니다. 이러한 규칙은 새로운 테크놀로지와 환경 변화에 따라 지속적으로 업데이트됩니다.
- 시험 일정은 여러 상충되는 우선순위 간의 타협점으로, 전체적으로 볼 때 가장 나은 선택지를 반영합니다.
- 예기치 못한 상황, 포용적인 접근성 조치, 불리한 상황 및 시험 일정 변경에 관한 자세한 정보는 관련 프로그램의 「평가 절차」에서 확인할 수 있습니다.

시험 응시 기간은 학생에게 심리적으로 매우 부담스럽고 긴장되는 시기입니다. 이 기간 동안 학생에게 가해지는 스트레스를 최소화하기 위한 IB의 주요 원칙은 다음과 같습니다.

- 시험 기수의 기간을 제한하기
- 가능한 경우, 학생이 응시할 수도 있는 다른 시험(예: 국가 시험)과의 일정이 겹치지 않도록 조정하기

이는 반드시 학생의 성과에 대한 타당한 결론을 도출할 수 있을 만큼 충분한 평가를 수행해야 할 필요성과 학업 부정행위를 방지해야 할 필요성 사이에서 균형을 이루어야 합니다.

시험장 준비 및 관리하기

학교는 반드시 관련 프로그램의 「IB 시험 시행 지침서」에 명시된 엄격한 규정에 따라 시험을 치러야 합니다.

시험 보안과 적절한 시행은 매우 중요합니다. 학교와 이해관계자 사이에 학업 부정행위가 발생했다는 인식이 생기면 학생 결과의 신뢰도가 크게 훼손되기 때문입니다. IB는 본 간행물의 ‘평가 공정성’ 섹션에 명시된 바와 같이, 국제적 신뢰를 유지하기 위해 다양한 정책과 절차를 마련해 두고 있습니다.

불가피하게 학교와 학생이 통제 밖의 사유로 IB가 채점해야 할 학생 과제에 접근할 수 없는 상황이 발생하기도 합니다. 이러한 경우 IB는 학생에게 미치는 영향을 완화하기 위해 학생의 점수를 추정할 수도 있으나, 이는 참고할 수 있는 다른 근거 자료가 있을 때에만 가능합니다. 서면 기반 시험의 경우 이러한 상황에 대비하기 위해, IB는 서로 다른 답안지를 서로 다른 날에 스캔 센터에 발송하여 모든 답안지가 분실될 가능성을 줄이고 있습니다.

디지털 평가의 경우, 업로드 과정의 각 단계마다 평가용 제출물의 사본이 남기 때문에, DP/CP 환경에서는 이러한 위험이 상당 부분 완화됩니다.

시험 일정 수립하기

시험 기수 기간을 최소화하기 위해 IB는 시험 일정에서 서로 다른 과목의 시험을 동일한 시간대에 배정합니다. 학생이 선택할 수 있는 과목 조합을 고려하기 위해 모든 노력을 기울이고 있지만, IB는 일부 일정 중복이 불가피하다는 점을 인지하고 있으며, 이에 대한 처리 절차는 관련 프로그램의 「평가 절차」에서 확인할 수 있습니다.

가능한 경우, 하루 동안 치르는 시험 시간이 총 6시간 30분을 초과하지 않도록 시험 일정을 수립합니다. 특정 과정의 다중 항목 시험은 일반적으로 2~3개의 시험이 연속으로 진행되며, 오후에 시작하여 다음 날 오전에 종료되는 일정입니다. 항상 가능한 것은 아니지만 한 과정의 모든 시험을 같은 날에 몰아 치르는 것보다는 이 방식이 더 바람직합니다.

이러한 방식은 대부분의 학생에게 2주(MYP) 또는 3주(DP/CP) 일정에 걸쳐 시험을 더 고르게 분산시킵니다. 이를 통해 학생은 특정 시험에서 자신의 실력을 충분히 발휘하지 못했다고 느낄 경우, 하룻밤 동안 컨디션을 가다듬고 다음 시험을 준비할 수 있는 기회를 제공합니다.

시험 일정 설계 원칙

다음은 시험 일정 수립의 기본 원칙을 요약한 내용입니다.

이 모든 원칙을 충족하는 것이 불가능할 때도 있으며, 그러한 경우에는 서로 상충하는 원칙 간의 절충이 필요합니다. IB는 시험이 시행되기 최소 1년 전에 시험 일정을 발표합니다.

- 많은 국가와 지역(영역)에서 IB 프로그램을 시행하고 있기 때문에 5월 1일을 제외한 공휴일, 국경일, 방학을 모두 시험 일정에 반영하는 것은 불가능합니다.
- 중등 지역 학교의 주말이 목요일 또는 금요일이라는 점을 고려하여 해당 요일에 시험을 배정하지 않는 것이 바람직하나, 현실적인 제약으로 인해 현재로서는 이를 전면적으로 적용하기 어렵습니다.
- 특정 지역적 또는 문화적 연관성을 지닌 과목이 있는 경우, IB는 시험 일정 수립 시 이러한 요소를 고려하기 위해 노력합니다. 예를 들어 아랍 문학이나 언어 시험은 금요일에 배정하지 않습니다.
- IB는 학생 등록 자료에 포함된 과목 조합 데이터를 활용하여, 전 세계적으로 과목 간 시험 일정이 겹치는 학생 수를 최소화하고자 합니다.
- 학생은 같은 날에 서로 다른 두 개의 외국어(시험 언어가 아닌 언어) 시험을 치르지 않아야 합니다.
- 여러 시험으로 구성된 과정의 경우, 예기치 못한 사건으로 학생이 모든 항목에서 불이익을 받을 위험을 최소화하고자 시험은 최소 이틀에 걸쳐 배정됩니다.
- 학생들은 각 과목 복습을 집중할 수 있도록 충분한 시간을 확보해야 합니다. 시험이 여러 날에 걸쳐 실시되는 경우 가능한 한 이틀 연속으로 시험을 배정할 수 있습니다.
- 학생의 시험은 짧은 기간에 집중되기보다는, 시험 기간 전체에 걸쳐 고르게 분산되어야 합니다. 이는 응시생 수가 가장 많은 과목(예: 영어, 역사, 수학)은 가능한 한 연속으로 배정하지 않습니다.
- 이와 같은 이유로 IB는 언어와 과학 시험을 시험 기간인 2주(MYP) 또는 3주(DP/CP) 동안 서로 다른 주에 배정하려 합니다.
- IB의 내부 시험 처리 절차상 특정 과목, 특히 응시자가 많은 과목은 일반적으로 시험 일정 초반에 배정됩니다.

예기치 못한 상황에 대처하기

대부분의 예기치 못한 상황에 대한 대응 방법은 관련 프로그램의 「평가 절차」에 명시되어 있습니다. 의문이 있을 경우 [IB 지원팀](#)에 문의하여 조언을 얻을 수 있습니다.

포용적인 접근성 조치 및 불리한 상황

이 주제는 본 간행물의 ‘모든 학생을 위한 공정성’ 섹션에서 자세히 다루어집니다.

IB는 모든 학생이 가능한 한 공정한 평가 조건에서 자신의 능력을 보여줄 수 있어야 한다고 믿습니다. 그러나 표준 평가 환경에서는 학습 지원이 필요한 학생이 자신의 학업 성취도를 온전히 보여주지 못해 불이익을 받을 수 있습니다. 이러한 경우 시험 시행 방식에 대한 특수 편의 조치나 특수 변형 시험지 제공과 같은 포용적인 접근성 조치가 승인될 수 있습니다.

시험 기수 중 발생하고, 학생이 통제할 수 없으며, 성과에 부정적인 영향을 줄 수 있는 사건을 ‘불리한 상황’이라고 합니다. 불리한 상황에는 일시적인 질병 또는 부상, 심각한 스트레스, 이례적으로 어려운 가정 상황, 가족 및 친지의 사망, 또는 학생의 건강이나 안전을 위협할 수 있는 사건 등이 포함됩니다. 또한 소요 사태, 자연재해처럼 학교 공동체 전체에 영향을 미치는 사건도 불리한 상황에 해당할 수 있습니다. 반대로 학교 측의 과실로 발생한 상황은 불리한 상황에 포함되지 않습니다.

이와 같은 조치에 관한 자세한 내용은 관련 프로그램의 「평가 절차」를 참고하십시오.

시험 일정 변경하기

시험 일정을 변경하는 것은 시험 과정의 공정성에 심각한 위험을 초래할 수 있습니다. 이는 일부 학생이 다른 학생보다 먼저 또는 나중에 시험을 치르게 되는 상황을 초래할 수 있으며, 그 결과 학업적 부정행위로 이어질 가능성이 커집니다. 이러한 이유로, 예정된 시험일을 기준으로 하루 전부터는 시험 일정 변경이 절대 허용되지 않습니다. 이는 해당 시험을 치르는 대다수 학생에게 영향을 미칠 위험이 지나치게 크기 때문입니다.

IB가 학생에게 IB 시험 일정과 다른 시간 및/또는 날짜에 하나 이상의 시험을 치르도록 승인할 수 있는 유일한 상황과, 시험 일정 변경 요청 절차는 [프로그램 자료실](#)의 「평가 절차」 및 「시험 준비 정책」에 상세히 규정되어 있습니다.

채점

- 성공적인 채점이란 학생 과제의 수준을 반영하여 일관되고 정확한 점수를 학생에게 부여하는 것을 의미합니다.
- 책임 채점관이 충분한 검토를 거쳐 설정한 판단 기준이 정확한 기준이 되며, 다른 모든 채점관은 반드시 이 기준을 적용하여 채점해야 합니다.
- 점수와 성적은 다릅니다. 학생이 더 쉬운 시험에서 더 높은 점수를 받을 수 있지만 그렇다고 해서 더 높은 성적을 받는 것은 아닙니다.

점수와 성적

학생의 총괄 평가를 시행하고 그 결과를 활용할 때에는 과제에 점수를 매기는 것과 성적을 부여하는 것의 차이를 이해하는 것이 매우 중요합니다.

- 채점은 채점 기준표 또는 이와 유사한 프레임워크에 따라 학생의 과제에 점수를 부여합니다. 이는 학생이 평가 과제를 얼마나 올바르게 수행했는지를 나타내는 지표입니다. 점수 그 자체에는 다른 의미가 내포되어 있지 않습니다.
- 반면 성적을 결정할 때는 채점관이 과제의 난이도와 완성도를 고려하여 정의된 기준에 따라 학생 과제의 질을 판단합니다. 따라서 성적은 어느 정도 의미와 관련성이 있으며, 일반적으로 다른 평가의 성과와 비교 가능하도록 설정됩니다.

난이도가 매우 높은 문제의 일부만 맞혀도 높은 성적을 받을 수 있는 경우가 있는 반면, 마찬가지로 난이도가 낮은 문제를 많이 맞혀도 이와 동일한 성적을 받을 수 있는 것은 아닙니다.

이후 섹션에서 다루어지듯이, 특정 성적이 의미하는 기준을 반드시 학생의 성과와 연관 지어 설명할 필요는 없으나 IB는 이러한 접근 방법을 사용해 왔습니다. 학생이 동료와 비교해 어느 정도 성과를 내는지에 따라 기준을 설정하는, 매우 일관적이며 신뢰할 수 있는 다른 시스템도 존재합니다.

IB 평가에서 점수는 일반적으로 학생의 전반적인 성과를 나타내는 지표로 사용됩니다. 그다음 해당 점수를 받은 학생들의 성과를 살펴본 후 그 점수 이상을 받은 학생에게 특정 성적을 부여하기 위한 경계점(성적 급간)을 설정합니다. 이 과정에 대한 자세한 설명은 '성적 부여 및 종합' 섹션에서 더욱 자세히 설명됩니다.

채점 접근 방법

IB는 다양한 평가 도구를 사용하므로, 각 평가에 적합한 유형의 채점 접근 방법을 선택하는 것이 매우 중요합니다. 채점 접근 방법에 대한 결정은 교육과정 개발 과정에서 이루어집니다.

분석적 채점 기준표

분석적 채점 기준표는 학생에게서 예상되는 답안의 범위가 좁은 평가에 사용됩니다.

이 채점 기준표는 문제의 총점을 답안의 여러 부분에 어떻게 배분할지에 대한 구체적인 지침을 채점관에게 제공할 수 있습니다. 매우 구체적인 답안이 요구되는 구조화된 문제에서도 채점 기준표는 학생들이 선택할 수 있는 다양한 접근 방식과 흔히 발생하는 오류를 고려하여 채점관이 일관되게 채점할 수 있을 만큼 충분한 정보를 제공해야 합니다. 예상치 못한 답안이나 대안적인 타당한 답안에 점수를 부여할 때 채점관의 전문적 판단이 불가피하지만, 채점 기준표는 반드시 이러한 판단을 어떻게 적용해야 하는지에 대해 가능한 한 상세한 지침을 제공해야 합니다.

학생은 문제의 어느 부분을 맞히고 틀리는지 예측 가능한 패턴을 따르지 않습니다. 이러한 문제는 특히 확장형 구조화 문제에서 발생하며, 앞부분의 실수가 이후 부분에 영향을 미칠 수 있습니다. IB는 학생이 문제의 초반 부분에서 실수를 하더라도 나머지 문제를 계속 풀 수 있도록 문제를 설계합니다. 그럼에도 분석적 채점 기준표는 특정 유형의 오답을 어떻게 채점해야 하는지, 학생이 문제의 일부에서 실수를 했을 때 풀이 과정을 어떻게 이어서 평가할 것인지에 대한 명확한 지침을 반드시 제공해야 합니다.

평가 기준

상세한 채점 기준표가 부적합한 평가에는 평가 기준표를 적용합니다.

평가 기준표는 답안의 상세 내용과 관계없이 학생이 보여주어야 할 성과 유형에 더욱 초점을 맞춥니다. 학생이 선보일 수 있는 다양한 수준의 전문성은 단계 기술어에 반영됩니다.

IB는 평가 기준별 단계 기술어를 적용할 때 ‘최적화’ 모델을 사용합니다. 평가 기준표를 적용하는 채점관은 반드시 채점하는 과제의 전반적인 성취도에 가장 잘 맞는 성취 단계를 선택해야 합니다. 특정 성취 단계가 부여되기 위해 그 단계에 제시된 모든 세부 측면이 모두 충족되어야 하는 것은 아니며, 어떤 기준에서도 가장 높은 단계가 완벽함을 나타내는 것은 아니라는 점에 유의해야 합니다.

평가 기준표를 사용하는 대부분의 평가 항목에서는 하나의 기준이 아니라 여러 기준이 함께 적용됩니다. 이때 각 기준은 서로 독립적이어야 합니다. 학생이 동일한 답변으로 여러 평가 항목에서 점수를 받는 것은 공정하지 않습니다.

일반적으로 출제 문제는 매년 달라지더라도, 동일한 평가 기준표가 매년 사용됩니다. 이는 평가의 본질적 내용은 변하지 않기 때문입니다. 해마다 달라지는 문제는 대체로 기준표에 명시되지 않은 구체적인 세부 사항과 관련됩니다. 책임 채점관은 각 문제에 평가 기준표를 어떻게 적용해야 하는지에 대한 지침과, 고려해야 할 관련 세부 내용의 예시를 담은 채점 노트를 제공하는 경우가 많습니다.

전체론적 기준: 점수 구간표

경우에 따라, 하나의 과제를 채점하기 위해 서로 다른 평가 기준을 분리하여 적용하는 것이 적절하지 않을 수 있습니다. 이는 평가 기준 간에 독립적인 구분이 불가능한 경우에 주로 발생합니다. 이러한 경우 독립된 기준 대신 점수 구간표를 사용합니다. 점수 구간표는 사실상 과제 전체를 하나의 통합된 기준으로 평가하는 것을 의미합니다. 학생의 성과를 변별하기 위해 합리적인 점수 범위가 필요하므로, 각 점수 구간의 단계 기술어는 특정 점수 범위와 연결됩니다.

<표 7> IB 점수 구간표 예시

점수	단계 기술어
0	아래 기술어에서 명시한 기준을 충족하지 못합니다.
1~3	<ul style="list-style-type: none"> 이슈에 대한 이해를 거의 보여주지 못합니다. 과목별 어휘를 사용하지 않거나 지속적으로 부적절하게 사용합니다. 이슈에 대한 설명이 매우 부족합니다. 논점이 피상적이며 자주 불분명합니다. 답변이 설명 위주입니다. 제시한 분석이 피상적이거나 일관되지 않습니다. 다양한 관점에 대한 논의가 거의 없거나 전혀 없습니다. 결론이 포함된 경우, 매우 피상적이거나 답변의 나머지 부분과 일관되지 않습니다.
4~6	<ul style="list-style-type: none"> 이슈에 대한 기본적인 이해를 보여줍니다. 과목별 어휘를 사용하지만 대개 부적절하게 사용합니다. 이슈에 대한 설명은 기본 수준이며 충분히 전개되지 않았습니다. 논점이 대개 부정확하거나 모호하며, 무엇을 전달하려는지 대개 불분명합니다. 제시한 분석이 제한적이며 답변이 전반적으로 분석보다는 설명에 치중되어 있습니다. 다양한 관점에 대한 논의가 제한적입니다. 단순한 결론이 포함되어 있습니다.
7~9	<ul style="list-style-type: none"> 이슈에 대한 이해를 어느 정도 보여줍니다. 과목별 어휘를 때때로 적절하게 사용합니다. 특정 이슈에 대한 설명이 만족스러운 수준이지만, 부분적으로 명확성과 전개가 부족합니다. 적절한 논점을 제시하지만 정확성과 상세 설명이 부족합니다. 답변에 분석이 포함되지만 전개가 부족합니다. 다양한 관점에 대해 어느 정도 논합니다. 결론이 포함되어 있습니다.
10~12	<ul style="list-style-type: none"> 이슈에 대해 우수한 이해를 보여줍니다. 과목별 어휘를 대부분 적절하게 사용합니다. 이슈에 대한 설명이 명확하지만 추가 전개가 필요합니다. 논점이 적절하고 정확하지만 상세 설명이 부족합니다. 답변에 비판적 분석이 포함되어 있지만 전개가 부족합니다. 다양한 관점에 대해 논합니다. 답변은 제시된 주장과 일관된 결론으로 이어집니다.
13~15	<ul style="list-style-type: none"> 이슈에 대해 매우 우수한 이해를 보여줍니다. 과목별 어휘를 정확하고 정밀하게 사용합니다. 이슈에 대한 설명이 명확하고 잘 전개되었습니다. 논점이 적절하고 정확하며 상세합니다. 답변에 잘 전개된 비판적 분석이 포함되어 있습니다. 다양한 관점에 대해 비판적으로 논합니다. 답변은 제시한 주장과 일관되며 명확하게 제시된 합리적 결론으로 이어집니다.

기술어는 그 자체로 상당히 길고 서술적이며, 학생 과제에서 드러나는 다양한 잠재적 자질을 다루고 과정 목표와 직접적으로 연관됩니다. 평가 기준과 마찬가지로 최적화 모델이 적용됩니다. 채점관은 학생 과제가 각 단계 기술어에 얼마나 부합하는지를 판단하여, 해당 단계에 설정된 점수 범위 내에서 점수를 결정합니다. 예를 들어 하나의 점수 구간은 7~9점 범위를 포함할 수 있습니다. 채점관은 학생 과제가 점수 구간표의 해당 단계 기술어에 얼마나 부합하는지를 고려하여 그 범위 내에서 점수를 부여합니다.

자동 채점

선다형 문제로 구성된 일부 평가 요소는 자동으로 채점됩니다. 자동 채점이란 사전에 정의된 채점 기준표에 따라 테크놀로지를 활용하여 학생의 과제를 평가하는 과정입니다. 이러한 선다형 문제의 경우, 채점 기준표에서 정답 여부가 객관적으로 명시되어 있습니다. 이러한 항목의 채점 기준표는 매우 명확하며 채점관의 판단이 개입될 필요가 없습니다.

채점관의 판단이 필요한 채점

일부 평가는 자동 채점이 가능하지만 분석적 채점 기준표 또는 채점 기준을 적용해야 하는 대부분의 IB 평가는 채점관이 직접 채점해야 합니다.

채점관은 ‘채점 도구(marking tool)’라 불리는 소프트웨어를 사용하여 학생 과제를 전자 방식으로 채점합니다. 이를 통해 오디오 및 시각 자료를 포함한 학생 과제의 전자 사본에 정답 표시, 점수 및 메모 등을 남길 수 있습니다. 이러한 방식은 공연 녹화나 구술 시험 녹음과 같이, 평가 근거로 제출할 수 있는 과제의 범위를 한층 넓혀 줍니다.

IB 채점 방식에 관한 자세한 내용은 “[The two methods of marking: How marking is carried out by examiners and teachers](#)”(영문 제공) 동영상을 참고하십시오.

표준화

표준화 과정에서는 평가 절차의 여러 요소를 포괄합니다. 표준화의 목적은 채점관이 신뢰성 있게 채점할 수 있도록 준비시키는 데 있습니다.

신뢰할 수 있는 결과를 얻으려면, 어느 채점관이 채점하더라도 각 답안지가 담고 있는 과제의 질에 따라 동일한 점수를 받는 것이 중요합니다. 이를 위해 채점관은 실제 IB 시험 기수에서 채점을 시작하기 전에, 각 항목에 대해 책임 채점관이 설정한 채점 기준을 이해하고 이와 일치함을 입증해야 합니다.

표준화 과정의 첫 번째 단계는 채점에 적합한 기능과 경험을 갖춘 채점관을 채용하는 것입니다. 이후 채점관은 채점의 원칙과 실제 수행 절차를 이해하기 위한 교육을 받습니다.

시험이 실시된 후에는 책임 채점관의 주도하에 선임 채점 팀이 모여 채점 기준표에 대해 논의하고 표본 답안지를 검토합니다. 이 표준화 회의의 목적은 채점 기준표가 채점관이 실제 접할 것으로 예상되는 다양한 답안을 충분히 포괄하고 있는지 확인하는 것입니다. 채점 기준표가 학생이 제시할 수 있는 모든 가능한 답안을 포함하기는 어렵지만, 회의에서 샘플 답안을 검토함으로써 책임 채점관과 선임 채점 팀은 지금까지 확인된 과제를 기반으로 각 점수에 요구되는 답안 유형을 채점관에게 알리기 위한 명확한 지침과 안내를 마련할 수 있습니다.

표준화 회의의 결과로, 책임 채점관이 확정적으로 채점한 답안 묶음이 도출됩니다. 이 답안들은 다른 채점관이 책임 채점관과 동일한 기준으로 채점할 수 있도록 교육하는 데 활용됩니다. 책임 채점관의 채점 기준과의 일치를 성공적으로 입증한 채점관만이 실제 시험 기수에서 답안지를 채점할 수 있으며, 채점 기간 동안 엄정한 품질 보증 절차를 통해 채점의 일관성을 지속적으로 점검받습니다.

품질 모델

학생의 과제를 어느 채점관이 채점하더라도, 해당 학생은 책임 채점관이 채점했을 때와 동일한 결과를 받아야 한다는 점이 중요합니다.

평가가 시행되고 일부 학생 답안이 수집되면, 책임 채점관은 다른 선임 채점관들과 함께 표준화 회의를 진행합니다. 이 회의에서 선임 채점관들은 품질 모델을 구축하는 데 활용될 여러 학생 답안의 점수에 대해 합의합니다. 확정적으로 채점된 답안지는 연습 답안지, 채점 자격 취득용 답안지, 시드 답안지의 세 가지 범주로 분류됩니다. 이 세 유형의 답안지가 품질 모델을 구성합니다.

연습 답안지

연습 답안지의 목적은 책임 채점관이 설정한 채점 기준을 채점관이 학습할 수 있도록 지원하는 것입니다. 연습 답안지는 대표적인 답안 사례를 어떻게 채점해야 하는지를 보여 주며, 채점 기준표에 대한 구체적 이해가 요구되는 일반적인 상황들을 식별할 수 있도록 해야 합니다.

우수한 연습 답안 묶음에는 낮은 점수부터 높은 점수까지 다양한 수준의 학생 성과가 포함되어야 하며, 이를 통해 채점관은 어느 부분에 점수를 부여해야 하는지, 좋은(또는 좋지 않은) 답안이 어떤 특징을 지니는지를 파악할 수 있도록 해야 합니다.

채점관은 채점 기준표 또는 채점 기준, 그리고 표준화 회의 이후 책임 채점관이 제공한 모든 채점 노트를 검토한 다음, 연습 단계의 채점을 시도할 수 있습니다. 연습 답안지는 책임 채점관의 채점 기준을 나타냅니다. 채점관은 연습 답안지를 채점하는 과정에서 필요한 경우 팀 리더에게 지도 요청을 할 수 있습니다.

연습 답안지 채점을 마친 후, 채점관은 책임 채점관의 채점 기준을 이해하고 이를 적용할 수 있다는 확신을 가져야 하며, 이후 채점 자격 취득용 답안지로 넘어갈 수 있습니다.

채점 자격 취득용 답안지

채점 자격 취득용 답안지의 목적은 채점관이 올바른 채점 기준에 따라 채점할 수 있음을 입증할 기회를 제공하는 데 있습니다. 이 단계를 성공적으로 완료한 채점관만이 실제 시험 기수에서 채점할 수 있게 됩니다.

일반적으로 하나의 채점 자격 취득용 답안지 묶음은 다섯 개의 답안지로 구성됩니다. 채점관이 채점 자격 취득용 답안지를 채점하는 과정에서 책임 채점관의 점수와 일치하지 않을 경우, 어떤 점수를 부여했어야 하는지와 그 이유에 대한 피드백을 받게 됩니다. 그 후 채점관은 다른 채점 자격 취득용 답안지 묶음을 사용하여 채점 기준을 올바르게 이해했음을 다시 한번 입증할 기회를 갖게 됩니다.

우수한 채점 자격 취득용 답안지 묶음에는 낮은 점수부터 높은 점수까지 다양한 수준의 학생 성과가 포함되어야 합니다. 또한 책임 채점관과 선임 채점 팀이 채점 과정에서 발생할 수 있는 예외적인 상황에 대비해, 채점 기준표의 세부 지침을 반드시 숙지해야만 해결할 수 있는 사례들도 함께 포함해야 합니다. 이러한 답안지는 표준화 회의에서 답안을 검토한 결과를 토대로 선정됩니다. 채점 자격 취득용 답안지는 채점관의 실수를 찾아내기 위해 선정되는 것이 아니라, 실제 채점 과정에서 채점관이 접하게 될 가능성이 높은 전형적인 답안들로 구성되어 있습니다.

이 단계에서 채점관이 책임 채점관의 채점 기준을 이해하고, 평가를 위해 제출된 과제에 이를 적용할 수 있음을 보여 줄 두 번의 기회가 제공됩니다. 채점관은 채점 자격 취득용 답안지 묶음을 성공적으로 완료해야 합니다.

각각의 채점 자격 취득용 답안지 묶음은 여러 개의 확정적으로 채점된 답안지로 구성됩니다. 채점관은 이 답안지 묶음을 채점하는 동안 해당 답안지에 대해 팀 리더와 논의할 수 없습니다. 첫 번째 채점 자격 취득용 답안지 묶음에서 채점관이 부여한 점수가 허용 오차 범위 안에 있으면 실제 채점 단계로 넘어갑니다.

그러나 채점관의 채점 결과가 허용 오차 범위를 벗어날 경우, 팀 리더로부터 책임 채점관의 기준에 맞추어 채점을 조정하는 방법에 대한 피드백을 받습니다. 채점관은 해당 피드백을 검토하고 반영한 후, 채점 기준을 이해했음을 보여 줄 두 번째 기회를 갖게 됩니다.

두 번째 채점 자격 취득용 답안지 묶음에서 채점관의 채점 결과가 허용 오차 범위 내에 들면, 이제 실제 채점을 시작할 수 있습니다.

실제 채점 및 시드 답안지

시드 답안지는 품질 보증 과정의 일부입니다. 이는 채점관의 수행 능력을 모니터링하고 채점관이 계속해서 올바른 기준에 따라 채점하고 있는지 확인하는 데 사용됩니다.

시간이 지남에 따라 책임 채점관이 설정한 기준과 일치하던 채점관의 기준이 점차 벗어 수 있습니다. 이러한 이유로 IB는 책임 채점관이 이미 확정적으로 채점한 시드 답안지를 활용하여 채점관의 채점이 기준에 부합하는지 주기적으로 점검합니다. 시드 답안지는 다른 답안지와 동일하게 보이므로, 채점관은 자신이 품질 보증을 위한 답안지를 채점하고 있다는 사실을 인지하지 못합니다. 일반적으로 채점관은 자신이 채점하는 답안지 10개마다 무작위로 선정된 1개의 시드 답안지를 채점하게 됩니다.

채점관이 시드 답안지에 책임 채점관과 동일한 점수를 부여하면, 채점을 계속 이어갈 수 있습니다. 그러나 잘못된 점수를 부여하여 요구되는 기준에 따라 채점하지 않고 있음이 드러날 경우, IB가 개입하게 됩니다.

우선 선임 채점관의 피드백과 지도를 통해, 해당 채점관이 기준에 부합하는 채점 역량을 다시 확립할 기회를 부여합니다. 이러한 추가 지원에도 불구하고 기준에 맞게 채점하지 못할 경우, 더 이상 답안지를 채점할 수 없습니다.

각 항목에는 채점 허용 오차가 설정되며, 이는 확정된 기준으로부터 일정 범위 내의 차이를 허용합니다. 자세한 내용은 ‘[허용 오차](#)’ 섹션을 참고하십시오.

성공적인 표준화의 지표

표준화가 성공적으로 이루어질 경우 다음과 같은 결과가 나타납니다.

- 채점 관련 질의를 지원하고 채점 일관성을 보장할 수 있는 경험이 풍부한 선임 채점 팀 확보
- 점수가 부여되는 근거를 명확히 설명하는 채점 기준표 또는 채점 노트 형태의 명료한 지침 제공
- 교육 및 품질 보증을 위해 확정적으로 채점한 연습 답안지, 채점 자격 취득용 답안지, 시드 답안지

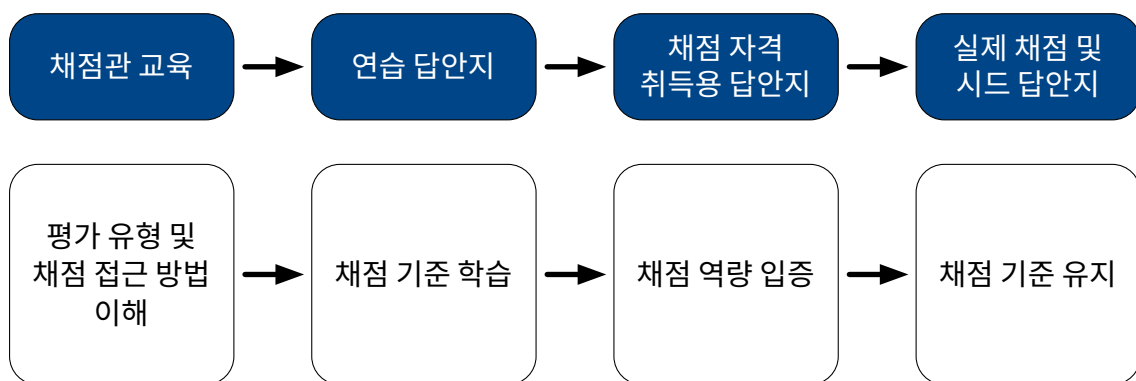
허용 오차

IB는 단순히 채점하기 쉬운 내용을 평가하는 것이 아니라, 중요한 내용을 평가하는 문제를 출제해야 한다는 신념을 가지고 있습니다. 정답과 오답이 명확히 갈리는 특정 유형의 문제에서는, 채점관이 책임 채점관과 완전히 동일한 점수를 부여할 것으로 기대하는 것이 타당합니다. 반면 이해력이나 분석 능력을 평가하는 문제의 경우, 일정 수준의 판단이 개입됩니다. 실제로 이는, 전문성과 문제 이해도가 대등한 두 채점관이라 하더라도, 동일한 과제에 정확히 같은 점수를 부여하지 않을 수 있음을 시사합니다. 두 채점관의 점수는 1~2점 정도 차이날 수 있습니다.

IB는 채점관의 성과를 검토할 때 이러한 타당한 의견 차이를 반영하기 위해 ‘허용 오차’라는 개념을 활용합니다. 예를 들어 책임 채점관이 한 에세이에 46점을 부여했고, 해당 문제의 허용 오차가 2점이라면 IB는 44점에서 48점 사이의 점수를 부여한 채점관이 적절한 기준에 따라 채점했다고 판단합니다.

IB는 채점관이 이러한 답안지를 채점할 때 확정적 점수(책임 채점관의 점수)에 어느 정도까지 근접해야 하는지를 규정합니다. 이때 허용되는 차이를 ‘허용 오차’라고 합니다. 여러 부분 또는 기준으로 구성된 문제를 검토할 때 IB는 전체 점수 차이뿐 아니라 각 부분 또는 기준별 점수 차이도 함께 모니터링합니다. 채점관은 이 두 가지 측면 모두에서 허용 오차 범위를 준수해야만 올바른 기준에 따라 채점하고 있음을 인정받을 수 있습니다.

[그림 27] 채점 기준의 일치



채점관 교육

평가 기수가 시작되기 전, 채점관 대상으로 교육이 실시됩니다. 이 교육의 목적은 채점관이 평가의 유형과 전반적인 채점 접근 방법을 이해하도록 돕는 것입니다. 다만 해당 교육에는 향후 기수의 시험에 관한 구체적인 정보는 포함되지 않습니다.

채점이 어렵거나 이례적인 답안지

채점관이 특정 답안지를 어떻게 채점해야 할지 판단하기 어려운 경우, 팀 리더에게 문의하여 조언을 구해야 합니다. 만약 해당 답안지를 공정하게 채점하기 어렵다고 판단될 경우, 상급 채점관에게 전달하여 채점을 요청할 수 있습니다. 특히 문제가 있는 답안지는 책임 채점관에게 보고되며, 책임 채점관이 해당 답안의 채점 방식에 대한 최종 결정을 내립니다.

이와 유사하게, 이례적인 답안지나 특수 변형 시험지로 시험을 본 학생의 답안지는 책임 채점관이 직접 채점합니다. 다른 학생들에게 적용된 기준과 동등한 수준의 채점이 이루어지도록 신중하게 균형을 맞추는 것이 책임 채점관의 역할입니다.

학교와의 관계

IB는 채점관이 가르치는 학생의 과제를 채점하지 않는 것을 원칙으로 합니다.

이를 관리하기 위해 IB는 모든 채점관에게 학생 및 학교와의 관계를 공개하도록 요구하며, 이해 상충이 발생할 수 있는 학교의 학생 과제를 배정받지 않도록 합니다. 이 과정에 대한 자세한 내용은 본 간행물의 ‘[이해의 상충](#)’ 섹션을 참고하십시오.

채점관의 메모 및 주석

IB 총괄 평가의 목적은 학생의 성과를 측정하는 것입니다. 따라서 채점관은 요구된 기준에 따라 학생 과제를 채점하는 데에만 전념해야 합니다. IB는 채점관이 채점을 수행하는 데 실질적으로 도움이 되는 경우에만 의견을 제시하도록 요청합니다.

학생 또는 교사를 위한 형성적 피드백을 작성하려면, 채점관은 과제의 적절한 점수를 판단하고, 그 과제가 어떻게 개선될 수 있었는지를 설명해야 합니다. IB는 이러한 피드백 과정이 훌륭한 교수의 핵심이라는 점을 인정하면서도, 이는 결코 단순한 작업이 아님을 인지하고 있습니다. 이 과정은 시간과 깊은 사고를 요구하며, 이는 일관된 기준에 의거하여 채점해야 하는 채점관의 핵심 임무에 지장을 줄 수 있습니다. 간단히 말해 IB는 채점관이 두 가지 과제(채점 및 피드백)를 낮은 수준으로 수행하는 것보다, 한 가지 과제(채점)를 높은 수준으로 수행하기를 기대합니다.

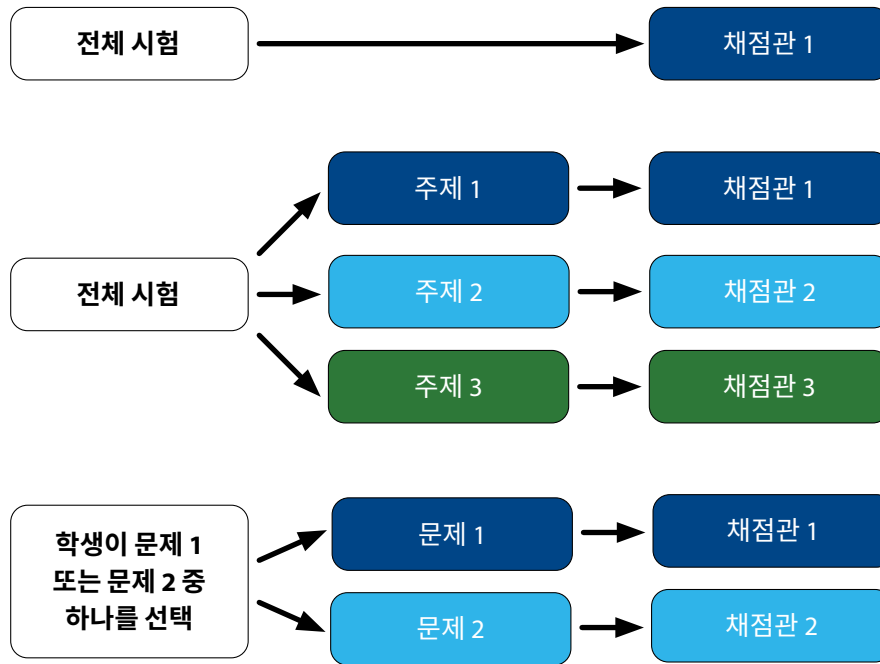
또한 채점관은 자신에게 주어진 단일 과제만을 근거로 판단할 수 있지만, 숙련된 교사는 학생에게 피드백을 제공할 때 훨씬 폭넓은 정보를 활용합니다. 따라서 채점관의 피드백은 교사의 피드백보다 통찰력 면에서 깊이가 부족할 수밖에 없습니다.

이러한 이유로 IB는 채점관에게 학생 또는 교사를 위한 피드백을 작성하지 말고, 오직 올바른 기준에 따라 채점할 것을 명시하고 있습니다.

다만 채점 과정을 지원하기 위해 채점관은 점수가 부여된 부분을 분명히 표시해야 하며, 모호함이 발생할 가능성이 있을 경우 적절한 의견을 남겨 이를 명확히 해야 합니다. 이렇게 하면 IB가 기준을 검증하는 데 도움이 되며 학교가 점수 부여의 근거를 명확히 확인할 수 있도록 투명성을 보장합니다.

문제 문항 그룹

[그림 28] 시험지를 문제 문항 그룹으로 구분하기



문제 문항 그룹의 개념은 채점관이 한 학생의 전체 답안지를 처음부터 끝까지 채점한 뒤 다음 학생의 답안지를 다시 채점하는 방식보다, 동일한 문제를 여러 번 반복적으로 채점하는 것이 훨씬 효율적이라는 발상에서 출발하였습니다.

각 문제 문항 그룹에는 별도의 연습 답안지, 채점 자격 취득용 답안지, 시드 답안지가 있으며 채점관은 각 문제 문항 그룹에서 요구되는 기준에 따라 채점할 수 있음을 입증해야 합니다. 이 방식은 도전적으로 보일 수 있지만, 채점관이 특정 문제의 채점 기준을 완벽히 이해하지 못하더라도, 전체 시험을 채점하지 못하게 되는 대신, 해당 문제를 제외한 나머지 문제를 계속 채점할 수 있음을 의미합니다.

종합

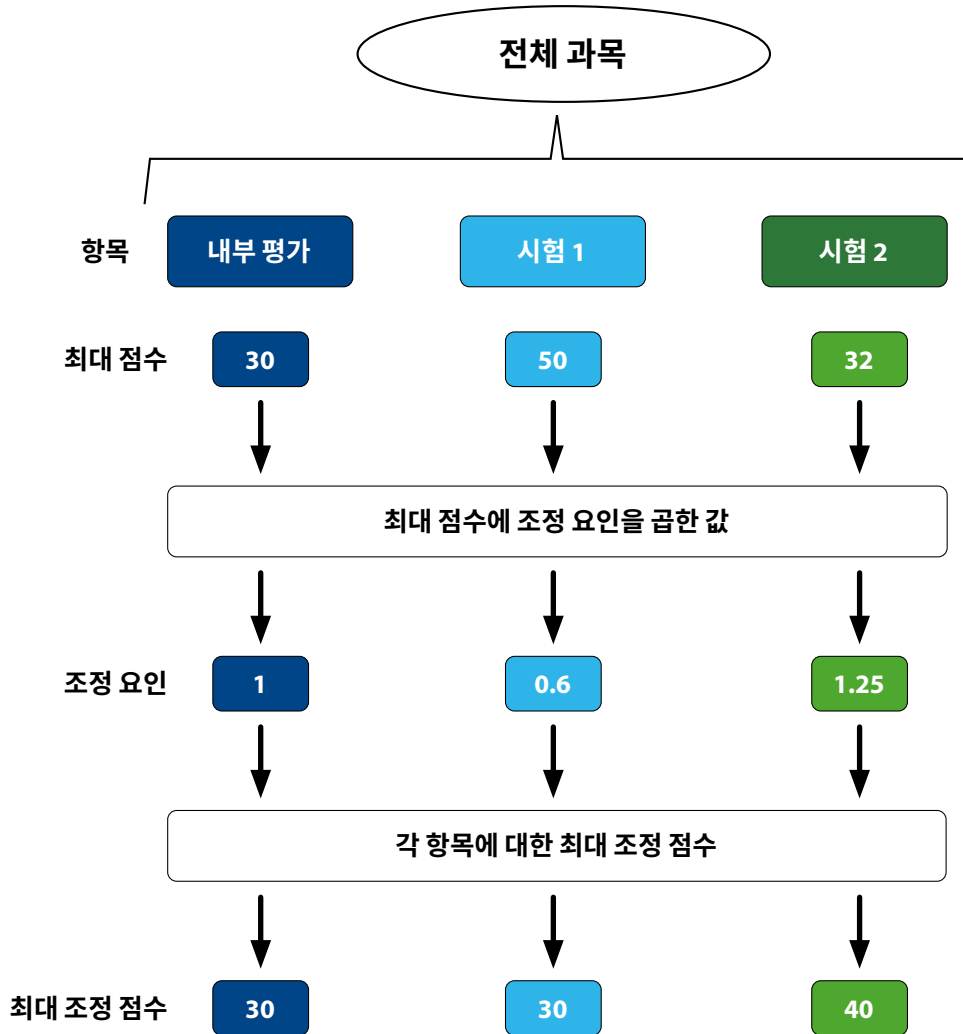
- 종합이란 여러 항목을 결합하여 전체 결과를 산출하는 과정입니다.
- 각 항목이 최종 점수에 올바른 비율(가중치)로 반영되도록 하기 위해 구성요소 점수를 환산해야 할 수도 있습니다.
- IB는 특정 문제나 항목에서 낮은 성과를 다른 문제 또는 항목에서의 높은 성과로 상쇄할 수 있는 ‘보상 모델’을 사용합니다.
- 학생의 최종 과목 성적은 항목별 성적이 아닌 항목 점수를 종합한 결과를 기준으로 결정됩니다.

종합은 서로 다른 항목의 점수(및 성적 급간)를 결합하여 최종 점수 또는 전체 성적 급간을 산출하는 과정을 의미합니다. 이를 실현하기 위해, 각 항목의 총점(또는 급간)을 환산해야 할 수도 있습니다.

환산은 각 항목이 교과 과정의 전체 평가에 기여하는 비율을 반영하여, 설정된 가중치를 유지하기 위해 수행됩니다. 이는 항목 점수를 곱하거나 나누어 과목 총점에 올바른 비율로 합산되도록 하는 것을 의미합니다. 항목별 최고점을 기준으로 처음에 설정되는 성적 급간의 경우에도 동일하게 적용됩니다.

각 항목의 최고 환산 점수는 교과 가이드에 명시된 해당 항목의 가중치와 같습니다.

[그림 29] 세 가지 항목이 있는 과목의 점수 환산 예시



가중치의 개념은 최종 결과에 기여하는 측면에서 각 평가 요소에 대해 IB가 부여하는 상대적 중요성을 반영하는 것입니다. 예를 들어, 한 항목이 주로 데이터 또는 자료의 해석을 평가하고 가중치가 30%라면 이는 해당 과정의 다른 목표와 비교했을 때 해석 능력이 전체 중요도의 약 30%를 차지함을 의미합니다. 다만, 여러 항목이 유사한 목표를 평가하는 경우가 많기 때문에 이러한 계산이 항상 실질적인 의미를 갖는 것은 아닙니다.

가중치의 또 다른 중요한 측면은 평가의 총점을 인위적으로 고정된 점수에 맞추려 하기보다 과제의 특성과 채점 기준에 적합한 총점을 설정할 수 있게 한다는 점입니다. 이러한 총점은 종합되어 최종 성적을 산출하는 데 사용됩니다.

평가에서 종합은 서로 다른 항목의 점수를 결합하여 학생의 최종 성적을 결정하는 과정을 의미합니다. 이러한 접근 방법은 항목 가중치에 따른 고정된 총점에 얽매이지 않고, 과제의 난이도와 기준, 평가 설계 의도에 따라 점수를 유연하게 설정할 수 있습니다. 즉, 항목 가중치는 고려 요소이지만 평가 또는 문제가 예기치 않게 쉽거나 어려운 경우 실제 결과가 의도와 다를 수 있음을 의미합니다. IB의 접근 방법은 개별 점수를 사전에 정해진 가중치에 정확히 일치시키는 것이 아닙니다. 가장 중요한 것은 전체 결과입니다.

‘보상 모델’로 알려진 이 방법은 개별 항목의 점수가 아닌 총점에 중점을 두며, 이를 통해 학생이 한 영역에서 낮은 성과를 다른 영역의 높은 성과로 상쇄할 수 있게 합니다. 따라서 최종 성적은 개별 항목 성적의 평균이 아니라 모든 항목 점수의 합으로 산출되며, 이에 따라 항목별 성적이 동일하더라도 최종 성적은 서로 다를 수 있습니다.

조정

- 조정은 학생의 과제를 재채점하는 과정이 아니라, 교사의 채점 기준을 검증하는 과정입니다.
- 성공적인 조정은 학생이 지구 반대편의 다른 학교로 전학을 가더라도 동일한 내부 평가 점수를 받을 수 있음을 의미합니다. IB는 이를 ‘세계적 기준’이라 부릅니다.
- 조정에서 사용되는 근거는 학생 결과물의 질뿐만 아니라, 교사가 해당 점수를 부여한 이유에 대한 설명도 포함됩니다.

역동적 표본 추출을 활용하는 조정은 DP 및 CP의 내부 평가 그리고 MYP의 e-포트폴리오에 적용됩니다.

조정 정의하기

많은 경우, IB가 평가하고자 하는 학생의 자질은 공식적인 시간 제한이 있는 시험에서 평가할 수 없습니다. 이러한 경우 가장 타당한 접근 방법은 교사가 내부 평가를 통해 평가를 수행하는 것입니다. ‘교실 기반 평가 및 내부 평가의 역할’ 섹션에서 이와 같은 접근 방법이 적절한 이유를 확인할 수 있습니다.

이 방법은 의미 있는 결과를 도출하지만, 동시에 교사마다 채점 기준을 다르게 해석할 위험을 내포하고 있습니다. 예를 들어, 서로 다른 두 학교의 교사가 동일한 과제에 서로 다른 점수를 부여할 수도 있습니다. IB는 채점관들이 요구되는 기준에 대한 공통된 이해를 갖도록 하기 위해 포괄적인 교육 및 평가 절차를 갖추고 있으나, 이는 모든 교사에게 적용하기에는 현실적으로 어렵습니다.

대신, IB는 교사들이 모든 학생을 일관된 기준으로 채점한다고 확신하며, 필요한 경우에만 세계적 기준에 맞추기 위해 점수를 조정하면 됩니다. 이를 위해 IB는 교사가 채점한 샘플을 요청하고, 이를 책임 채점관의 기준과 비교하여 검토합니다. 이 비교에서 얻은 데이터를 바탕으로 IB는 필요 시 각 교사가 부여한 점수를 조정하는 수학 공식을 산출합니다. 두 점수 집합의 통계적 비교 결과에 따라, 필요한 경우 해당 교사가 부여한 점수를(해당 항목에 한해) 학교의 모든 학생에게 일괄적으로 조정합니다. 교사가 지속적으로 점수를 낮게 또는 높게 부여한다면 모든 점수에 동일한 조정이 적용됩니다. 그러나, 교사가 점수 범위의 상위권 또는 하위권에서 낮게 또는 높게 점수를 부여하는 경우, 조정은 교사의 점수 범위 전반에 걸쳐 달라질 수 있습니다.

유념해야 할 중요한 사항

- 조정의 목적은 교사가 학생의 과제를 채점할 때 평가 기준을 얼마나 정확하고 일관되게 적용했는지를 확인하는 것입니다.
- 조정의 결과에 따라 학교의 점수가 하향되거나 상향될 수 있습니다. 또는 그대로 유지될 수도 있습니다.
- 조정 인자가 적용된다고 해서 교사의 채점의 질이 낮다는 뜻은 아니며, 단지 그 채점이 세계적 기준과 일치하지 않음을 뜻합니다.

실질적인 이유로 IB는 개별 교사 단위가 아니라 학교 단위로 조정을 수행하며, 따라서 학교의 모든 교사가 동일한 기준으로 채점하는 것이 매우 중요합니다. 조정 인자는 모든 교사의 점수에 적용되므로, 조정이 적절히 이루어질 수 있도록 샘플이 학교의 채점 방식의 공정한 대표 표본이어야 합니다.

코디네이터가 조정 대상인 내부 평가 점수를 검토할 때 고려해야 할 사항은 다음과 같습니다.

- 학교의 모든 과목 교사(내부 평가 점수를 부여하는 담당자)가 동일한 기준으로 채점하고 있습니까?
- 교사들이 모든 학생을 동일한 기준에 따라 일관성 있게 채점하고 있습니까?
- 교사들이 특정 점수를 부여한 이유에 대해 명확한 설명을 제공합니까?

평가 기준을 정확하게 적용하는 방법에 관한 교사용 추가 지침은 본 간행물의 ‘학교에 제공되는 평가 피드백’ 섹션 및 현재 사용되는 MYP, DP, CP 「평가 절차」 간행물에서 확인할 수 있습니다.

조정 인자의 계산과 관련된 기술적 측면에 대한 자세한 정보는 본 간행물 부록의 ‘내부 평가의 조정: 자세히 알아보기’에서 확인할 수 있습니다.

조정: 교사에게 기대되는 사항

평가를 위해 제출된 과제는 교사가 내부적으로 평가하고 IB가 외부적으로 조정합니다.

교사는 내부 평가 기준이 상세히 제시된 관련 교과 가이드를 숙지해야 합니다. 각 평가 기준에는 특정 범위의 점수와 함께 그에 상응하는 성취 단계를 설명하는 단계 기술어가 제시되어 있습니다. 단계 기술어는 성취의 긍정적 측면에 초점을 두고 있으며, 낮은 단계의 경우 성취 미달에 대한 기술이 포함될 수 있습니다.

교사는 단계 기술어와 관련 설명을 참고하여 내부 평가 과제를 기준에 따라 판단해야 합니다. 기준은 최적화 모델에 따라 적용되어야 합니다. 과제가 서로 다른 수준에서 기준의 여러 요소를 충족하는 경우, 부여되는 점수는 전체적으로 해당 기준에 대한 성취도를 가장 공정하게 반영해야 합니다. 단계 기술어의 모든 측면을 충족해야만 해당 점수를 받는 것은 아닙니다. 가장 높은 단계 기술어가 학생의 과제가 완벽하다는 것을 의미하지 않습니다.

단계 내에 두 개 이상의 점수가 존재할 경우, 해당 단계의 기술된 특성을 대부분 충족하고 상위 단계의 성취 수준에 근접해 있다면 두 점수 중 높은 점수를 주어야 합니다. 반대로, 학생의 과제가 단계의 요건을 충족하지 못하고, 하위 단계에 더 가까울 경우 교사는 낮은 점수를 부여해야 합니다.

점수는 반드시 정수로만 기록해야 하며, 분수나 소수점 등의 부분 점수는 허용되지 않습니다.

기준은 독립적으로 고려되어야 합니다. 한 기준에서 높은 성취 단계를 달성한 학생이 다른 기준에서도 반드시 높은 성취 단계를 달성하는 것은 아닙니다. 마찬가지로 한 기준에서 낮은 성취 단계를 보이는 학생이 다른 기준에서도 반드시 낮은 성취 단계를 보이는 것은 아닙니다. 교사는 학생의 전반적인 평가 점수가 특정 점수 분포를 보일 것이라고 가정해서는 안 됩니다.

일부 과목에서는 단계 기술어에 특정 지시어가 포함될 수 있습니다. 이러한 지시어는 가이드의 관련 섹션에 명시된 대로 해석해야 합니다.

내부 평가 과제에 조정 인자가 적용되더라도, 이는 교사가 채점 절차를 잘못 따랐음을 의미하지 않습니다. 조정은 단지 부여된 점수를 세계적 기준에 맞추기 위한 것입니다. 매 시험 기수 이후 발표되는 과목 보고서에는 내부 평가 수행에 관한 유용한 정보가 포함되어 있는 경우가 많습니다. 또한 교사용 보조 자료는 프로그램 자료실에서 확인할 수 있습니다.

각 학교에서 제출한 모든 평가 과제에는 동일한 조정 인자가 적용됩니다. 따라서 학교에서 동일한 과목의 내부 평가를 여러 교사가 분담하는 경우, 채점 기준을 최대한 일치시키기 위한 내부 표준화가 반드시 이루어져야 합니다. 이는 채점 시작 전 평가 기준에 대해 논의하고 제출된 과제를 검토하는 과정을 포함할 수 있습니다.

교사와 학생은 학문적 진실성과 관련된 개념, 특히 진위성 및 지식 재산권의 개념을 반드시 이해해야 합니다. 학생의 과제는 전적으로 본인이 직접 수행한 것이어야 하며, 각 교과 가이드에 명시된 요건 목록에 따라 준비되어야 합니다. 협력이 허용되는 경우, 협력과 결탁의 차이를 모든 학생에게 명확히 안내해야 합니다. 자세한 내용은 ‘[윤리적 사고방식 기르기](#)’ 섹션을 참고하십시오.

이 내용의 요약본을 확인하고자 하는 교사는 본 간행물 부록의 ‘인쇄 가능한 자료’에서 ‘내부 평가 채점하기: 조정 과정에서 교사에게 기대되는 사항’ 포스터를 인쇄하여 활용할 수 있습니다.

학생 과제의 선택

- 모든 학생에게 공정한 점수가 부여되어야 합니다. 따라서 어느 학생의 과제라도 조정을 위한 교사의 기준을 보여주는 적절한 예시로 사용될 수 있습니다.
- 교사마다 매우 우수한 과제나 낮은 수준의 과제가 무엇인지에 대한 기준이 다를 수 있으므로, 학교의 샘플에는 부여된 점수의 전체 범위가 반드시 포함되어야 합니다.
- IB는 모든 학교 간의 투명성과 공정성을 보장하기 위해 학생의 최종 점수에 기여한 모든 과제의 근거를 확인할 수 있어야 합니다.

투명성을 보장하고 학업 부정행위에 대한 의혹을 방지하기 위해 조정을 위한 샘플은 학교가 아닌 IB가 선정하는 것이 중요합니다. IB는 조정 인자의 신뢰도를 극대화하기 위해 광범위한 지침을 활용하며 이러한 지침의 범위 내에서 실제 학생 선정이 무작위로 이루어지도록 합니다.

첫 번째 원칙은 IB에서 어떤 학교에 대해서도 신뢰할 수 있는 조정 인자를 얻기 위해 가능한 최소한의 샘플을 사용해야 한다는 것입니다. 이렇게 하면 학교 측의 부담을 최소화할 수 있을 뿐 아니라, 시험 응시료를 통해 학교에 전가될 수 있는 IB의 비용 또한 줄일 수 있습니다. 그러나 IB가 보다 견고한 조정 인자를 산출하기 위해 더 많은 교사의 채점 예시를 필요로 하는 경우 학교별로 샘플에 포함되는 학생 과제 수가 다를 수 있습니다.

두 번째 원칙은 조정 인자가 전체 점수 범위에 걸쳐 모든 학생에게 공정하다는 확신을 IB가 가질 수 있어야 한다는 것입니다. IB의 경험에 따르면 교사는 점수 범위의 서로 다른 구간에서 서로 다른 기대 수준을 보일 수 있으며, 낮은 수준의 과제에는 세계적 기준보다 관대하면서 높은 수준의 과제에는 오히려 세계적 기준보다 더 엄격할 수도 있습니다. 이러한 이유로 학교의 전체 점수 범위가 적절히 반영될 수 있도록 내부 평가 샘플이 신중하게 선정됩니다.

또한 IB는 일반적으로 만점을 받은 학생의 과제를 조정 샘플로 선정하지 않는 경향이 있습니다. 이는 교사가 지나치게 엄격하게 채점했을 경우, 높은 점수대에 있는 학생들의 점수가 상향 조정될 수 있는 여지를 남겨두기 위함입니다. 마찬가지로 0점을 받은 학생의 과제 역시 일반적으로 조정 샘플로 선정되지 않습니다.

특수한 상황

조정은 특정 과제가 채점하기 매우 어려운 경우에도 교사가 세계적 기준에 맞게 채점하고 있는지를 확인하기 위해 설계된 절차입니다. 이와 같은 상황에서는 교사가 해당 점수를 부여한 이유를 보다 상세히 설명해야 할 수 있습니다. 예를 들어 학생이 과제를 수행하는 과정에서 교사로부터 추가적인 도움을 받은 경우, 교사는 이를 채점 의견에 명시해야 하며 채점관은 교사의 채점을 검토할 때 이를 고려해야 합니다.

조정 인자를 산출할 수 없는 경우

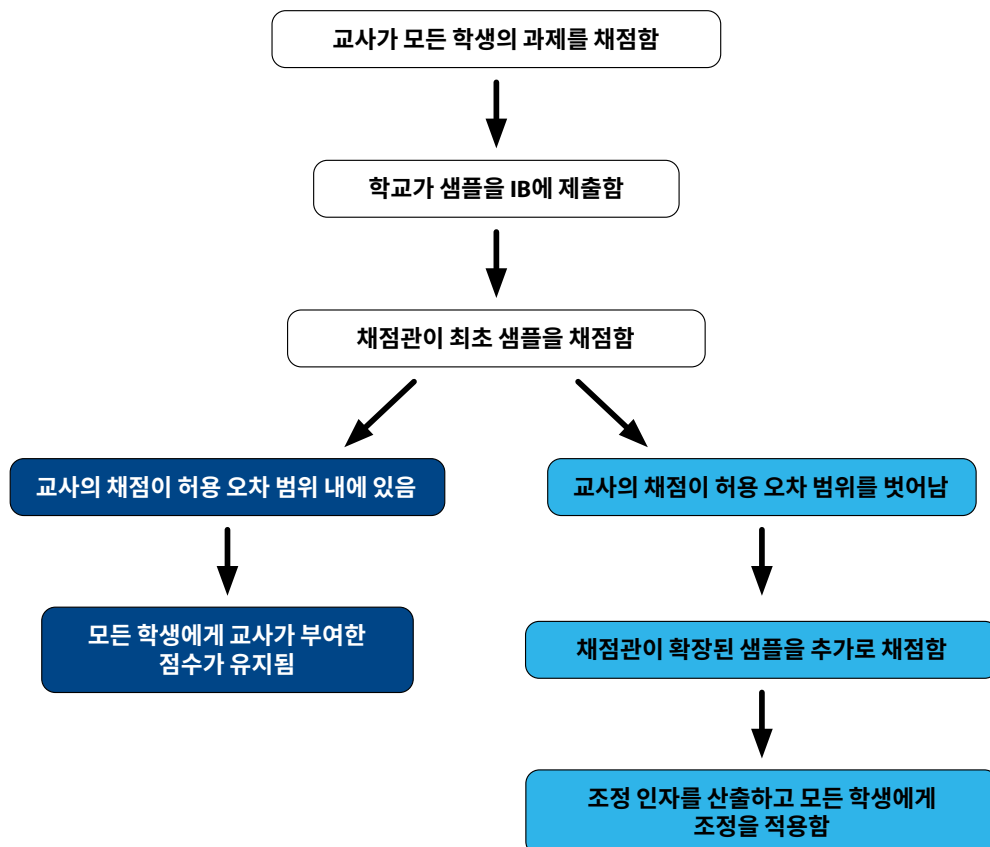
제출된 샘플 과제만으로 조정 인자를 산출할 수 없는 경우가 발생할 수 있습니다. 이는 채점관의 점수와 교사의 점수 간 차이가 일관되지 않거나, 점수를 기준으로 볼 때 교사의 채점이 세계적 기준에 비해 지나치게 관대하거나 지나치게 엄격하다는 경향이 나타나는 경우에 발생합니다. 이러한 경우 IB는 공정한 조정 인자를 적용할 수 있도록 학교에 추가 과제 제출을 요청합니다. 이러한 이유로 모든 학생의 과제는 시험 기수 종료 시점까지 보관되어 있어야 합니다. 아주 드문 경우이지만, 공정한 조정 인자를 적용할 수 없는 경우에는 채점관의 점수가 공식 점수로 부여됩니다.

역동적 표본 추출

- 역동적 표본 추출을 활용하는 조정은 교사의 채점이 허용 오차 범위 내에 있을 경우, IB가 해당 점수를 그대로 인정함을 의미합니다.
- 교사의 점수가 허용 오차 범위를 벗어난 경우, IB는 조정 인자를 적용합니다.
- 교사의 내부 평가 채점을 검토하는 모든 채점관은 품질 보증을 위해 모니터링됩니다.

역동적 표본 추출을 활용하는 조정의 원칙은 교사가 샘플에서 허용 오차 범위 내의 점수를 부여함으로써 세계적 기준을 잘 이해하고 있음을 입증할 경우, 해당 교사가 부여한 모든 점수가 요구되는 기준을 충족한 것으로 간주한다는 것입니다. 반대로 허용 오차를 벗어난 경우, 조정 인자가 산출되어 적용됩니다. 이는 [그림 30]에서 설명되어 있습니다.

[그림 30] 역동적 표본 추출을 활용하는 조정



조정이 공정하게 이루어지기 위해서는 책임 채점관이 설정한 세계적 기준을 충분히 이해하고 이를 바탕으로 조정을 수행해야 합니다. 조정을 담당하는 채점관은 세계적 기준에 대한 교육을 받고, 해당 기준에 따라 조정을 수행하고 있는지를 점검받습니다. ‘표준화’ 섹션에서 설명된 바와 같이, 책임 채점관은 세 가지 유형의 확정적으로 채점된 내부 평가 과제를 준비합니다.

- **연습 답안지:** 세계적 기준을 설명하기 위함
- **채점 자격 취득용 답안지:** 채점관이 세계적 기준을 이해했는지 확인하기 위함
- **시드 답안지:** 채점관이 세계적 기준을 유지하고 있는지 확인하기 위함

한 학교의 내부 평가 과제는 동일한 채점관이 모두 검토합니다. 그러나 채점관에게 제시되는 과제는 여러 학교의 과제가 특정 순서 없이 무작위로 제공됩니다. 이는 채점관이 첫 번째 학생의 과제물만 보고 교사가 높게 또는 낮게 채점했다는 선입견을 가진 채, 나머지 샘플에서도 같은 패턴을 찾으려 하는 것을 방지합니다. 또한 이 방식은 시드 답안지를 포함하더라도 그것이 시드 답안지임을 드러나지 않도록 합니다. 조정 인자의 필요 여부가 결정되면 채점관에게 학교의 채점에 대한 요약된 의견을 제시할 것을 요청할 수 있습니다.

역동적 표본 추출의 품질 모델은 다른 유형의 채점 모델보다 더욱 복잡합니다. IB는 채점관이 교사의 점수가 엄격한지, 관대한지, 혹은 적절한지를 검토할 때 확신을 가지고 판단하는지를 점검해야 합니다. [그림 31]은 모든 시나리오를 포괄하기 위해 필요한 품질 모델 내부 평가 답안지의 범위를 보여줍니다.

[그림 31] 요구되는 내부 평가 모델 답안지의 품질 범위

	시나리오	
채점관의 점수가 점수 범위의 상위권에 해당하는 경우	1	교사가 매우 엄격하게 채점함
	2	교사가 세계적 기준에 맞게 채점함
	3	교사가 매우 관대하게 채점함
채점관의 점수가 점수 범위의 중위권에 해당하는 경우	4	교사가 매우 엄격하게 채점함
	5	교사가 세계적 기준에 맞게 채점함
	6	교사가 매우 관대하게 채점함
채점관의 점수가 점수 범위의 하위권에 해당하는 경우	7	교사가 매우 엄격하게 채점함
	8	교사가 세계적 기준에 맞게 채점함
	9	교사가 매우 관대하게 채점함

IB 조정 방식에 관한 자세한 정보는 “[The two methods of marking: How marking is carried out by teachers and examiners](#)”(영문 제공, 동영상)를 참고하십시오.

성적 부여 및 종합

- 학생이 어느 기수에서 시험을 치르든 성적의 의미는 동일해야 합니다.
- 이를 위해 성적 부여 과정에서는 점수를 성적으로 환산하는 방식을 결정합니다.
- 성적 급간은 전체 학생 집단의 결과와 제출된 평가 과제에 대한 전문가의 판단을 포함한 다양한 근거를 토대로 결정됩니다.
- 학생의 성취에 있어 가장 중요한 것은 개별 항목의 점수가 아니라 전체 성적입니다.

점수와 성적은 서로 다른 개념이라는 점을 유념해야 합니다. 이에 대한 자세한 설명은 ‘[점수와 성적의 차이점](#)’ 섹션을 참고하십시오. 성적 부여 과정은 특정 시험에서 점수가 어떤 성적에 해당하는지를 결정하는 절차를 의미합니다.

이 결정은 책임 채점관, 수석 채점관, 선임 채점관 및 IB가 며칠간의 심도 있는 논의를 거쳐 이루어집니다. 이 과정에서는 해당 연도뿐 아니라 이전 연도에 응시한 학생들에게도 공정한 결론을 도출하기 위해 다양한 근거(자세한 내용은 ‘[성적 부여 시 적용되는 근거](#)’ 섹션 참조)를 활용합니다. 최종적으로 수석 채점관은 이번 시험 기수의 결과에 대한 권고안을 IB에 제시합니다.

이 논의의 주요 결과물은 각 성적을 받기 위해 학생이 획득해야 하는 최소 점수를 나타내는 성적 급간입니다. 모든 과목의 모든 등급에 대해 세부적인 판단을 내리는 것은 현실적으로 불가능하므로, IB는 채점관에게 몇 가지 ‘[판단 기반 성적 급간](#)’에 대한 제안을 요청하고 나머지 급간은 산술적으로 계산합니다. 자세한 내용은 ‘[판단 기반 성적 급간과 보간된 성적 급간](#)’ 섹션을 참고하십시오.

성적 급간은 시험 기수마다 동일하지 않을 수 있습니다. 이는 시험 기수마다 학생에게 제시되는 문제가 달라지며, 시험의 난이도에 변화를 반영하기 위해 성적 급간이 조정되어야 하기 때문입니다. IB는 매년 시험 난이도를 일관된 수준으로 유지하기 위해 최선의 노력을 다합니다. 그러나 시험의 중요성과 보안상의 이유로, 문제가 유출될 위험이 있는 사전 모의 시험은 시행하지 않습니다.

성적 급간의 위치를 결정할 때 채점관과 IB는 [그림 32]에 제시된 요소들을 고려합니다.

[그림 32] 성적 부여 과정에서 고려되는 측면

학생 집단	<ul style="list-style-type: none"> 이런 시험에 응시한 학생 집단은 전년도와 비교하여 어떠합니까?
시험	<ul style="list-style-type: none"> 출제자의 예상과 비교하여 어떠한 성과를 보였습니까? 시험의 난이도에 대한 교사의 관점은 어땠습니까?
답안지	<ul style="list-style-type: none"> 전년도 및 성적 기준표와 비교하여 응시생의 과제는 어떠합니까?
결과	<ul style="list-style-type: none"> 전년도와 비교하여 올해 응시생의 결과는 어땠습니까?
균형	<ul style="list-style-type: none"> 존재하는 모든 근거가 하나의 결론으로 이어집니까? 아니면 서로 상충합니까?

IB는 대부분의 상황에서 여러 평가 항목을 종합하여 전체 성적을 산출합니다. 이에 대한 자세한 내용은 본 간행물의 ‘종합’ 섹션에서 확인할 수 있습니다. 성적 부여 과정에서 채점관이 반드시 유념해야 할 가장 중요한 점은 **전체 결과**가 핵심이라는 사실입니다. 필요한 경우 전체적으로 공정한 결과를 얻기 위해 개별 항목의 성적 급간이 완벽하지 않을 수도 있습니다.

성적 부여 과정의 공식적인 목적은 다음과 같습니다.

- 평가를 위해 제출된 과제의 분포에서 과제의 질을 가장 잘 설명하는 성적 기술어가 달라지는 지점을 설정하는 것
- 각 항목에서 성적 급간이 될 점수를 결정하는 것
- 과목 전체 수준에서 이 성적 급간들의 조합이 공정한 성적 부여를 반영하도록 하는 것

성적 부여 과정이 성공적이었을 경우 다음과 같은 결과를 얻을 수 있습니다.

- 해당 학년의 정보를 고려했을 때, 답안지에 대한 판단과 결과에 대한 근거가 넓은 의미에서 일치하는 경우
- 학교별 성과에서 나타나는 유의미한 편차가 설명할 수 있는 경우
- 수석 채점관과 IB 수석 평가 담당자(Chief Assessment Officer)가 평가 기준이 유지되었음을 확인하는 경우

판단 기반 성적 급간과 보간된 성적 급간

판단 기반 성적 급간은 성적 부여 과정에서의 논의를 바탕으로 수석 채점관이 제안하는 성적 급간을 의미합니다. MYP, DP, CP에서 이러한 판단 기반 성적 급간은 2/3, 3/4, 6/7입니다. 나머지 성적 급간인 1/2, 4/5 및 5/6은 판단 기반 성적 급간을 바탕으로 산술적으로 계산되며 **보간된 성적 급간**이라고 합니다.

5점 또는 1점을 받은 학생의 비율이 크게 바뀐 경우 전체 학년 집단의 맥락에서 이에 대해 논의해야 합니다. 이로 인해 판단 기반 성적 급간이 다시 고려되거나 예외적인 경우 해당 급간에 포함된 평가 제출 과제가 다시 검토될 수 있습니다.

성적 부여 과정에서 활용되는 근거

성적 부여는 근거에 기반하여 이루어지며 다음을 포함한 다양한 정보를 참고합니다.

- 평가에 대한 교사의 피드백
- 학생 과제 예시에 대한 채점관의 전문적인 판단
- 올해의 학생 성과를 이전 연도의 결과와 비교하는 통계 정보의 검토

특히 더 중요한 유형의 근거는 없습니다. 모든 근거를 균등하게 고려하여 결론을 내려야 합니다.

이번 연도의 학생 집단 고려하기

성적 부여 과정에서 가장 먼저 수행해야 하는 과제는 이번 연도의 평가 응시 학년 집단이 이전 연도의 집단과 얼마나 유사한지를 검토하는 것입니다. 올해 평가를 수행하는 학생이 속한 학교가 전년도와 대체로 동일하다면 학생 성과의 차이는 시험 난이도의 차이에서 비롯된 것일 가능성이 높습니다. 하지만 상황이 이와 다를 경우(예: 다수의 신규 학교가 처음으로 해당 과목을 시행하거나, 재응시 학생 비율이 매우 높은 경우), IB는 이러한 성과 차이가 성적 결과에 반영되어야 한다고 판단할 수 있습니다.

평가를 수행하는 전체 학년 집단을 비교할 때 고려되는 요인의 예시는 다음과 같습니다.

- 평가를 수행하는 학생 수의 변화
- 영어, 프랑스어, 스페인어 및 적절한 경우 다른 언어로 평가를 수행하는 학생 비율의 변화
- 새로운 학교의 수와 해당 학교의 학생 수
- 학생이 선택한 옵션에서의 모든 변화

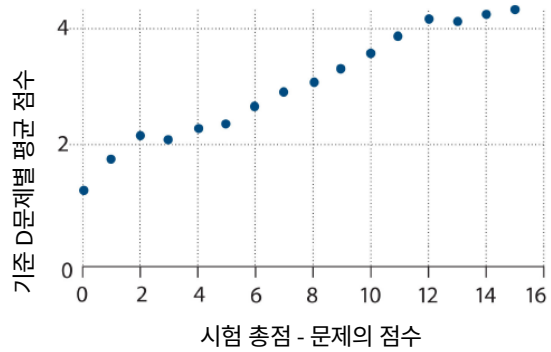
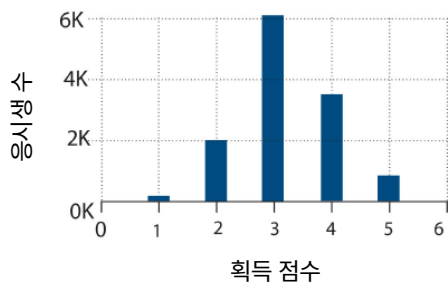
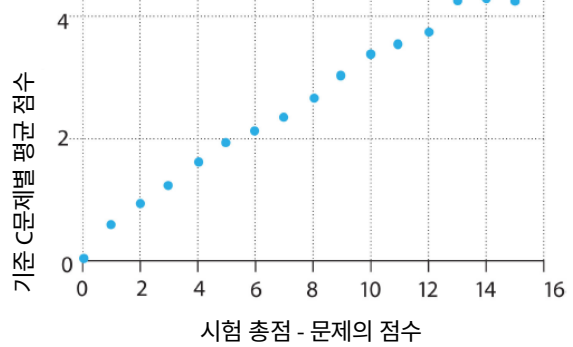
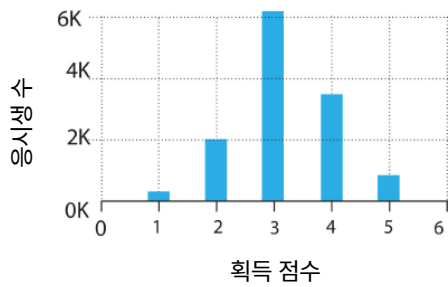
평가에 대한 피드백

다음 과제는 기대에 비해 평가가 사전에 예측된 기대 수준과 비교하여 어떻게 수행되었는지를 검토합니다. 특정 문제가 예상보다 훨씬 어려웠거나, 6점과 7점 학생을 변별할 것으로 기대된 문제가 이를 달성하지 못한 경우, 성적 부여 팀은 성적 급간을 설정할 때 이러한 요소를 고려해야 합니다.

이들은 시험에 대한 교사의 피드백과 개별 문제 또는 문항의 성과에 관한 통계([그림 33] 참조)를 바탕으로 논의를 진행합니다. 또한 성적 부여 팀은 개별 채점관의 보고서를 요약하고 해당 기수에서 학생 과제를 채점한 자신의 경험도 함께 반영합니다.

[그림 33] 채점관에게 제공되는 문항 수준의 통계 예시

문항 요약 통계							해당 문항과 나머지 구성요소의 상관관계	
문제	문제	총점	해당 문제에 답한 학생 비율	문제 문항별 평균 점수	문항 최고점	모표준	문제	상관 계수
01 기준 A	12,071	42,042	28.71%	2.7	5	0.95	01 기준 A	0.8044
01 기준 B	12,071	42,042	28.71%	2.6	5	0.94	01 기준 B	0.8018
01 기준 C	12,071	42,042	28.71%	2.8	5	0.84	01 기준 C	0.7983
01 기준 D	12,071	42,042	28.71%	3.2	5	0.79	01 기준 D	0.7253
02 기준 A	29,848	42,042	71.00%	2.8	5	0.98	02 기준 A	0.7954
02 기준 B	29,848	42,042	71.00%	2.7	5	0.93	02 기준 B	0.7944
02 기준 C	29,848	42,042	71.00%	2.9	5	0.83	02 기준 C	0.7865
02 기준 D	29,848	42,042	71.00%	3.2	5	0.79	02 기준 D	0.7320



답안지 근거 검토하기

다음으로 선임 채점관은 각 성적의 성적 기준표를 예상 성적 급간과 가까운 점수를 받은 답안과 비교합니다. 성적 부여 과정 단계에서 채점관은 부여된 점수가 아니라 학생 답안의 특성과 그 답안이 성적 기준표와 얼마나 부합하는지에 초점을 맞추어야 합니다.

답안 검토를 시작하기 전에, 채점관이 전년도 학생 과제 예시를 검토하여 각 성적 수준에 대한 기대치를 다시 상기하는 것이 유용합니다. 이러한 예시 답안은 성적 부여 과정 중 반드시 접근할 수 있어야 합니다.

성적 협의 위원회 이전에 선임 채점 팀 구성원, 특히 책임 채점관들은 각 항목의 예상 성적 급간을 제출합니다. 채점관들은 과거 기준에 대한 경험을 바탕으로, 성적 급간이 어느 점수에 위치해야 하는지 제시합니다. 이러한 잠정 성적 급간은 각 시험의 기능에 대한 합의와 전체 점수 분포에 대한 인식과 함께, IB 과목 담당자가 성적 부여 샘플에 포함해야 할 점수 범위를 제시하는 데 필요한 정보와 근거를 제공합니다.

각 선임 채점관은 선정된 학생 답안지를 검토하고 과제의 질을 가장 정확히 반영하는 성적 기준표가 무엇인지 결정해야 합니다. 비교적 일관된 수준의 답안을 작성한 학생이라 하더라도 답안 전반에서 다양한 성적 수준의 특성이 혼재되는 경우가 많기 때문에 이는 매우 어려운 과정입니다. 특정 답안이 성적 기준표 내에 겨우 해당되었는지 또는 근소한 차이로 상위 성적에 거의 도달했는지를 표시하는 것은 허용되며(예를 들어 '7-' 또는 '6+'로 표시) 이러한 표시는 종종 유용합니다.

채점관은 이를 수행할 때 어떠한 선입견이나 편향도 최소화해야 합니다. 따라서 모든 채점관이 자신의 판단을 독립적으로 기록하기 전까지는 서로의 의견을 공유하거나 논의하지 않는 것이 중요합니다. 마찬가지로, 수석 채점관을 제외한 채점관들이 이 단계를 완료하기 전까지는, 전체 통계 분석 결과 대신 학생 집단의 문항별 성과와 관련된 필수 정보만 제공받는 것이 바람직합니다.

답안에 가장 적합한 성적을 결정하는 일은 매우 주관적이므로 채점관 간의 차이가 발생하는 것은 자연스러운 일입니다. 또한 우수한 이해를 보여 주었지만 여러 실수를 한 학생이, 가장 쉬운 과제에서 높은 점수를 얻었으나 동일한 깊이의 이해를 보여 주지 못한 학생보다 다소 높은 성적을 받는 것은 충분히 합리적입니다. 이는 후자의 학생이 점수가 약간 더 높더라도 마찬가지입니다. 이 답안 검토 과정은 전체 학생 과제 중 일부 예시에만 근거하고 있다는 점을 유념해야 합니다.

연구에 따르면, 채점관은 자신이 검토 중인 답안들 사이에서 질적으로 차별화되는 답안을 식별하는데 가장 높은 전문성을 보입니다(본 간행물의 '채점 접근 방법' 섹션 참고). IB는 채점관에게 샘플 묶음 중 최고 점수의 답안부터 검토를 시작하여, 상위 성적에 해당하는 근거가 더 이상 일관되게 나타나지 않는 지점까지 점차 낮은 점수로 내려가며 검토할 것을 요청합니다. 그런 다음 가장 낮은 점수의 샘플부터 시작해 상위 성적의 근거가 일관되게 나타나기 시작하는 지점까지 상향 검토를 진행해야 합니다. 필요한 경우, 샘플의 점수 범위는 확대될 수 있습니다.

모든 선임 채점관의 성적 결정이 종합되면, 성적 급간이 위치해야 할 점수 범위를 지정해야 합니다. 이는 '불확실의 구간'이라 하며, 합리적인 근거에 따라 성적 급간을 설정할 수 있는 최저 점수부터 최고 점수까지의 범위를 의미합니다. 불확실의 구간을 설정하는 데 정해진 공식은 없으며, 선임 채점관이 각자의 판단을 바탕으로 논의하여 합의해야 합니다.

[그림 34] 답안 검토 결과의 예시

답안지 번호	점수	채점관 1(책임 채점관)	채점관 2	채점관 3	채점관 4
1	51	4	4	4	4
2	51	4	4	4	4
3	50	3+	4	4	3+
4	50	4-	4	4	3+
5	49	4-	4-	4-	4-
6	49	3+	3	4-	3+
7	48	4-	3+	4-	3+
8	48	3+	4-	4-	3+
9	47	3	3	3+	3
10	47	3	3	3	4-

이 과정이 유일하게 적용되지 않는 경우는 선다형 문제 시험입니다. 경험에 따르면 선다형 문제로만 구성된 시험에서 제출된 평가 과제의 질을 근거로 성적 급간을 판단하는 것은 매우 어려운 일입니다. 이는 답안에 학생이 실제로 무엇을 수행했는지 판단할 근거가 답안에 거의 드러나지 않기 때문일 수 있습니다. 따라서 이러한 시험의 경우, 각 성적에 해당하는 학생 비율이 해당 과목과 가장 밀접하게 연관된 다른 시험의 비율과 최대한 일치하도록 성적 급간을 산출합니다.

결과에 대한 통계 검토하기

전문가의 판단에 의존하는 준거 기반 시스템에서는 선임 채점관이 시험 문제와 각 문항이 학생에게 요구하는 답안의 수준을 고려하여 성적 급간을 설정할 수 있어야 한다고 주장할 수도 있습니다. 그러나 실제로는 학생이 어떻게 답변했는지 참고하지 않고 이러한 판단을 내리는 것은 매우 어렵습니다. 크레스웰(Cresswell, 2000)은 채점자가 한 기수의 시험이 다른 기수의 시험보다 더 쉽거나 어려운지를 구별하는 데는 대체로 정확하지만, 그 난이도의 차이가 어느 정도인지 추정하는 데는 대체로 정확하지 않다고 결론지었습니다.

성적 부여의 목적이 일관성 있는 성적의 의미 또는 기준을 유지하는 것이므로, 크레스웰이 제시한 비교 가능성의 정의 중 하나를 살펴보는 것은 의미가 있습니다.

동일한 능력과 사전 성취도를 가진 두 학생 집단이 동일한 입학 정책을 가진 학교에 재학하고, 동등한 역량을 갖춘 교사의 지도를 받으며, 동등한 수준의 동기를 지닌 상태에서, 각자 교수요목을 학습하고 시험을 치른 뒤, 동일한 성적 분포를 보인다면 두 시험은 비교 가능한 기준을 가진 것으로 정의될 수 있다.

(Cresswell, 1996, pp. 57-84)

성적 부여 과정을 지원하기 위해 선임 채점관에게는 **통계적 권장 급간**이 제공됩니다. 이는 특정 성적까지의 누적 백분율(특정 성적 또는 그 이상을 받은 학생의 비율)이 전년도와 동일하도록 설정된 성적 급간을 의미합니다.

성적 부여 과정에 참여하는 채점관에게는 학생의 평균 점수와 실제 점수 분포를 나타내는 히스토그램에 대한 정보도 제공됩니다. 이 정보는 통계적 권장 급간을 산출할 때 이미 반영되지만, 보다 상세한 정보를 직접 검토하는 것도 종종 유용합니다.

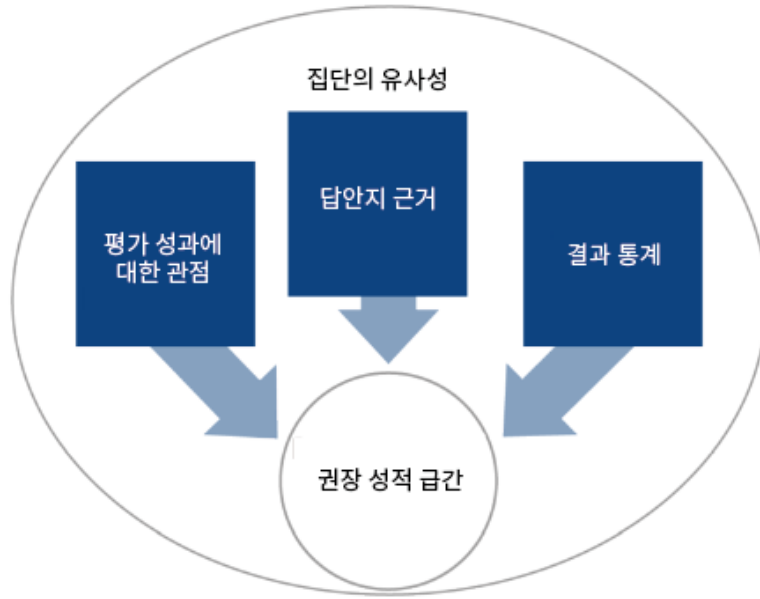
기수마다 시험을 치르는 학생 집단은 서로 다르다는 점을 반드시 기억해야 합니다. 이에 따라 앞서 제시된 비교 가능성의 정의에 내포된 가정이 완전히 충족되지 않을 가능성이 높습니다. 이는 특히 두 집단이 서로 매우 다를 때도 마찬가지입니다. IB 월드 스쿨과 같이 유사한 교육 경험을 지닌 대규모 학생 집단을 다루는 경우, 성취 결과에서 나타나는 유의미한 차이가 학생 개인의 성취도 차이보다 서로 다른 문제에 답했기 때문일 가능성이 더 높습니다.

통계적 권장 급간의 타당성을 평가할 때 학생 집단의 규모를 반드시 고려해야 하지만, 다른 요인 또한 중요하므로 IB는 이에 대한 공식적인 규정을 두고 있지는 않습니다. 일반적으로 학생 수가 수십 명에 불과한 경우에는 전체 결과에서 상당한 편차가 발생할 수 있지만, 학생 수가 수천 명에 달할 경우 이러한 수준의 편차가 나타날 가능성은 상대적으로 낮습니다.

근거 간의 균형 맞추기

어떤 유형의 근거도 다른 근거보다 더 중요하지 않으며, 성적 부여 팀의 과제는 권장 사항을 마련할 때 모든 근거를 균형 있게 종합하는 것입니다.

[그림 35] 성적 급간의 선정을 뒷받침하는 근거



종종 서로 다른 근거가 동일한 결론을 시사하기도 합니다. 예를 들어 통계적 권장 급간이 불확실의 구간에 위치하는 경우가 이에 해당합니다. 그러나 때로는 이러한 근거가 상충하는 경우도 발생합니다. 이러한 경우에는 최종 결정을 정당화할 때 불일치를 반드시 설명해야 합니다.

성적 부여 과정의 마지막 단계에서는 서로 다른 결정이 집단의 전반적인 결과에 어떤 영향을 미칠지를 모형화할 수 있습니다. 개별 항목의 결과가 아닌 전체 결과가 중요하다는 점을 상기하면서 성적 부여 팀은 전체 학생 집단의 평균 성적이 교사의 예측 성적과 어떻게 다른지 비교해볼 수 있으며, 이러한 비교가 과거에도 얼마나 신뢰할 수 있었는지 확인하기 위해 이전 연도의 자료를 참고할 수 있습니다. 또 다른 접근 방법은 전체 결과에서 신규 학교를 제외해 봄으로써 이들이 기존 학교와 매우 다른 성과 양상을 보이는지 확인하는 것입니다.

마지막으로 성적 부여 팀은 권장 사항과 그에 대한 근거를 함께 제시하여 승인을 받아야 합니다. 그런 다음 종합 과정이 진행됩니다. 종합이란 서로 다른 항목의 점수와 급간을 결합하여 최종 점수 또는 전체 성적 급간을 형성하는 과정입니다. 이를 위해 전체 항목 점수 또는 급간을 환산해야 할 수 있습니다. 자세한 내용은 본 간행물의 '종합' 섹션을 참고하십시오.

고정된 성적 급간

다수의 내부 평가 과제와 일부 외부 채점 항목의 경우, 학생이 수행해야 하는 과제가 기수마다 본질적으로 동일합니다. 예를 들어 예술 포트폴리오나 개인 프로젝트의 준비 등이 이에 해당합니다. 이러한 경우 IB가 매년(전년도 과제의 표준화를 통해) 동일한 채점 기준을 유지하는 만큼 성적 급간 역시 일관되게 유지될 것이라고 합리적으로 가정할 수 있습니다.

성적 부여 과정에서 책임 채점관과 수석 채점관은 기존 성적 급간이 더 이상 적절하지 않다는 근거가 있는지를 검토하도록 요청받으며, 대부분의 경우 IB는 기존 성적 급간을 그대로 유지하는 것이 타당하다는 결론에 이를 것으로 예상합니다.

그럼에도 불구하고 이러한 성적 급간은 매년 고려되며, 기수 간 성적의 비교 가능성을 유지하는 데 더 이상 효과적이지 않다는 근거가 있을 경우 급간이 조정될 수 있다는 점을 분명히 하는 것이 중요합니다. 이러한 상황의 예로는 과제 수행과 관련한 테크놀로지의 새로운 발전과 같이 학생의 행동에 중대한 변화가 생긴 경우, 과제의 전반적인 질은 향상되지 않았음에도 제출된 과제가 채점 기준표와 더 잘 부합하게 된 경우, 또는 과제 채점이나 조정에 대한 IB의 접근 방법이 수정되어 채점 기준 자체가 변동된 경우 등이 있습니다.

성적 급간이 변경되는 이유와 그 의미에 대한 자세한 내용은 “[Setting grade boundaries: Why grade boundaries can change, and what it means if they do](#)”(영문 제공, 동영상)를 참고하십시오.

성적 급간 권장 사항 검토 및 승인

성적 부여 팀이 성적 급간의 위치에 대한 권장 사항을 도출하면 이 권장 사항과 그 근거는 **24시간 배포 과목 보고서**에 제시됩니다. 이 보고서에는 과정을 수강한 집단의 변화, 제안된 급간을 적용했을 때의 성적 결과와 전년도 결과의 비교 등 관련 데이터가 포함됩니다. 이후 이 보고서는 IB 평가 부서의 선임 구성원들에 의해 검토되어 제시된 논거가 충분히 타당한지 판단됩니다.

IB 평가 리더십 팀이 권장 사항에 이의가 있는 경우, 과목 담당자와 우려 사항을 논의하여 특정 근거에 더 큰 비중을 두거나 권장 사항을 뒷받침하는 추가 분석을 제공하도록 요청합니다.

IB의 수석 평가 담당자는 성적 부여 팀의 권장 사항을 바탕으로 성적 급간의 최종 위치를 결정할 권한을 가집니다.

프로그램 수료증 부여하기

IB 프로그램 수료증(MYP 수료증, IB 디플로마, IB CP 수료증)은 성적 부여 과정만으로 결정되는 것은 아닙니다. 수료증 수여 여부는 결과 발표 준비 단계에서 결정되며 이에 대한 절차는 ‘[결과 발표 준비](#)’ 섹션에서 확인할 수 있습니다.

교사 관찰자

IB는 평가 과정의 투명성을 높이고 성적 부여 방식에 대한 일반적인 이해를 증진하기 위해 노력하고 있습니다. 이러한 노력의 일환으로 교사 관찰자가 성적 협의 위원회에 참석합니다(구체적인 사항은 회의가 대면으로 또는 비대면으로 열리는지에 따라 달라집니다). 교사 관찰자는 자신의 경험을 동료 교사에게 공유하고, 더 넓은 IB 교육 전문가 네트워크(IBEN) 공동체를 위한 보고서를 작성할 의무가 있습니다. 자세한 내용은 [IB 지원팀](#)에 문의해 주시기 바랍니다.

성적 부여 과정의 원칙

IB 성적 부여 과정의 기본 원칙은 다음과 같습니다.

- 3/4, 6/7, 2/3 성적 급간은(이 순서대로) 사용 가능한 모든 근거(판단에 근거한 증거 및 통계적 증거)를 활용하여 결정됩니다. 이러한 성적 급간을 설정하기 위한 평가 과제가 제출되지 않은 경우, 사용 가능한 근거를 활용하여 가장 적절한 급간을 결정합니다.

- 그 밖의 성적 급간은 정해진 절차에 따라 산술적으로 결정됩니다.
- 성적 급간은 채점관 판단의 증거, 통계적 증거, 학생 집단에 대한 정보를 균등하게 고려하여 결정됩니다. 모든 근거는 서로 균형 있게 고려되어야 하며 이를 바탕으로 절충안이 도출됩니다.
- 평가 대상 집단이 전년도와 대체로 유사한 경우 IB는 평가 결과도 전반적으로 유사할 것으로 예상합니다. 그러나 다음 사항을 고려해야 합니다.
 - 학생 집단은 보통 해마다 달라지며, 특히 학생 수가 적은 과목일수록 차이가 뚜렷해질 수 있습니다.
 - 결과가 전년도와 다른 경우 IB는 그 차이를 설명할 수 있는 충분한 근거를 요구합니다.
 - 평가 과제가 전년도와 대체로 유사하다면 IB는 성적 급간 또한 전년도와 유사할 것으로 예상합니다.
 - 모든 성적 급간은 해마다 변경될 수 있으며, 내부 평가 과제도 예외가 아닙니다.
 - IB는 매년 평가 난이도의 일관성을 유지하기 위해 최선을 다하지만, 개별 시험의 요구 수준이 달라질 수 있음을 인지하고 있습니다.
 - 전체 과정 성적 급간이 가장 우선시됩니다. 이는 학생에게 가장 중요한 결과이며, 이해관계자가 의사 결정을 내릴 때 활용됩니다.
 - 항목 성적 급간은 타당한 전체 과정 결과를 도출하는 데 중요한 단계이지만, 항목 수준에서의 작은 차이들이 누적되면 전체 과정 결과에 상당한 영향을 미칠 수 있습니다.

품질 평가

- 평가의 가장 중요한 결과는 학생이 받게되는 성적이기 때문에 최종 검토는 성적에 초점을 맞춥니다.
- 이러한 최종 검토의 목적은 일반적이지 않은 요소를 확인하는 데 있습니다.

성적 급간이 설정된 이후 학생에게 부여되는 성적이 공정하고 올바른지 확인하기 위해 추가적인 품질 평가가 이루어집니다.

이러한 평가 중 하나는 ‘이상사례’의 재채점입니다. 여기서의 예상보다 낮은 최종 성적을 받은 학생을 식별합니다. ‘이상사례’ 재채점 중 일부는 추가 검토가 필요한 것으로 파악된 채점관의 채점 결과에 초점을 둡니다. 재채점은 책임 채점관의 채점 기준을 일관되게 적용할 수 있는 것으로 입증된 채점관에 의해 수행됩니다. ‘이상사례’ 재채점은 추가 점수를 발굴하기 위한 것이 아니라, 현재 점수가 적절한지 확인하여 학생이 올바른 성적을 받도록 하기 위해 수행됩니다.

또한 IB 평가 직원도 학교의 과목 성적이 적절해 보이는지 확인하는 점검을 수행합니다.

IB 최종 학위 수여 위원회

- IB 최종 성적 부여 위원회는 IB가 MYP, DP, CP의 평가 기수의 개요를 설명하고, 초대된 다양한 이해관계자들이 이를 면밀히 검토하는 공식적인 자리입니다. PYP에는 검증하거나 인증할 최종 결과가 없으므로, IB 최종 학위 수여 위원회가 존재하지 않습니다.
- IB 최종 학위 수여 위원회는 선임 IB 직원 및 수석 채점관을 포함하여 조직 전반의 대표자들로 구성됩니다. 이 위원회는 결과 발표를 최종 승인할 뿐만 아니라, IB의 학문적 진실성 정책과 관련된 사안 및 기타 관련 이슈도 검토합니다.

IB 최종 학위 수여 위원회는 IB에 제출한 학생의 IB 프로그램의 평가 결과 승인의 마지막 단계를 담당합니다. 이 위원회는 시험 기수가 IB 기준을 충족하였으며 IB 이사회를 대신하여 결과를 수여해야 한다는 수석 교육 책임자(Chief Education Officer)와 평가 담당 책임자(Director of Assessment)의 권고 사항을 검토합니다. 또한 IB 자격 수여와 관련된 정책과 선례를 수립하는 공식 논의의 장이기도 합니다.

이 위원회의 정확한 소관 업무와 구성은 프로그램별로 약간 다르지만, 일반적으로 다음 출신의 의결권을 가진 동일 인원으로 구성됩니다.

- 교육 부서 및 학교 부서를 담당하는 선임 IB 직원
- 다양한 교과군의 수석 채점관

성적 급간 승인은 선임 평가 담당자의 책임이지만, IB 최종 학위 수여 위원회는 감독 이사회로서 프로그램 전체 이수율과 IB 평가 부서 동료들이 제기한 주요 이슈 등 거시적 결과를 검토합니다.

또한 학문적 부정행위와 행정 실책 사안도 검토하며, 이는 학문적 진실성 하위 위원회에서 심도 있게 논의된 사안입니다.

마지막으로, IB 최종 학위 수여 위원회는 시험 기수의 성과를 검토하고 차기 시험 기수를 위한 권장 사항을 IB에 제시하는 것입니다.

이해의 상충

IB 최종 학위 수여 위원회는 중대한 의사 결정 기구로서, 투명성과 독립성을 유지하는 것이 매우 중요합니다. 특정 안건에서 이해 상충 가능성이 있는 것으로 보이는 구성원은 해당 논의가 진행되는 동안 회의장에서 퇴장해야 하며, 이 원칙은 매 회의 시작 시 반복적으로 고지됩니다.

관찰자

IB는 다음과 같은 목적을 위해 관찰자의 IB 최종 학위 수여 위원회 회의 참석을 장려합니다.

- 절차의 투명성을 강화하기 위함
- 과정의 개선 및 변경 사항을 제안할 기회를 제공하기 위함

- IB 구성원 간의 파트너십을 공식적으로 인정하기 위함

이러한 목표를 달성하기 위해 IB의 지원을 돕는 차원에서, 관찰자는 회의 종료 후 2주 이내에 수석 교육 책임자에게 참석 보고서를 제출해야 합니다. 이 보고서에는 현재 시행 중인 절차와 과정에 대한 일반적인 관찰 내용과 필요한 경우 제안할 변경 사항 및 개선점이 포함되어야 합니다. 위원회에서 다루어진 개별 사례에 대한 의견을 제시해서는 안 됩니다.

위원회 논의 사안의 민감성을 고려하여, 관찰자는 기밀 유지와 개인적인 이해 상충 관련 적절한 규제를 준수해야 합니다. 관찰자는 위원회의 결정에 대한 투표권을 갖지 않습니다.

IB 최종 학위 수여 위원회 회의 관찰자에 관한 자세한 정보는 support@ibo.org로 문의하시기 바랍니다.

결과 발표 준비

- 모든 평가가 채점되고 성적 급간이 결정된 후 학생의 결과가 발표될 준비를 마칠 때까지 여러 절차를 완료해야 합니다.
- 특정 상황에서 평가를 위해 제출된 과제의 근거가 없는 경우, IB는 분실 답안 채점 절차를 통해 점수를 추정할 수 있습니다.
- 전체 프로그램에 등록된 학생들의 과목별 결과가 종합되어 MYP 수료증, IB CP 수료증, IB 디플로마 취득 여부가 결정됩니다.
- 결과 발표 과정에는 엄격한 변경 관리 시스템(change-control system)이 적용됩니다. 성적 변경 사항은 모두 식별되고, 관련 이해관계자(주로 학교와 대학)에게 전달됩니다.

성적 급간이 설정된 이후부터 학생에게 결과를 발표하기까지 여러 과정과 절차가 수행되어야 합니다. 예를 들어, 예상 과목 이월과 같은 일부 과정은 특정 학생에게만 적용되며, 그 외 과정은 모든 학생에게 영향을 미칠 수 있습니다.

분실 답안 채점

- 학교나 학생의 책임이 아님에도 IB가 학생 과제에 접근할 수 없는 경우, 학생에게 발생할 수 있는 불이익을 최소화하기 위해 IB는 점수를 추정합니다.
- 이러한 추정은 반드시 근거에 기반해야 합니다. 점수 추정의 근거로 활용할 평가 과제가 거의 없는 경우, 분실 답안에 대한 다른 해결책을 찾아야 합니다.
- 분실 답안 채점 절차는 평균 학생 성과에 기반하며 이에 따라 이러한 추정으로 이익을 받는 학생의 수와 불이익을 받는 학생의 수는 대체로 비슷합니다. 가장 공정한 결과는 항상 학생의 실제 과제를 채점하는 것입니다.

때때로 학교와 학생의 통제를 벗어난 이유로 학생의 과제를 IB가 채점할 수 없는 경우가 있습니다. 예를 들어 서면 기반 시험지가 우편 발송 중 분실되었거나, 시험 당일 학생이 갑작스럽게 아픈 경우 등이 이에 해당합니다.

분실 답안 채점 절차는 이러한 상황에서 IB가 해당 학생의 점수를 추정하기 위해 사용할 수 있는 메커니즘입니다.

학생 과제가 IB 또는 제3자(학교 제외)의 행위로 인해 분실되어, 학생에게 다른 시기에 동일한 평가를 다시 치르도록 요구하는 것이 합리적이지 않은 경우에 이 절차를 적용하는 것이 적절합니다.

모든 경우에 분실 답안 채점 절차는 반드시 학생의 성과를 입증할 수 있는 근거에 기반해야 합니다. IB가 학생 성과에 대한 정보를 충분히 보유하지 못한 경우, 공정한 추정 점수를 부여하기 매우 어렵습니다.

분실 답안 채점 절차는 특정 평가 항목에서 전체 학생의 평균 성취도와 다른 항목에서의 성과를 비교한 결과에 기반하여 이루어집니다. 이와 같이, 평균치를 활용함으로써 IB는 이 추정으로 인해 성과가

더 좋아지는 학생 수와 더 나빠지는 학생 수가 대체로 비슷할 것으로 예상합니다. 따라서 IB가 학생의 실제 과제를 채점할 수 있을 때 항상 더 공정한 결과를 얻을 수 있습니다.

학교에서 특정 과목과 수준에 등록된 학생이 5명 이상인 경우, IB는 해당 학교의 데이터를 활용하여 분실 답안 점수를 결정합니다. 등록 학생 수가 5명 미만일 경우, IB는 전 세계 데이터를 활용하여 분실 답안 점수를 산출합니다.

[그림 36] 학생의 분실 답안에 대한 점수 결정하기

응시생의 분실된 점수를 추정하는 방법:

$$\text{비율} = \frac{\text{다른 항목에서의 응시생의 환산 점수}}{\text{다른 항목에서의 세계(학교)의 총 환산 점수}}$$

$$\text{응시생의 분실된 점수} = \text{분실된 평가 항목에 대한 세계(학교)의 환산 점수} \times \text{비율}$$

비율은 학생이 완료한 항목에 대해 학생의 점수를 세계 평균 또는 학교 평균과 비교합니다.

- 학생이 평균보다 높은 성과를 보인 경우 이 값은 1보다 큽니다.
- 학생이 평균보다 낮은 성과를 보인 경우 이 값은 1보다 작습니다.

이 산출은 다른 모든 분실 답안 채점 절차와 마찬가지로, 사용 가능한 근거에 기반한 ‘최선의 추정치’일 뿐입니다. 따라서 항상 학생의 실제 과제를 채점하는 것이 가장 바람직합니다.

분실 성적 절차

MYP의 경우, 특정 과목의 평가 모델에 항목이 하나만 있는 경우에는 점수를 추정할 근거가 없으므로 분실 답안 채점 절차를 적용할 수 없습니다. 이러한 상황에서는 IB가 ‘MYP 분실 성적 절차’라는 별도의 접근 방법을 적용합니다.

학생의 부담을 줄이기 위해 평가 항목을 하나만 두는 구조의 단점은, MYP에서 활용할 수 있는 근거가 극히 제한된다는 점입니다. 이는 분실 성적 절차에 대한 IB의 신뢰도가 다른 프로그램의 분실 답안 채점 절차보다 낮음을 의미하며, 따라서 이 절차는 예외적인 경우에만 적용해야 합니다.

학생이 요구 사항에 따라 평가를 성실히 수행하고 제출하였으나, 학생이나 학교가 통제할 수 없는 요인으로 IB가 해당 과제를 채점할 수 없는 경우, IB는 분실 성적 절차를 적용합니다.

이 절차를 적용하려면 학생이 MYP 내 최소 4개 이상의 다른 과정에서 최종 성적을 받았어야 합니다. 그보다 적은 과정의 성적만 있는 경우에는, 정보에 기반한 추정을 할 수 있는 충분한 자료가 없으므로 IB는 결과를 산출할 수 없습니다.

분실 성적 산출 방식은 지난 18개월간 평가를 위해 제출된 과제를 기반으로 성적이 부여된 다른 모든 과목의 평균 성적을 산출합니다.

평균 성적이 정수보다 0.5 이상 높을 경우 반올림합니다.

평균 성적이 정수보다 0.5 미만 높을 경우 내림합니다.

이 산출은 다른 모든 분실 답안 채점 절차와 마찬가지로, 사용 가능한 근거에 기반한 ‘최선의 추정치’일 뿐입니다. 따라서 항상 학생의 실제 과제를 채점하는 것이 가장 바람직합니다.

프로그램 결과

- IB는 개인의 판단이 아닌 기준의 충족 여부에 따라 디플로마 및 수료증을 수여합니다.
- 각 프로그램마다 디플로마 또는 수료증 수여를 위한 고유한 기준이 있습니다.

IB가 학생의 모든 과목 결과를 취합하면 해당 학생이 MYP 수료증, IB 디플로마 또는 IB CP 수료증 취득 자격을 갖추었는지 계산할 수 있습니다.

MYP, DP, CP 결과 산출 방법과 합격 기준에 관한 자세한 내용은 [IB 웹사이트](#) 및 관련 프로그램의 「평가 절차」에서 확인할 수 있습니다.

결과 발표

IB의 관점에서 결과 발표는 매우 엄격한 변경 관리 프로토콜을 실행하는 것을 의미합니다.

시험 기수 동안 관련 정보는 지속적으로 업데이트되어 IB가 전체 진행 상황을 명확히 파악할 수 있습니다. 학생과 학교에 결과가 발표된 이후에는, 모든 관련 당사자에게 적절히 공지되지 않은 상태에서 결과가 변경되지 않도록 명확히 관리해야 합니다. DP와 CP의 경우 특히 학생의 성적 증명서를 수령하는 대학과 직접적으로 관련됩니다.

결과 발표 후 승인된 변경은 성적에 대한 문의, 학교의 보류 중인 결과 확인, 또는 학업 진실성 사례 해결로 인해 발생할 수 있습니다.

성적에 대한 문의 및 평가에 대한 이의 제기

- 성적에 대한 문의 서비스의 목적은 학교가 채점 과정에서 실수가 일어났다고 생각하는 부분을 IB에 제기할 수 있도록 하는 데 있습니다.
- 성적에 대한 문의 과정에서도 주 시험 기수와 동일한 기준이 적용되어야 하며, IB는 이를 보장하기 위해 시드 작업 품질 모델을 활용합니다.
- 외부 평가는 형성 평가가 아닌 총괄 평가 목적이기 때문에, IB는 채점관에게 채점을 뒷받침하는 경우에만 의견을 기재할 것을 요구합니다.
- 또한 학교가 올바른 절차가 준수되지 않았다고 판단하는 경우를 대비해 IB는 공식적인 이의 제기 절차를 마련해 두고 있습니다.

제공되는 성적에 대한 문의 서비스의 법적, 절차적 설명은 관련 프로그램의 「평가 절차」를 참고하십시오.

성적에 대한 문의 카테고리

결과가 발표된 이후 코디네이터는 다음 중 하나를 요청할 수 있습니다.

- **카테고리 1 성적에 대한 문의:** 특정 과목에서 학생의 모든 외부 평가 항목에 대한 재채점
- **카테고리 1 보고서:** 카테고리 1 성적에 대한 문의의 채점에 관한 보고서
- **카테고리 2 성적에 대한 문의:** 외부 평가 항목 자료의 사본
- **카테고리 3 성적에 대한 문의:** 내부 평가 항목의 재조정

각 카테고리에는 수수료가 부과됩니다(카테고리 1 재채점 동안 성적이 변경된 경우는 제외).

성적에 대한 문의 서비스를 제공하는 이유

성적에 대한 문의 과정은 채점 시스템에서 발생할 수 있는 오류를 방지하기 위한 최종 안전장치입니다. 이를 통해 학교는 오류가 발생했다고 판단되는 경우 이를 IB에 제기할 수 있으며, IB는 필요한 경우 해당 오류를 조사하고 수정할 수 있습니다.

실제 채점부터 성적에 대한 문의 채점까지 기준 유지하기

성적에 대한 문의의 목적은 학생에게 IB 기준에 따라 학생 과제의 질을 반영한 점수를 제공하는 것입니다. 채점관은 학생이 초기 점수에 만족하지 못하거나 성적 급간에 근접해 있다는 사실에 부적절한 영향을 받지 않도록 하는 것이 매우 중요합니다.

성적에 대한 문의 과정에서 IB는 가장 일관된 선임 채점관만을 배정하며, 이 재채점 결과를 최종적으로 올바른 점수로 간주합니다.

채점관이 기준을 유지할 수 있도록 IB는 성적에 대한 문의 과정에 시드 답안지를 포함하여 기준에서 벗어날 경우 전자 형식으로 알림을 보냅니다.

IB 평가 직원은 성적에 대한 문의 과정에서 제안된 모든 성적 변경을 검토하여 변경된 성적이 기존 성적보다 더 신뢰할 수 있는지 확인합니다. 두 채점관 간 점수 차이가 크고 그 차이에 대한 설명이 충분하지 않으며 성적 변경이 발생한 경우, 제3의 채점관에게 추가 의견을 요청할 수 있습니다.

카테고리 1 성적에 대한 문의: 재채점

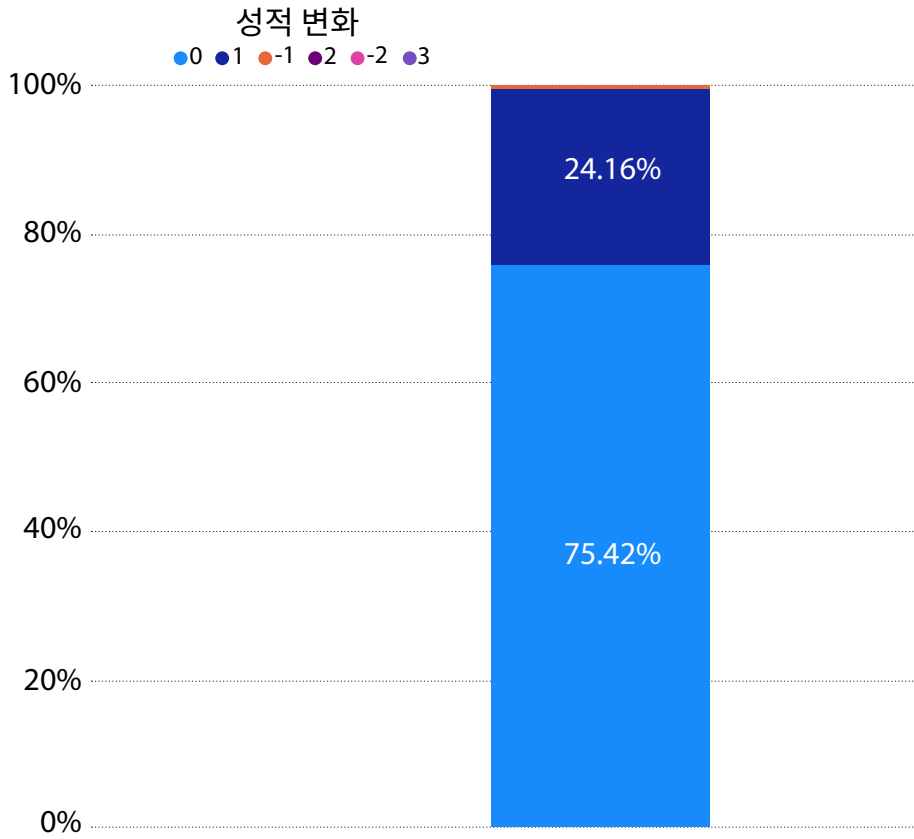
‘채점’ 섹션에서 설명한 바와 같이, 동일한 문제에 대해 두 채점관이 허용 오차 범위 내에서 서로 다른 점수를 부여할 수 있습니다. 따라서 성적에 대한 문의 과정에서 점수가 소폭 변경되었다면, 이는 채점 과정에 실수가 없었을 가능성이 큼니다. IB에 접수되는 성적에 대한 문의의 대부분이 이러한 경우에 해당합니다.

채점관 간의 작지만 허용 가능한 점수 차이가 결과에 미치는 영향을 최소화하기 위해 IB는 카테고리 1 성적에 대한 문의가 접수되면 학생이 수행한 모든 외부 평가를 재채점합니다. 이는 변화가 시스템적 문제에서 비롯된 것이 아니라는 전제하에, 여러 시험에 걸쳐 발생한 작은 차이가 서로 상쇄될 것으로 IB는 기대합니다.

IB는 성적에 대한 문의 점수가 학생의 과제를 가장 정확히 반영한다고 판단하며, 이를 최종 점수로 부여합니다. 따라서 카테고리 1 성적에 대한 문의 결과에 따라 학생의 성적은 하향되거나 상향될 수도 있습니다. 이 때문에 학교는 카테고리 1 성적에 대한 문의를 제출하기 전에 반드시 학생의 동의를 받아야 합니다.

성적 급간 설정의 특성상, 급간보다 1점 낮거나 1점 높은 학생은 항상 존재합니다. 이러한 경우 채점 과정에 실수가 없더라도 성적에 대한 문의의 결과로 성적이 변경될 수 있습니다. 이와 같은 상황에서 IB는 해당 학생이 두 성적 경계에 위치하며, 어느 성적이든 학생의 성과를 공정하게 반영할 수 있음을 강조합니다.

[그림 37] 5개 DP 시험 기수에 걸친 카테고리 1 성적에 대한 문의의 평균 성적 변화



카테고리 2 성적에 대한 문의: 학생 자료의 반환

카테고리 2 성적에 대한 문의의 대상이 되는 평가는 총괄 평가입니다. 반환된 답안지에는 배정된 채점관의 유용한 의견이 포함될 수 있습니다. 그러나 채점관이 학생 과제를 채점할 때 의견을 반드시 남겨야 하는 것은 아니므로, 의견이 포함되지 않을 수도 있습니다.

IB는 학생이 제출한 과제를 반영하여 공정한 결과를 얻을 수 있도록 채점관이 적절히 채점할 것을 요구합니다. 채점관은 채점을 뒷받침하는 데 필요한 의견만을 작성하며, 학생의 향후 개선을 위한 피드백을 제공하지 않습니다.

과제가 공정하게 채점되지 않았다고 판단되는 경우 어떻게 해야 하는지에 대한 자세한 내용은 “Challenging results: What you can do if you think your work hasn’t been marked fairly”(영문 제공, 동영상)를 참고하십시오.

카테고리 3 성적에 대한 문의: 재조정

카테고리 3 성적에 대한 문의(재조정)의 경우 점수 변경이 발생하면 전체 학생 집단에 동일하게 적용됩니다. 만약 학생을 개별적으로 동의를 구하는 원칙을 적용한다면, 한 명의 학생이 동의하지 않거나 요청에 답하지 않는 것만으로도 전체 학생 집단의 재조정이 이루어지지 못하는 상황이 발생할 수 있습니다. 이는 명백히 불합리하기 때문에, 이러한 위험을 학생이 공식적으로 수용하지 않은 상황에서

성적이 하향되는 것을 방지하기 위해 카테고리 3 성적에 대한 문의의 결과로 성적은 **상향만** 가능하고 하향은 불가합니다.

재조정은 조정된 내부 평가 점수의 평균과 학생의 원점수(교사가 부여한 점수)평균 간의 차이가 해당 항목 최고점의 **15% 이상**일 때만 수행됩니다.

이 기준이 마련된 것은 IB가 아무런 제한을 두지 않으면 일부 학교들이 모든 과목에 대해 무분별하게 재조정을 요청하여 성적 상승을 기대하는 상황을 방지하기 위한 것입니다. IB에는 선임 조정자의 수가 제한되어 있기 때문에 이처럼 과도한 요구를 감당할 수 없습니다. 재조정 요청을 수용하기 위한 기준을 설정함으로써, 명확한 사유가 있는 학교만 이 서비스를 이용할 수 있도록 한 것입니다. 이를 통해 IB는 제한된 인원의 선임 조정자의 업무량을 관리하는 동시에 조정 결과가 학생 성적에 심각한 영향이 있었던 학교에게 재조정 기회를 제공할 수 있게 합니다.

[그림 38] 카테고리 3 성적에 대한 문의 자격이 있는 전체 학생 집단

학생 번호	샘플 사용 여부	원점수	조정 점수	조정 차이
2	Y	19	14	-5
10		15	12	-3
11	Y	16	13	-3
12		20	15	-5
13		18	14	-4
16	Y	12	11	-1
36	Y	17	13	-4
45	Y	10	10	0
점수 기준 평균 변화				-3.125
최대 점수 기준 평균 변화 비율				-15.63%

[그림 38]의 예시에서 해당 항목에서 받을 수 있는 최대 점수는 20점입니다.

평균 조정 점수는 -3.125점으로, 이는 최대 점수의 15.63%에 해당합니다.

이는 해당 샘플이 카테고리 3 재조정의 자격을 갖추었음을 의미합니다.

평가에 대한 이의 제기

평가에 대한 이의 제기는 평가 과정에서 내려진 결정 즉, 학생의 IB 성적 산정 방식이나 학업적 부정행위 대처 방식을 IB에 검토 요청할 수 있는 절차입니다.

이의 제기를 통해 IB 과정에 대한 예외 적용을 요청할 수는 없지만, IB 규정의 해석이 합리적인지에 대한 이의 제기는 할 수 있습니다. 이의 제기로 IB가 공개적으로 채택한 정책을 변경할 수는 없으며, 이의 제기의 결과로 IB 규정을 명확하게 수정해야 하는 경우, 해당 기수의 모든 학생에게 적용되어야 합니다. 이는 IB 평가 과정에 대한 우려의 최종 단계를 나타냅니다.

자세한 내용은 프로그램 자료실에 게시된 MYP, DP 또는 CP의 「평가 절차」간행물을 참고하십시오.

재응시생

학생이 받은 성적에 만족하지 못할 경우 학생은 해당 과목을 다시 응시할 수 있습니다. 이는 6개월 이후의 시험 기수 또는 원칙적으로 그 이후의 모든 시험 기수에서 재응시가 가능합니다. 그러나 중요한 새 교육과정이나 평가 요건이 도입되었다면, 재응시 학생은 반드시 새로운 요구 사항을 준수해야 합니다. 1개 또는 그 이상의 과목을 다시 응시하려는 학생은 해당 과목을 원래 수강했던 학교에 다시 등록하지 않아도 됩니다. 학교에서 1개 또는 그 이상의 과목을 재응시하려는 학생을 수용할 경우 학교는 해당 학생에 대한 학업적, 행정적 책임을 부담해야 합니다.

교육과정 및/또는 평가 요구 사항에 중대한 변화가 없을 경우, 재응시 학생은 비시험 요소의 결과를 이월할 수 있습니다. 서면 시험 점수는 이월할 수 없으므로, 해당 과목의 모든 시험을 재응시해야 합니다.

비시험 요소 과제를 다시 제출하려는 경우, 학생은 재응시 기수에 등록된 학교의 수업을 수강해야 합니다. 이는 과목 교사가 학업적인 지도를 하고, 내부 평가 과제를 채점하며, 모든 과제의 진위성을 확인해야 하기 때문입니다. 학생이 이전에 제출한 비시험 요소 과제를 일부 수정하는 것만으로는 충분하지 않습니다. 원칙적으로 평가를 위해 내용이 완전히 다른 과제를 제출해야 합니다. 그러나 과제가 매우 방대한 경우, 완전히 새로운 과제를 제출하는 것이 시간 제약 때문에 어렵다는 것을 IB는 인지하고 있습니다. 이러한 경우, 원래 과제의 상당한 수정이 허용됩니다. 수정된 과제가 제출되면 새로운 과제로 간주되어 채점되며 학생은 더 낮은 점수를 받을 수도 있음을 인지해야 합니다. 수정된 과제의 경우 학생은 반드시 과제에 할애했던 총 시간의 최소 20% 이상을 추가로 투입해야 합니다. 또한 과제는 다음 기준 중 적어도 하나 이상을 충족해야 하며, 다음을 포함해야 합니다.

- 변경된 연구 질문, 논점 또는 논지
- 결론에 영향을 미치는 새로운 주제, 근거 구성요소 또는 관점
- 실질적 추가 분석/계산
- 저자 또는 예술가의 작품에 대한 새로운(추가된) 측면 탐구

학생이 이전에 이미 완료했지만 평가의 일부로써 제출하지 않은 과제(예: DP 예술 포트폴리오)가 있는 경우, 해당 과제를 만드는 데 이미 투입한 시간은 새로 제출하는 과제의 총 소요 시간의 20% 요건을 충족하는 데 포함될 수 있습니다.

이 접근 방법은 두 가지를 균형 있게 고려하여 선택되었습니다. 하나는 재응시 학생이 과제물에 대한 추가 피드백으로 혜택을 받아서는 안 된다는 기대(새 과제물을 제출하는 학생에게 불리하므로)이고, 다른 하나는 재응시를 사실상 불가능하게 만들 완전히 새로운 데이터나 연구 확보 등 과도한 부담을 학생이나 학교에 지우는 것은 공정하지 않다는 인식입니다.

비교 가능성에 대한 IB의 접근 방법은 본 간행물의 **‘비교 가능성’** 섹션에서 확인할 수 있습니다.

IB는 시간이 지나면 과정의 내용이 변경될 필요가 있음을 알고 있습니다. 새로운 과정을 도입하기 전 IB는 관련 프로그램의 최신 정보와 새로운 교과 가이드의 제작을 포함한 다양한 방식으로 학교에 모든 변경 사항을 전달합니다. 이전 과정의 마지막 시험 기수가 종료된 후 IB는 재응시생에게 기존 과정 시험에 응시할 기회를 제공할 수 없습니다. 이는 동일한 시험 기수에서 동일한 과목에 대해 서로 다른 두 가지 과정을 기준으로 평가하는 것이 학교와 IB 모두에게 관리상 불가능하기 때문입니다. 다만, 내부 평가 점수 또는 외부 평가 과제 점수를 이월하는 것이 가능할 수도 있습니다. 새로운 과정이 이전 과정과 상당히 다른 경우, 이러한 선택지가 제공되지 않을 가능성이 높습니다.

재응시생에 관한 자세한 내용은 해당 프로그램의 「평가 절차」 간행물을 참조하십시오.

학교에 제공되는 평가 피드백

- IB 총괄 평가의 목적은 학생의 성과를 측정하는 것이며, 모든 관련 과정은 결과의 타당성을 극대화하도록 설계되어 있습니다.
- IB는 학교 및 학생에게 평가 결과를 전달할 때 투명성을 최우선으로 하여, 점수와 성적이 일관되고 편향 없이 부여되었음을 확인할 수 있도록 합니다.
- 투명성은 교사 채점의 결과를 기반으로 결과가 도출되는 내부 평가에서 특히 중요합니다.
- 성적에 대한 문의 과정을 포함한 IB 총괄 평가는 타당한 평가에 필요한 데이터의 부산물로 얻어지는 것을 제외하고는, 학생 성과 향상을 위한 지침을 학교에 제공할 목적으로 설계되지 않았습니다.
- IB는 평가 과정 외에도 학교의 교수법과 전문성 신장을 지원하는 다양한 서비스를 제공합니다.

IB 총괄 평가의 목적은 학습 과정을 마친 이후 학생의 성과를 평가하고, 성적의 형태로 신뢰할 수 있는 측정 결과를 제공하는 것입니다. 이는 학생의 학습을 지원하기 위해 피드백을 제공하는 것을 목적으로 하는 형성 평가와는 다릅니다. 유의미한 피드백은 형성 평가와 효과적인 교수에서 필수적인 요소이지만, 총괄 평가의 결과를 형성적 목적으로 활용하는 것은 적절하지 않습니다. 그 이유 중 하나는 IB 채점관이 채점 중인 하나의 과제(시험 답안지)를 기반으로 판단을 내리는 반면, 교실에서 IB 교사는 학생에게 양질의 피드백을 제공하는 방법에 대해 결정할 때 광범위한 정보와 개인적 통찰력을 활용하기 때문입니다. 이에 따라 채점관은 채점할 때 학생이나 교사에게 서면 피드백을 제공하지 않습니다.

채점관은 점수가 어디에 부여되었는지 확실하게 명시하고 모호한 부분이 있을 경우 적절한 의견을 제시해 이를 명확히 해야 합니다. 이렇게 하면 IB가 기준을 검증하는 데 도움이 되며 학교가 점수 부여의 근거를 명확히 확인할 수 있도록 투명성을 보장합니다.

과목 보고서

각 시험 기수가 종료되면 IB는 보통 수석 채점관이 작성한 과목 보고서를 발행합니다. 과목 보고서에는 각 평가 항목에 대한 정보와 함께, 어떤 문제 또는 주제에서 학생들이 우수한 성과를 보였는지, 또는 어려움을 겪었는지에 대한 분석이 포함됩니다. 또한 향후 기수에서 교사들이 유사한 문제나 주제에 대해 학생들을 어떻게 준비시킬 수 있는지에 관한 채점 팀의 일반적인 권장 사항이 담겨 있습니다. 과목 보고서에서는 해당 과목의 성적 급간과 항목별 급간도 확인할 수 있습니다.

개별 교사는 과목 보고서의 정보를 자신의 수업 맥락에 맞게 적용해야 합니다. 예를 들어 특정 문제가 전반적으로 낮은 성과를 보였더라도, 해당 교사의 모든 학생이 높은 점수를 받았을 가능성도 있습니다.

내부 평가 피드백 양식

내부 평가에 교사가 부여한 점수는 조정 과정을 거칩니다. 교사는 내부 평가 기준을 IB 채점관과 다르게 해석할 수 있습니다. 예를 들어 교사가 더 엄격하게 채점하여 더 낮은 점수를 부여할 수 있습니다. 내부 평가 조정의 목적은 교사의 채점 기준을 책임 채점관 또는 세계적 기준과 일치하도록 조정하는 것입니다. 교사가 자신의 기준이 세계적 기준과 어떻게 다른지 이해할 수 있도록, IB는 교사의 점수가 허용 오차를 벗어날 경우 내부 평가 채점에 대한 피드백을 제공합니다. 이를 통해 교사는 왜 조정 인자가 적용되었는지 파악할 수 있습니다.

학교에 제공되는 피드백은 샘플에 포함된 학생들이 어떻게 하면 더 나은 결과를 얻을 수 있었는지 설명하기 위한 것이 아닙니다. 그보다는 교사가 각 기준을 얼마나 잘 채점했는지, 평가 기준을 전반적으로 어떻게 적용했는지, 그리고 제출된 과제가 적합했는지에 대한 정보를 제공하는 데 중점을 둡니다.

교과서, 워크숍, 시험 관련 유의사항

각 IB 과정의 교육과정은 교과 가이드에 명시되어 있으며 이를 기반으로 평가가 설계됩니다. IB의 승인을 받은 교과서를 포함한 모든 교과서는 학생과 교사가 교과 가이드에 제시된 과정을 이수하도록 지원하기 위한 보조 자료이며 교육과정의 범위를 정의하기 위해 쓰여진 것이 아닙니다.

따라서 특정 교과서에는 수록되지 않았더라도 교과 가이드에 포함된 주제 또는 그 일부라면, 해당 내용은 언제든지 시험 문제로 출제될 수 있습니다. 그러므로 모든 교사는 교육과정의 범위를 고려할 때 교과 가이드만을 유일한 기준으로 참고할 것을 권장합니다. 시험 출제 과정에서 IB는 흔히 사용되는 교과서들을 대조하여, 시험 문제가 해당 자료의 문제와 중복되지 않도록 점검합니다.

이와 마찬가지로 IB 워크숍에서 제공되는 의견이나 유인물은 교사의 교과 가이드 해석을 도울 수는 있지만, 교과 가이드에 명시된 공식 설명을 대체할 수 없습니다. 교과 가이드에 대한 모든 변경 사항은 IB가 공식적으로 발표하며, 이를 명확히 안내합니다.

프로그램별 과정 정의하기

- 모든 국제 바칼로레아(IB) 프로그램은 동일한 교육철학을 기반으로 하고 있으므로, 모든 IB 평가는 이러한 철학에 부합하기 위해 동일한 원칙과 폭넓은 수행을 따라야 합니다. 평가의 목적은 학생이 서로 다른 교육 단계에서 다양한 프로그램을 진행함에 따라 자연스럽게 변화합니다.
- IB 프로그램에서 외부 총괄 평가의 포괄적인 목적은 학생의 학업 성취도에 대한 타당한 요약 정보를 제공하여, 향후 학업이나 취업을 지원하는 데 있습니다.
- 평가는 초등교육 프로그램(PYP)의 중요한 요소이지만, 이 프로그램에 외부 총괄 평가를 도입하는 것은 적절하지 않습니다.
- 각 프로그램에는 평가 수행 프레임워크 내에서 특정 과정을 요구하는 고유한 특징을 가지고 있습니다.

모든 프로그램의 공통 요소

- IB 프로그램은 폭넓고 균형 잡혀 있으며, 개념 중심적이고, 서로 연계된 교육과정 또는 교육과정 프레임워크를 제공합니다.
- 모든 IB 평가는 설계 단계에서 이러한 IB 교육의 근간을 이루는 측면을 고려해야 하며, 명시적으로 평가되지 않는 경우에도 학습 및 교수에 긍정적인 역류 효과를 가져올 수 있도록 해야 합니다.

[그림 39] IB 프로그램의 연계성



간행물 『IB 교육이란 무엇인가?』(2019)에서 IB 프로그램은 개념적 학습을 장려하며, 여러 교과군에 적용 가능한 강력한 조직화 개념을 중심으로 학습을 통합하고 교육과정의 일관성에 초점을 맞춘다는 점을 강조합니다. 또한 이 프로그램들은 학문 간의 관계를 탐구하고 개별 과목의 범위를 넘어서는 방식으로 세계를 이해하는 학습의 중요성을 부각합니다. IB 프로그램은 학생들에게 폭넓고 균형 잡히며 개념적이고 서로 연계된 학문 및 학습의 경험을 제공합니다.

- **폭넓고 균형 잡힌:** IB 교육은 다양한 교과군에 걸친 폭넓은 학습 내용을 학생이 접할 수 있도록 하는 균형 잡힌 접근 방법을 제시합니다.
- **개념적:** 개념적 학습은 교과군 내외에서 관련성이 있는 폭넓고 강력한 조직화 개념에 초점을 맞춥니다. 이러한 개념은 국가적, 문화적 경계를 뛰어넘습니다. 개념은 학습을 통합하고, 교육과정에 일관성을 더하며, 학문에 대한 이해를 심화시키고, 복잡한 아이디어를 다룰 수 있는 능력을 기르며 새로운 맥락으로의 학습 전이를 가능하게 합니다.
- **연계된:** IB 교육과정 프레임워크는 학습의 동시성을 중요하게 여깁니다. 학생들은 프로그램 기간 동안 여러 과목을 동시에 학습하며, 다양한 분야에 걸친 지식과 경험의 상호관계에 대한 풍부한 이해를 추구하고 연결점을 도출하는 법을 배웁니다. 교과 과정의 목표와 프로그램 요구 사항을

통해 개별 과목의 범위를 뛰어넘는 다양한 방법으로 세상에 대해 배울 수 있는 실질적인 학습의 기회를 제공합니다.

개별 평가를 설계할 때 교사, 학교, IB 출제자는 이러한 근본적인 목표를 성찰하여 우수한 학습 및 교수를 저해하는 문제가 발생하지 않도록 해야 합니다. 항상 긍정적인 역류 효과를 창출하는 것을 목표로 삼아야 합니다.

국제적 차원

IB의 교육 목표는 학생들이 '서로 다른 문화를 이해하고 존중하며, 더 나은 평화로운 세상을 실현하는데 기여할 수 있는, 지식이 풍부하고 탐구심과 배려심이 많은 청소년을 기르는 것'입니다(IB 교육 목표, 2024). IB 프로그램은 여러 국가와 지역(영역)에서 다양한 국적의 학생들이 함께 학습하고 있습니다. 따라서 IB 학습 및 교수에는 국제적 맥락과 국제적 소양을 기르는 데 중점을 두며, 이는 평가에서도 반드시 반영되어야 합니다. 국제적 소양에 대한 IB의 접근 방법의 자세한 내용은 본 간행물의 '[국제적 소양 및 문화 간의 이해](#)' 섹션을 참고하십시오.

추가 문헌

더 폭넓은 IB 교육 프로그램의 근간을 이루는 가치에 대한 자세한 내용은 다음 자료를 참조하십시오.

- 「IB 교육이란 무엇인가?」
- 「IB 학습자상」
- 「중등교육 프로그램: 원리부터 실천까지」, 「디플로마 프로그램: 원리부터 실천까지」, 「Career-related Programme: From principles into practice」(영문 제공)

프로그램별 요구와 해결책

본 간행물의 앞선 섹션에서는 평가 주기를 구성하는 과정들을 설명했습니다. 이러한 일반화된 수행은 모든 IB 프로그램에 공통적으로 적용됩니다. 그러나 각 프로그램은 서로 다른 목적을 지니고 있으므로, 프로그램별 평가의 일부 측면과 프로그램 수료증 수여 여부를 결정하는 기준에도 차이가 존재합니다.

이처럼 프로그램별로 특화된 평가 접근 방법은 매우 중요합니다. 이 섹션에서는 이러한 접근 방법이 무엇이며, 어떻게 관리되는지를 설명합니다.

디플로마 프로그램

디플로마 프로그램(DP) 외부 총괄 평가는 다음과 같은 고유한 특징을 갖습니다.

- 디플로마를 취득하려면 학생은 지정된 과목의 모음을 반드시 수강해야 합니다.
- 디플로마에서의 전체 성과는 최대 45점 만점의 점수로 표현됩니다.
- 핵심 과정은 점수표에 따라 전체 디플로마 결과에 최대 3점까지 반영됩니다.
- 거의 모든 과목은 외부 평가와 내부 평가를 포함한 여러 평가 항목으로 구성됩니다.
- 거의 모든 과목은 표준 수준(SL) 또는 심화 수준(HL)으로 제공되며 전체 디플로마 결과에 동등하게 반영됩니다.

디플로마 프로그램의 목적

평가 결과는 과정 및 프로그램의 목적과 명확히 연결될 때에만 타당성을 가집니다.

DP의 주요 목적은 16세에서 19세 학생에게 도전적이면서 국제적 관점을 지니고 폭넓고 균형 잡힌 교육 경험을 제공하는 것입니다. 이러한 목적을 뒷받침하기 위해, 학생들은 2년에 걸쳐 6개 교과 과목과 핵심 과정을 동시에 학습해야 합니다. 또한 DP는 학생들이 대학 진학, 향후 교육 및 선택한 직업에 필요한 기본적인 학문적 기능을 갖추도록 돕고, 의미 있고 목적 있는 삶을 영위하는 데 필요한 삶의 기능과 가치의 개발을 지원하는 것을 목표로 합니다. 즉, DP와 그 과정 및 핵심 요소에서 평가 결과의 타당성은 학생 개개인의 전체 교육 경험에 대한 고려를 근간으로 합니다.

DP 평가 결과의 타당한 활용

평가 모델과 교육과정을 개발할 때, IB는 DP 과정의 성적과 전체 디플로마 점수가 다음을 결정하는 데 활용될 수 있도록 합니다.

- 대학 입학 선발 또는 취업 적격성
- 학생이 대학 프로그램의 일부 요건을 이미 충족했는지 여부(자격 취득을 위한 학점 인정이나 특정 과목 수강 면제로 인정될 수 있음)
- 학생이 특정 언어로 추가 학습을 수행할 자격이 있음을 보여주는 증거

이러한 타당한 활용이 왜 중요한 고려 사항인지에 대해서는 본 간행물의 ‘타당성’ 섹션을 참조하십시오.

DP의 구성 및 DP 과정

IB 디플로마를 취득하기 위해 학생은 반드시 지식이론(TOK), 소논문(EE), 창의, 활동, 봉사(CAS)라는 핵심 요소와 함께 여섯 가지 과목을 수강해야 합니다.

학생들은 일부 과목을 표준 수준(SL)에서, 일부는 심화 수준(HL)에서 수강하게 됩니다. 각 학생은 HL에서 최소 3개(최대 4개)의 과목을 수강하며 나머지는 SL에서 수강합니다. SL 과정과 HL 과정의 범위는 서로 다르지만, 공통된 교과군 성적 기준표 및 평가 목표를 통해 이 과정들의 성과는 일치됩니다. 학생은 HL에서 더 높은 수준의 지식, 이해 및 기능을 발휘해야 합니다.

제한된 수의 통합교과 과정은 여러 교과군에 걸쳐 인정됩니다. 예를 들어 환경 시스템과 사회는 개인과 사회 교과군과 과학 교과군의 요구 사항을 동시에 충족합니다.

[그림 41] 디플로마 프로그램 설계모형



DP 결과 산출 방법

DP의 타당성은 학생의 성과가 측정되는 방식에 반영됩니다. 전체 디플로마 점수는 6개 과목 각각에서 획득한 성적(1~7)을 합산한 다음 핵심 요소(지식이론, 소논문, CAS)의 성과 점수인 0~3점과 종합하여 산출됩니다. 학생의 최종 점수를 산정 시 SL과 HL 과목은 동일한 비중으로 반영됩니다. 학생이 받을 수 있는 최고 점수는 45점입니다.

핵심 요소 점수표

일반 과목과 달리 지식이론과 소논문은 A~E까지의 성적이 부여되며, 세 번째 핵심 요소인 CAS는 특성상 별도의 성적이 부여되지 않습니다.

핵심 요소의 성과는 전체 디플로마 점수 중 0점~3점에 해당합니다. 지식이론 또는 소논문에서 E 성적을 받거나 CAS 요구 사항을 충족하지 못한 경우 학생은 IB 디플로마 취득 자격을 잃게 됩니다. 핵심 요소에 부여되는 점수는 [그림 42]의 방식에 따라 산출됩니다.

[그림 42] 지식이론 및 소논문 점수의 부여

		지식이론(TOK)				
만 년 성	취득 성적	A	B	C	D	E 또는 N
	A	3	3	2	2	디플로마 취득 자격을 충족하지 못함
	B	3	2	2	1	디플로마 취득 자격을 충족하지 못함
	C	2	2	1	0	
	D	2	1	0	0	
E 또는 N	디플로마 취득 자격을 충족하지 못함					

IB 디플로마 수여 요건

다음 조건들이 모두 충족되어야 학생은 IB 디플로마를 수여받을 수 있습니다.

- CAS 요건을 충족합니다.
- 학생의 총 점수가 24점 이상입니다.
- 지식이론, 소논문 또는 기여 과목에서 N(취득 성적 없음)을 받지 않았습니다.
- 지식이론과 소논문 모두에서 D 이상을 받았습니다.
- 어떤 과목/수준에서도 1점을 받지 않았습니다.
- HL 또는 SL에서 2점을 3회 이상 받지 않았습니다.
- HL 또는 SL에서 3점 이하를 4회 이상 받지 않았습니다.
- HL 과목에서 최소 12점을 받았습니다(4개의 HL 과목을 등록한 학생의 경우, 상위 3개의 성적을 합산함).
- SL 과목에서 최소 9점을 받았습니다(2개의 SL 과목을 등록한 학생은 SL에서 최소 5점을 받아야 함).

이중 언어 디플로마

표준 IB 디플로마의 대안으로, 다음 기준 중 하나 이상을 충족한 학생에게 ‘이중 언어 디플로마’ 수료증이 수여될 수 있습니다.

- 언어와 문학에서 선택한 두 가지 언어 과정을 모두 3점 이상의 성적으로 이수했습니다.
- 개인과 사회 또는 과학의 과목 중 하나를 언어와 문학에서 선택한 언어와 다른 언어로 이수했습니다. 학생은 반드시 언어와 문학의 언어와 개인과 사회 또는 과학에서 선택한 과목에서 모두 3점 이상을 받아야 합니다.

위의 조건을 충족한다면 시범 과목과 통합 교과 과목은 이중 언어 디플로마 수료증 수여에 기여할 수 있습니다.

다음 사항들은 이중 언어 디플로마 수료증 수여에 영향을 주지 않습니다.

- 소논문
- 학교 개설 교육과정 과목
- IB 디플로마를 위한 여섯 가지 과목 외에 학생이 추가로 수강한 과목(‘추가 과목’)

전체 학생 집단에 걸쳐 기준 관리하기

DP 과목의 전체 학생 집단은 한 명의 학생부터 4만 명 이상에 이르기까지 매우 다양합니다. 모든 과목은 동일하게 중요하지만, 실질적인 이유로 학생 수가 많고 적은 과목에는 조금씩 다른 과정이 적용됩니다.

채점관 수준 유지하기

채점관 수가 매우 적은 과목의 경우 본 간행물의 ‘**표준화**’ 섹션에서 설명한 연습 답안지, 채점 자격 취득용 답안지, 시드 답안지를 사용하는 **품질 모델**을 그대로 적용하는 것이 적절하지 않을 수 있습니다.

모든 채점관이 표준화 논의에 참여하는 경우, 채점관들은 이미 채점 기준 설정 과정에 동참한 것이므로, 별도의 연습 답안지나 채점 자격 취득용 답안지를 제공할 필요가 없습니다. 다만, 채점관이 계속해서 기대되는 기준에 맞춰 채점하는지 확인하기 위해 시드 답안지는 여전히 사용됩니다. 그럼에도 불구하고, 실제 사용 여부와 관계없이 일반적으로는 전체 품질 모델이 제작됩니다.

여기서 중요한 점은 해당 과목의 학생 수가 아무리 적더라도 채점의 품질은 반드시 유지되어야 한다는 것입니다.

성적: 전체 성적 부여(가상 또는 대면)

수석 채점관 또는 책임 채점관 아래 여러 명의 채점관이 있는 과목의 경우, 본 간행물의 ‘**성적 부여 및 종합**’ 섹션에서 명시된 수행을 따르기 위해 공식적인 회의가 열립니다. IB 과목 담당자가 회의에 참여하여 채점관을 지원하고, 해당 과정이 품질관리 기준에 부합하는지 검증합니다. 이러한 회의는 대면 또는 가상 방식으로 진행될 수 있습니다.

성적: 학생 수가 적은 과목의 ‘가이드형’ 성적 부여(가상)

각 평가 항목별 채점관 수가 매우 제한적인 경우, 선임 채점관들이 온라인으로 논의를 진행합니다. IB 과목 담당자는 성적 부여를 뒷받침할 수 있는 모든 범위의 근거 자료를 제공하며, 품질 보장을 위해 전체 과정을 모니터링합니다.

성적: 학생 수가 적은 과목의 '표준형' 성적 부여

학생 수가 너무 적어 연도별 통계 자료의 의미가 크게 떨어질 경우, 채점관은 판단 기반 성적 급간에 걸쳐 있는 샘플만 검토하는 것이 아니라 제출된 모든 학생 과제를 검토하게 됩니다.

IB 과목 담당자는 요청에 따라 이러한 회의를 지원하며, 회의를 매번 참관하기보다는 무작위 품질 점검을 실시합니다.

학생 수가 적은 모든 과목(지도형 및 표준형 성적 부여)에서 채점관을 지원하기 위해 동일 교과군 내의 다른 과목을 채점관들과의 논의를 장려합니다. 이를 통해 해당 교과군 전체에서 성적의 의미에 대한 공통된 이해가 형성되도록 돕습니다.

최종 승인

시험을 치른 학생의 수와 관계없이 성적 부여 과정의 최종 단계는 수석 채점관 또는 책임 채점관이 IB 선임 평가 교직원에게 권장 사항을 제시하는 것입니다. 이후 선임 평가 교직원은 해당 권장사항을 뒷받침하는 근거가 충분한지 확인하고 이를 최종 승인합니다.

전체 학생 집단의 학생 수와 관계없이 모든 과목은 동일한 방식으로 철저한 검증을 거칩니다.

추가 문헌

DP에 관한 자세한 정보는 다음 자료를 참조하십시오.

- 「디플로마 프로그램: 원리부터 실천까지」
- 「디플로마 프로그램의 평가 절차」
- 「IB 월드 스쿨 규정」
- 교과 가이드(프로그램 자료실에서 각 과목 페이지 확인)
- 교사용 보조 자료(프로그램 자료실에서 각 과목 페이지 확인)
- 「디플로마 프로그램의 평가를 활용한 교수 학습법」

직업연계 프로그램

직업연계 프로그램(CP) 평가는 다음과 같은 고유한 특징을 갖습니다.

- 학생은 IB CP 수료증을 취득하기 위해 일련의 요건을 충족해야 합니다.
- IB CP 수료증에는 종합 점수 제도가 적용되지 않습니다.
- 디플로마 프로그램(DP)과 공유된 과정을 수강하는 CP 학생은 DP 학생들과 함께 평가됩니다.
- CP 프레임워크는 직업교육 과정이 필수적으로 포함되지만, IB에서 이를 제공하거나 수여하는 것은 아닙니다.

직업연계 프로그램의 목적

평가 결과는 과정 및 프로그램의 목적과 명확히 연결될 때에만 타당성을 가집니다.

CP의 고유한 특징은 학생들이 직업을 가질 준비가 된 학습자로 성장할 수 있도록 지원한다는 점입니다. 이 프로그램은 학생들이 어떤 진로를 선택하든 향후 학업이나 취업 전반에 걸쳐 도움이 될 전이 가능한 평생 기능을 개발합니다.

직업연계 프로그램을 통해 학생들은 다음을 할 수 있습니다.

- DP 과정과 직업교육을 통해 다양한 역량을 함양하고, 특정 지식 영역에 대한 이해 심화
- 다양한 맥락에서의 지식 및 기능 습득 또는 향상을 위해 유연한 전략 개발
- 다양한 관점과 기회에 열려 있고 도전 과제에 참여할 줄 아는 평생 학습자로 성장할 수 있도록 태도와 습관 함양
- 변화하는 세계에 효과적으로 참여하기 위한 준비
- 자신의 역량과 의지를 개발하는 학습에 참여하여 건설적인 방식으로 기여

CP 평가 결과의 타당한 사용

평가 모델과 교육과정을 개발할 때 CP 과정의 성과와 IB CP 수료증은 다음을 결정하는 데 활용될 수 있도록 합니다.

- 수습과정과 같은 고용 및 고용 프로그램에서의 선별
- 적절한 직업 교육에서의 향후 교육을 위한 선별
- 대학 입학 선별
- 학생이 대학 프로그램의 요건을 충족했는지 여부(추가 학점 인정 또는 특정 수업/과정 면제 등)

학생이 특정(시험) 언어로 평가를 수행한 경우 해당 언어로 과목 또는 직업 분야의 추가 학습을 수행할 수 있다는 증거가 될 수 있습니다.

이러한 타당한 사용이 왜 중요한 고려 사항인지에 대한 자세한 내용은 본 간행물의 ‘타당성’ 섹션을 참조하십시오.

CP의 구성

CP는 세 부분으로 이루어진 교육 프레임워크입니다. CP는 다음으로 구성됩니다.

- DP에서 SL 또는 HL 수준으로 최소 두 과정
- CP 핵심 과정
- 직업교육

[그림 43] 직업연계 프로그램 설계모형



CP 핵심 과정

CP 핵심 과정의 네 가지 요소는 그 본질, 학습 성과, 의도적 연결성에 따라 서로 밀접하게 연관되어 있습니다. 전체적으로 보면 핵심 과정의 목적은 다음과 같습니다.

- IB 학습자상 및 국제적 소양의 개발을 통해 프로그램을 IB 교육 목표에 확고히 안착

- DP 과정과 직업교육을 맥락화하고 강화하며, 프레임워크의 모든 요소들을 하나로 연결
 - 개인적, 학문적, 전문적인 지식과 기능, 태도의 지속적 개발을 장려
 - 학습이 반복적이고 상호 연결되어 있으며, 개인과 공동체의 지속적인 웰빙을 위한 초석임을 반영
- CP 핵심 과정의 네 가지 구성 요소는 필수로 이수해야 합니다. 네 가지 요소는 다음과 같습니다.

- 공동체 참여
- 언어 및 문화 학습
- 자기개발능력 및 전문기술
- 성찰 프로젝트

성찰 프로젝트는 평가 대상입니다. 즉, 최종 프로젝트와 성찰 결과물은 채점 과정을 거치며, 학생은 이 항목에 대해 IB로부터 최종 성적을 부여받습니다. 다른 핵심 과정의 경우, 학교는 IB에 이러한 요소가 만족할 만한 수준으로 이수 되었음을 입증해야 하지만 IB에서 학교가 수행한 평가를 검증하지는 않습니다. CP 학교에서 적용하는 절차와 관행은 학교 평가방문 기간 동안 검토됩니다.

CP의 일부로서의 DP 과정

각 학생은 DP와 공통인 과정을 최소 2개 이상 이수해야 합니다. CP 학생은 DP 학생과 동일한 평가 과정을 거칩니다. CP 학생만을 위한 별도의 시험이나 성적 부여 절차는 없습니다.

일부 과정을 공유함에도 불구하고, 학생이 DP와 CP를 동시에 등록하는 것은 불가능합니다. 각 프로그램의 폭넓은 요구 사항으로 인해 두 과정을 동시에 이수하는 것은 불가능합니다.

직업교육

IB는 CP의 직업교육 요소에 대해 별도의 평가를 내리거나 어떤 방식의 품질 관리도 적용하지 않습니다. 유일한 요구 사항은 학교가 학생의 직업교육 이수 여부를 확인하는 것입니다. DP 과정은 CP의 직업교육 요소의 일부를 구성하기에 적합하지 않습니다.

CP 결과 산출 방법

IB CP 수료증에는 점수 제도가 적용되지 않습니다.

다음 모든 요구 사항을 충족한 학생에게 IB CP 수료증이 수여됩니다.

- 학교에서 학생이 특정 직업교육을 이수했음을 확인했습니다.
- 학생이 최소 두 개의 DP 과정에서 3점 이상의 성적을 받았습니다.
- 학생이 성찰 프로젝트에서 D 이상의 성적을 받았습니다.
- 학교에서 모든 공동체 참여, 언어 및 문화 학습, 자기개발능력 및 전문기술에 대한 요구 사항이 충족되었음을 확인했습니다.
- 학생이 IB 최종 학위 수여 위원회로부터 학업적 부정행위로 인한 징계를 받지 않았습니다.

CP 성적 및 결과는 DP와 동일하게 IB 최종 학위 수여 위원회에 의해 확정됩니다.

CP 이중 언어 수료증

IB CP 수료증의 대안으로 다음 기준 중 **하나 이상**을 충족한 학생에게 CP 이중 언어 수료증이 부여될 수 있습니다.

- 언어와 문학에서 선택한 두 가지 DP 언어 과정을 모두 3점 이상의 성적으로 이수했습니다.
- 언어와 문학의 DP 언어 과정을 이수하고, 개인과 사회 또는 과학의 DP 과정을 언어와 문학에서 선택한 언어와 다른 시험 언어로 이수했습니다. 학생은 반드시 두 가지 과정 모두에서 3점 이상의 성적을 받아야 합니다.

기준 및 분실된 과제 관리하기

성찰 프로젝트와 DP에 공통으로 적용되는 모든 과정은 분실 답안 채점 절차와 학생 집단의 규모에 따른 기준 관리 방식에 있어 DP와 동일한 접근법을 사용합니다.

분실된 과제 및 분실 답안 채점에 관한 자세한 내용은 본 간행물의 ‘**분실 답안 채점**’ 섹션을 참고하십시오.

추가 문헌

직업연계 프로그램에 관한 자세한 정보는 다음 자료를 참조하십시오.

- “[Career-related Programme: From principles into practice](#)”(영문 제공)
- “[Career-related Programme Assessment procedures](#)”(영문 제공)
- “[Overview of the Career-related Programme](#)”(영문 제공)
- “[Reflective project guide](#)”(영문제공)
- “[Language and cultural studies guide](#)”(영문 제공)
- “[Personal and professional skills guide](#)”(영문 제공)
- “[Community engagement guide](#)”(영문 제공)

중등교육 프로그램

MYP 수료증은 학생이 중등교육 프로그램(MYP)의 모든 측면을 성취했음을 보여주는 증서입니다. MYP 수료증을 취득하려면 학생은 다음 요건을 반드시 충족해야 합니다.

- MYP 컴퓨터 기반 평가에서 1점 또는 2점을 받지 않고 최소 28점의 총점을 획득함
- 권장 기간인 최소 2년 이상(최저 1년) 프로그램에서 공부하고 5년 차의 요구 사항을 충족함
- 필수 과목을 포함하여 서로 다른 교과군에서 최소 5개 학문, 최대 8개 교과군의 8개 학문에서 내부 평가 및 시험을 완료함
- 예술, 체육과 보건 또는 디자인 그룹에서 하나 이상의 학문적 e-포트폴리오를 완료함
- 통합교과 온 스크린 시험을 완료함
- 개인 프로젝트를 완료하고 제출함
- 공동체 참여에 대한 IB의 최소 요구 사항을 충족함

중등교육 프로그램의 목적

평가 결과는 과정 및 프로그램의 목적과 명확히 연결될 때에만 타당성을 가집니다.

MYP는 만 11세~16세 학생들에게 학업적 도전 과제를 제시하고 이를 통해 살아가는 데 필요한 기능을 개발하도록 일관성 있고 종합적인 교육과정 프레임워크로서 설계되었습니다. 만 11세부터 16세는 청소년의 발달에 매우 중요한 시기입니다. 학교에서의 성취는 개인적, 사회적, 정서적 웰빙과 긴밀히 연관됩니다. 학생들이 자신의 정체성을 확립하고 자존감을 형성하는 시기에, MYP는 학생들에게 동기를 부여하여 학교생활은 물론 졸업 이후의 삶에서도 성공을 거둘 수 있도록 지원합니다. MYP 프로그램을 통해 학생들은 자신의 강점을 발전시키고 자신이 뛰어나지 않을 수 있는 교과와 도전 과제를 받아들일 수 있습니다. MYP는 학생들의 잠재력을 일깨우고, 학습 선호를 탐구하며, 적절한 도전을 감수하고, 성찰하며, 자신의 정체성에 대한 강한 인식을 확립할 수 있는 기회를 제공합니다.

(「중등교육 프로그램: 원리부터 실천까지」, 2014, p. 3)

MYP에서는 MYP 교과군 목표와 채점 기준이 서로 명시적으로 일치합니다. 모든 MYP 교과군에는 네 개의 목표에 부합하는 네 개의 평가 기준이 있습니다. 각 기준은 최종 결과에 동등하게 반영됩니다.

컴퓨터 기반 평가에 세계적 맥락 포함하기

MYP의 학습 맥락은 진정성 있는 세계적 환경, 사건, 상황이어야 (또는 이에 대한 본보기여야) 합니다. MYP에서의 학습을 위한 맥락은 국제적 소양과 세계적 참여를 장려하기 위해 프로그램 내 세계적 맥락에서 선택됩니다. MYP는 PYP의 초학문적 주제에서 개발되고 확장된 교수와 학습을 위한 여섯 가지 세계적 맥락을 구분합니다.

(「중등교육 프로그램: 원리부터 실천까지」, 2014, p. 18)

각 시험 기수는 「중등교육 프로그램: 원리부터 실천까지」에서 공개한 목록에서 선택된 특정 세계적 맥락과 탐구를 바탕으로 구성되며 영향을 받습니다.

각 학문적 온 스크린 시험에서 약 1/3의 과제는 선택된 세계적 맥락과 연관되거나, 그로부터 영향을 받거나, 그에 기반하여 파생됩니다. 통합교과 온 스크린 시험 전체는 선택된 세계적 맥락의 영향을 받습니다.

언어습득, 예술, 체육과 보건 과목의 부분적으로 완성된 단위 계획서는 선택된 세계적 맥락을 참조하여 작성됩니다.

MYP 평가 결과의 타당한 사용

평가 모델과 교육과정을 개발할 때 IB는 MYP 과정의 성적과 수료증 점수가 다음과 같은 용도로 활용될 수 있도록 합니다.

- 향후 교육의 기회 또는 직업을 위한 선별
- 교육을 지속하는 학생에게 긍정적인 피드백과 개인의 강점을 보여주는 지표 제공

학생이 특정(시험) 언어로 평가를 수행한 경우, 이는 프랑스어, 스페인어, 영어를 사용하는 학교에서 학습하는 것과 같이 해당 언어로 해당 과목의 추가 학습을 수행할 수 있음을 보여주는 증거가 됩니다.

본 간행물의 ‘타당성’ 섹션에서 왜 이러한 점이 중요한 고려 사항인지 자세히 알아보십시오.

MYP의 구성

[그림 44] 중등교육 프로그램 설계모형



MYP 결과 산출 방법

학생은 MYP 5년 차 학습을 마친 후 IB 외부 평가에 참여할 수 있습니다. 이러한 평가의 결과는 ‘MYP 과정 결과’ 간행물에 기록됩니다. 또한 학생은 MYP 수료증 수여로 이어질 수 있는 평가를 선택하여 응시할 수 있습니다.

학교에서는 MYP 참여 증서를 발급할 수도 있습니다. 이는 최소 2년 이상 프로그램을 학습하고 3년 차 또는 4년 차의 요구 사항을 완료한 MYP 학생을 위한 것입니다. 이러한 학생은 IB의 어떠한 평가에도 등록하지 않습니다. MYP 참여 증서는 학교에서 발급하는 문서이며 IB의 검증을 받지 않습니다.

MYP 수료증을 취득하려면 학생은 반드시 프로그램의 마지막 학년에 참여해야 하며(권장 참여 기간 2년), 다음 요건을 충족해야 합니다.

- 언어와 문학, 언어습득(또는 제2 언어 및 문학), 개인과 사회, 수학, 과학 그리고 예술, 체육과 보건, 디자인 과목 중 한 과목을 포함하여 총 여섯 가지 과목에서 온 스크린 평가를 완수하거나 e-포트폴리오를 작성합니다.
- 위의 항목에 나열된 여섯 가지 과목 각각에서 최소 3점 이상의 성적을 받습니다.
- 교과통합 온 스크린 시험을 수행하고 최소 3점 이상의 성적을 받습니다.
- 개인 프로젝트를 최소 3점 이상의 성적으로 완수합니다.
- 총 28점을 취득합니다.
- 공동체 참여에 대한 학교의 요구 사항을 충족합니다.

IB MYP 이중언어 수료증은 다음 중 하나에 해당하는 온 스크린 시험에서의 성공적인 결과를 추가로 요구합니다.

- 제2 언어 및 문학 과정(언어습득 과정의 대안)
- 학생이 선택한 언어와 문학 과정의 언어가 아닌 다른 언어로 과학, 개인과 사회, 통합교과 시험을 하나(또는 그 이상) 수행

외부 총괄 평가 수행하기(MYP 컴퓨터 기반 평가)

선택 사항인 컴퓨터 기반 평가는 학생들이 알고 있는 것과 할 수 있는 것을 평가하기 위해 다음과 같이 두 가지 방법으로 구성됩니다.

- 언어습득, 예술, 디자인, 체육과 보건 과목에서 평가를 위해 제출되는 e-포트폴리오는 세계적 기준의 일관성을 유지하기 위해 조정을 거칩니다.
- 언어와 문학, 언어습득, 개인과 사회, 과학, 수학, 통합교과 학습의 과정에 대한 2시간 분량의 온스크린 시험입니다.

또한, 개인 프로젝트는 전자 형식으로 IB에 제출되어 조정 과정을 거칩니다 다른 컴퓨터 기반 평가는 학교에 따라 선택적으로 수행되지만 모든 MYP 5년 차 학생은 반드시 개인 프로젝트 컴퓨터 기반 평가에 참여해야 합니다.

시험 계획안

IB는 컴퓨터 기반 평가의 구성과 형식을 학교에 명확히 안내하기 위해 시험 계획안을 공개합니다. 이러한 계획안을 통해 교사와 학생은 MYP 컴퓨터 기반 평가의 특징과 목적을 이해할 수 있습니다. 또한 시험 계획안은 학생이 온 스크린 시험을 준비하고 교과군별 기준과 평가 전략에 집중할 수 있도록 지원합니다. 계획안에는 항상 네 가지 기준이 존재하며, 각 기준은 동일한 비중으로 반영됩니다.

IB는 어떤 시험 기수에서도 시험이 계획안에서 3점 이상 벗어나지 않도록 보장합니다.

e-포트폴리오와 부분적으로 완성된 단원 계획서

e-포트폴리오는 시험 형식으로 평가하기 어려운 확장 과제물(결과물) 또는 공연을 평가할 수 있도록 합니다. e-포트폴리오는 부분적으로 완성된 단원 계획서를 기반으로 하며, 이는 교사가 지역적 맥락에 맞는 유연성을 발휘하는 동시에, 공정하고 의미 있는 판단을 내릴 수 있도록 적절한 학생 증거를 산출하게끔 안내하는 역할을 합니다. 기수마다 부분적으로 완성된 단원 계획서가 새롭게 제공됩니다.

단원 계획서는 교사가 설정한 과제가 학생들이 MYP 성적의 전체 범위에 걸쳐 증거를 제시할 수 있도록 해야 합니다. 교사가 고안한 평가에서는 과제가 매우 쉽거나 어려울 경우, 학생에게 불리하게 작용할 위험이 분명히 존재합니다. 조정 과정에서 IB는 평가를 위해 제출된 과제만을 기반으로 성적을 부여할 수 있습니다. 교사가 고안한 과제가 성적 기준표의 일부 범위만을 다루는 경우, 그 범위 밖의 성적은 부여될 수 없습니다.

단일 평가-학생의 부담 관리하기

시험을 치르고 과제물을 수행하는 과정은 학생에게 스트레스와 부담을 줍니다. 또한 이는 교수에 할애할 시간을 빼앗기도 합니다. MYP 5년 차 학생의 경우 IB는 모든 요소를 고려했을 때 총괄 평가의 양을 최소화하는 것이 가장 적합하다고 판단합니다. 이로 인해 학생이 자신의 성취를 보여줄 기회가 한 번뿐이라는 문제가 생길 수 있으나, IB는 학생의 전반적인 복지를 위해 이러한 문제를 감수합니다.

분실 성적에 대한 접근 방법

학생의 부담을 줄이기 위해 평가 항목을 하나만 두는 구조의 단점은, MYP에서 활용할 수 있는 근거가 극히 제한된다는 점입니다. 이는 분실 성적 절차에 대한 IB의 신뢰도가 다른 프로그램의 분실 답안 채점 절차보다 낮음을 의미하며, 따라서 예외적인 경우에만 적용되어야 합니다.

분실 성적 추정 과정에 관한 자세한 내용은 본 간행물의 ‘[분실 성적 절차](#)’를 참고하십시오.

추가 문헌

중등교육 프로그램에 관한 자세한 정보는 다음 자료를 참조하십시오.

- [「중등교육 프로그램: 원리부터 실천까지」](#)
- [“Guide to the MYP exam session”](#)(영문 제공)
- [MYP 교과 가이드](#)
- [MYP 프로젝트 가이드](#)

- “Middle Years Programme Assessment procedures”(영문 제공)
- “MYP on-screen examinations: IT requirements and school responsibilities”(영문 제공)
- “MYP on-screen familiarization”(영문 제공)

초등교육 프로그램

초등교육 프로그램(PYP) 평가는 다음과 같은 고유한 특징을 갖습니다.

- 평가에는 교사와 학생이 협력하여 학습을 관찰, 기록, 측정, 보고 및 수정하는 과정이 포함됩니다.
- 평가 문화를 조성하는 일은 학습 공동체의 모든 구성원이 평가 역량을 개발하도록 하는 것을 포함합니다.
- IB 외부 총괄 평가 요구 사항은 존재하지 않습니다.

초등교육 프로그램의 목적

평가 결과는 과정 및 프로그램의 목적과 명확히 연결될 때에만 타당성을 가집니다.

PYP 교육과정 프레임워크는 초학문적 학습을 교육과정 구성 원리로 삼아, 학생들이 전통적인 과목 간의, 과목에 걸쳐, 과목을 초월하여 학습을 경험할 수 있도록 합니다.

구성주의 및 사회구성주의 학습 이론에 기반하여, PYP 교육과정 프레임워크는 PYP 학생이 학습의 주체자이며 학습 과정의 동반자라는 전제로 시작됩니다. 이들에게는 자신, 타인, 주변 세상에 대해 탐구하고, 질문하며, 궁금해하고, 이론화할 수 있는 타고난 잠재력을 지니고 있습니다.

학생들이 학습하는 방법에 대한 이러한 이해는 탐구 기반 및 개념 주도의 초학문적 학습 및 교수 모델의 근간이 됩니다. 탐구 프로그램에 참여하고 자신의 학습을 성찰하는 과정을 통해 PYP 학습자들은 지식, 기능 및 태도를 발전시켜 세상과 소통하며 인류와 지구의 웰빙을 위한 실천을 이끌어냅니다.

PYP의 구성

[그림 45] 초등교육 프로그램 설계모형

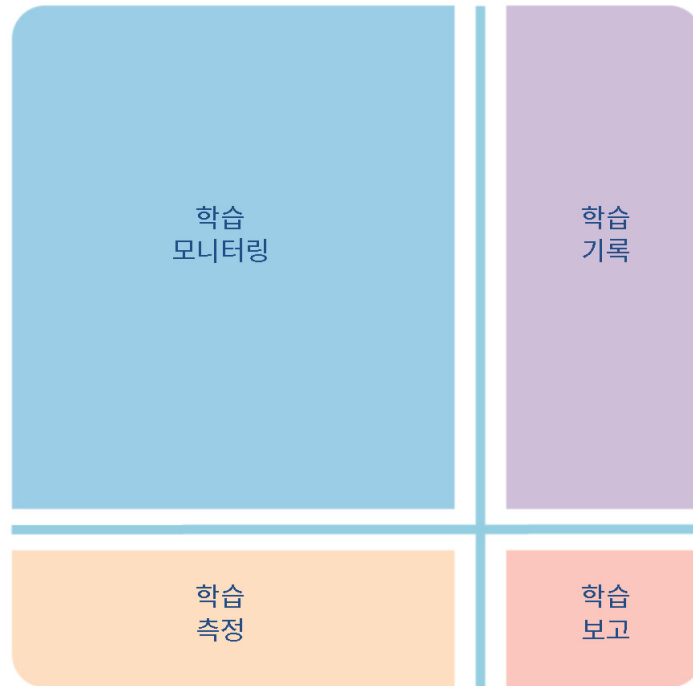


PYP에서의 평가

평가는 숙고에 기반해 효과적으로 학생이 교과별 지식과 기능을 습득하고, 개념을 이해하며, 학습 접근 방법과 IB 학습자상 자질을 개발하는 것을 지원한다는 PYP의 목표에서 매우 중요한 역할을 합니다.

PYP 평가는 학습에 대한 모니터링, 기록, 측정, 보고라는 네 가지 차원으로 구성됩니다. 이 네 가지 차원은 각각 고유한 기능을 지니지만, 모두 교수와 학습에 반영할 수 있는 근거를 제공하는 것을 목적으로 합니다. 평가의 네 가지 차원은 동일한 비중을 갖지는 않지만, 각 차원은 고유한 중요성과 가치를 지닙니다. PYP에서 학습의 모니터링과 기록에 중점을 두는 이유는 학습자에게 실행 가능한 피드백을 제공하는 데 이러한 차원이 중요하기 때문입니다.

[그림 46] PYP에서 평가의 여러 차원



다음은 PYP 평가의 근간을 이루는 가정입니다.

- 평가는 교사와 학생, 그리고 학생 동료 간에 지속적으로 이루어지는 협력적 과정입니다. 이를 통해 학습의 증거를 수집, 분석, 성찰하고, 그 결과를 다시 교수 활동에 반영합니다.
- 학생 개개인이 자신의 경험을 이해하는 방식은 매우 다양하고 복잡하며 정교합니다. 이를 충분히 고려하여 과제 평가를 위한 다각적인 전략을 활용해야 합니다.
- 학생은 학습을 평가하고 성찰하며, 동료와 교사에게 받은 피드백을 바탕으로 행동하고, 그 결과를 다음 학습 단계에 피드포워드하는 과정에 능동적으로 참여합니다.
- 학습 목표와 성취 기준은 공동으로 구성되고 명확하게 공유됩니다.
- 학습 성과와 학습 과정이 모두 평가됩니다.
- 평가 설계는 과거 지향적인 동시에 미래 지향적입니다.
- 평가 문화를 조성하는 일은 학습 공동체의 모든 구성원이 평가 역량을 개발하도록 하는 것을 포함합니다.

추가 문헌

PYP에 관한 자세한 내용은 프로그램 자료실의 「초등교육 프로그램: 원리부터 실천까지」를 참고하십시오.

내부 평가의 조정: 자세히 알아보기

조정은 모든 학교에서 중등교육 프로그램(MYP), 디플로마 프로그램(DP), 직업연계 프로그램(CP)의 내부 평가 과제에 공통 기준이 적용되었는지 확인하기 위한 방법입니다. 조정의 결과에 따라 학교의 점수가 하향되거나 상향되거나 그대로 유지될 수 있습니다. 조정의 목적은 교사가 학생의 과제를 채점할 때 평가 기준을 얼마나 정확하고 일관되게 적용했는지를 확인하는 것입니다.

표본 추출

학교가 내부 평가(IA) 점수를 제출하면 국제 바칼로레아(IB)에서 조정용 샘플에 포함될 학생 과제를 선정합니다. 이러한 샘플은 학교의 점수 범위가 적절히 반영되도록 신중하게 선정됩니다. 조정용 샘플은 해당 과목의 학생 수에 따라 10개, 8개, 5개 또는 5개 미만으로 결정됩니다. 샘플 중 세 명은 초기 표본으로 배정되며, 이는 샘플의 점수 범위를 반영하고 교사가 세계적 기준의 허용 오차 범위 내에서 채점하고 있는지 판단하기 위한 것입니다. 그렇지 않은 경우 나머지 샘플[‘확장 샘플(enhanced sample)’]도 채점관에 의해 채점됩니다. IB는 만점을 받은 학생의 과제를 조정용 샘플로 선정할 경우 상위권 점수가 조정의 결과로 더 높아질 수 있기 때문에 만점을 받은 학생 과제를 조정용 샘플로 선정하지 않는 경향이 있습니다.

특정 과정에 대한 학생 수가 여러 반으로 나뉘는 만큼 많고, 내부 평가를 수행하는 교사가 둘 이상인 경우, IB는 해당 교사들이 내부 평가를 공유하고 함께 협력하여 평가 기준 적용 방식을 서로 표준화하도록 요구합니다. 학교에는 하나의 조정용 샘플이 요청되며, 이 샘플에는 여러 교사가 채점한 과제가 포함될 가능성이 매우 높습니다. 그러나 한 학교 내 동일 과목에서 서로 다른 시험 언어를 사용하는 학급이 있는 경우, 각 언어별로 별도의 조정용 샘플이 필요합니다.

조정 인자 결정하기

모든 내부 평가 항목은 평가 기준표나 점수 구간표를 적용하여 채점되며, 대부분의 경우 교사는 평가 과제에 내포된 맥락과 과정에 대해 채점관보다 훨씬 더 많은 정보를 가지고 있습니다. 이러한 이유로 내부 평가 항목을 조정하는 채점관은 단순히 교사가 부여한 점수를 무시하고 다시 채점하기보다는 교사의 채점이 적절한지 여부를 판단하도록 요청됩니다. 교사의 점수는 조정자가 해당 점수가 부적절하며 세계적 채점 기준에 부합하지 않는다고 확신할 때만 변경되어야 합니다.

교사의 채점 샘플은 채점관에 의해 조정되며, 두 점수 집합 간의 통계적 비교 결과에 따라 필요한 경우, 학교 내 해당 과목의 모든 학생에 대해 교사가 부여한 점수가 조정됩니다.

교사가 지속적으로 점수를 낮게 또는 높게 부여하는 경우 교사가 부여한 모든 점수가 동일하게 조정됩니다.

다음 예시(<표 8>)에서 교사는 학생에게 평균 5점을 적게 부여한 경우, 해당 학생 집단 전체에서 교사가 부여한 모든 점수에 5점이 더해집니다.

<표 8> 일관된 과소 채점

교사의 점수	채점관의 점수	교사의 점수와 채점관 점수의 차이	최종 조정 점수
15	20	5	20
11	17	6	16
10	14	4	15
8	12	4	13
4	10	6	9

교사가 점수 범위의 상위권 또는 하위권에서 과소 또는 과대 채점하는 경우 조정은 점수 범위 전반에 걸쳐 달라질 수 있습니다.

다음 예시(<표 9>)에서 교사는 국제적 기준과 비교했을 때 일관되지 않게 채점했습니다. 점수 범위의 상위권에서는 지나치게 관대했지만 점수 범위의 하위권에서는 지나치게 엄격했기 때문에, 교사 점수에 적용되는 조정은 점수 범위에 따라 달라집니다.

<표 9> 일관되지 않은 채점

교사의 점수	채점관의 점수	교사의 점수와 채점관 점수의 차이	최종 조정 점수
37	35	-2	35
25	22	-3	24
17	18	1	17
12	14	2	13
6	8	2	8

교사가 올바른 기준에 따라 채점하는 경우 조정이 이루어지지 않습니다.

<표 10>에서 교사의 점수는 모두 세계적 기준의 허용 오차 내에 있기 때문에 모든 학생이 교사가 부여한 점수를 받게 됩니다.

<표 10> 올바른 기준에 따른 채점

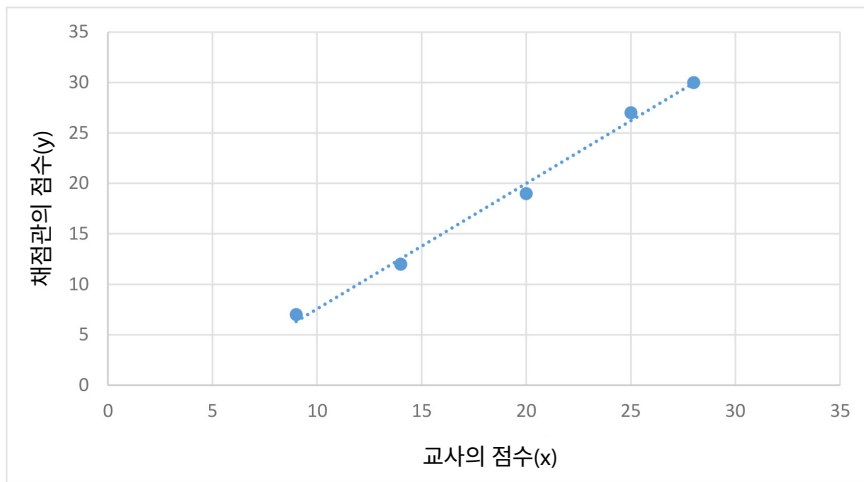
교사의 점수	채점관의 점수	교사의 점수와 채점관 점수의 차이	최종 조정 점수
32	31	-1	32
29	30	1	29
8	8	0	8

선형 회귀

각 조정용 샘플 데이터를 분석하여 교사의 채점이 허용 오차 내에 있는지를 먼저 판단합니다. 허용 오차 범위를 벗어나는 경우, 샘플에서 나타난 일반적인 추세에 기반하여 교사가 부여한 모든 점수에 적절한 조정이 적용됩니다. 이 기법을 '선형 회귀'라 부르며, 이는 교사와 채점관이 부여한 샘플 점수에서 파생된 데이터 점수 모음을 통해 최적선이 계산됩니다. 선형 회귀와 이에 상응하는 점수 모음의 예시는 [그림 47]에 제시되어 있습니다.

[그림 47]의 조정 회귀직선은 점수 범위의 상위권에서는 다소 엄격하게, 하위권에서는 너무 관대하게 채점한 교사의 사례를 보여줍니다. 각 점은 교사와 채점관이 하나의 샘플 과제에 부여한 두 점수를 나타냅니다. 이 연속된 회귀직선을 사용하여 교사의 점수를 조정 점수로 변환합니다.

[그림 47] 선형 회귀와 이에 상응하는 점수 모음



교사의 점수	채점관의 점수	최종 조정 점수
28	30	30
25	27	26
20	19	20
14	12	12
9	7	6

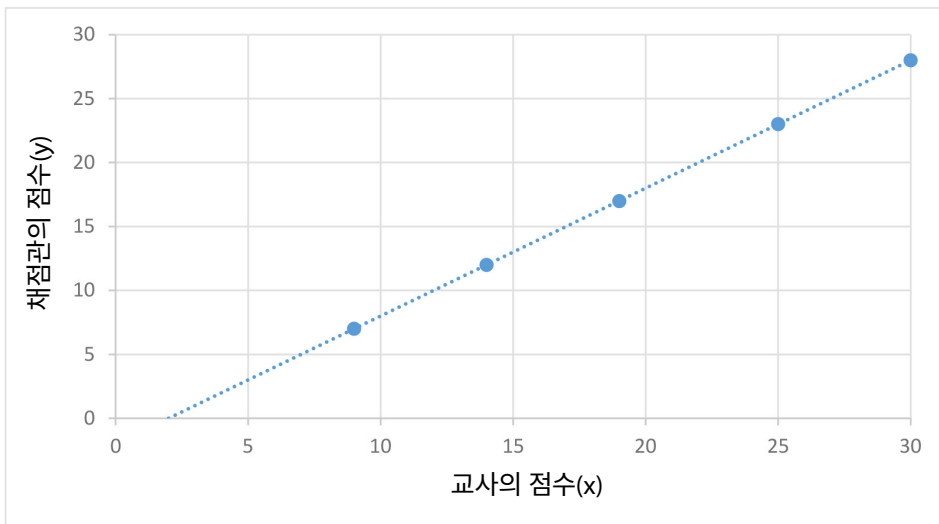
샘플 데이터로부터 계산된 회귀직선 방정식을 사용하면, 교사가 부여한 점수(x)를 채점관이 평균적으로 해당 학생에게 부여했을 가능성이 가장 높은 점수(y)로 변환할 수 있습니다. 이와 같은 조정은 샘플에서 얻은 결과를 더 큰 점수 모음으로 외삽하는 방식이므로, 채점 과정에서 나타나는 일반적인 추세만을 반영할 수 있습니다. 조정의 목적은 전체적으로 학생의 점수가 보다 적절한 수준으로 조정되도록 하는 것입니다. 이는 최적화 모델이므로 모든 학생이 채점관 점수를 그대로 조정 점수로 받는 것은 아닙니다.

조정 실패

선형 회귀 최적선(계산된 조정 인자)을 모든 교사의 점수에 적용하기 전에 이 선이 특정 조건을 충족하는지 확인하기 위해 자동으로 검사가 수행됩니다. 경우에 따라 제출된 샘플 과제만으로는 적절한 조정 차이를 계산하는 것이 불가능할 수도 있습니다. 이때 사용하는 통계적 측정법 중 하나가 상관계수(적률 상관계수)입니다. 이는 교사가 부여한 점수와 채점관이 부여한 점수 간의 관계가 얼마나 일관적인지를 측정합니다.

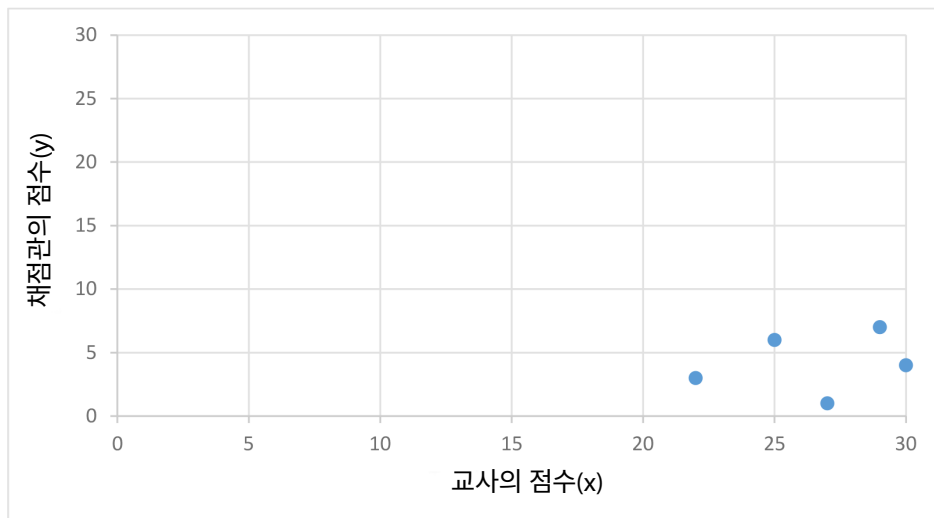
상관계수 1은 채점 간의 관계가 완벽하게 일관되며, 학생들의 순위(최상위에서 최하위까지)를 매기는데 있어 완전한 일치를 보였음을 의미합니다(단, 점수가 반드시 동일하다는 뜻은 아닙니다). 이는 [그림 48]에서 확인할 수 있습니다.

[그림 48] 상관계수: 1



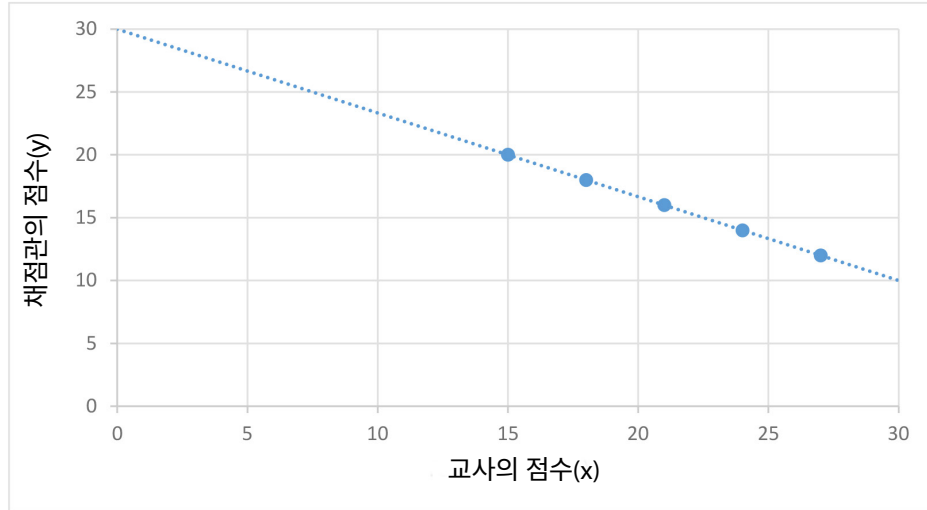
상관계수 0은 두 점수 간에 아무런 관계가 없음을 의미합니다([그림 49] 참조).

[그림 49] 상관계수: 0



상관계수 -1 은 학생 과제의 상대적 수준에 대해 교사와 채점관이 지속적으로 정반대의 견해를 가지고 있음을 나타냅니다. 즉, 교사와 채점관이 매긴 학생들의 순위가 정반대라는 의미입니다([그림 50] 참조).

[그림 50] 상관계수: -1



계산된 조정 인자가 허용 가능하려면 상관계수가 최소 0.85 이상이어야 하며, 이는 교사와 채점관 사이에 높은 수준의 일치가 있었음을 의미합니다. 그러나 높은 상관계수만으로는 조정 인자가 적절하다고 보장하기에 충분하지 않습니다. 회귀직선의 기울기(경사)가 0.5 와 1.5 사이인지 확인하는 추가 검사가 수행됩니다. 기울기가 너무 낮은 경우(또는 너무 완만한 경우) 교사가 일관된 기준으로 채점했다라도 학생들의 점수를 너무 넓게 분산시켰음을 의미합니다. 즉, 성과가 낮은 과제에는 상대적으로 너무 적은 점수를, 성과가 좋은 과제에는 너무 많은 점수를 부여한 것입니다. 이러한 경우 채점관은 교사의 점수 범위를 좁게 줄이게 됩니다. 기울기가 1.5 보다 크면 직선이 지나치게 가파르고 반대의 상황이 적용됩니다. 즉, 교사가 평가를 위해 제출된 과제에서 성과가 낮은 과제와 높은 과제를 충분히 구별하지 않았고 채점관이 부여된 점수 범위를 더 확장해야 함을 의미합니다.

상관계수가 0.85 미만이거나 기울기가 $0.5 \sim 1.5$ 범위를 벗어난 경우, 샘플은 자동 조정 검사에서 실패합니다.

조정 검사에 실패한 모든 학교 샘플은 IB 평가 직원이 개별적으로 검토하며, 기반 데이터를 신중히 고려한 후 다음을 결정할 수 있습니다.

- 계산된 회귀직선이 교사의 특정 점수 범위에 적합한지 판단하기
- 교사의 점수 범위에 적합한 다른 조정 방식을 적용하기
- 추세를 명확히 하기 위해 추가 샘플 데이터 요청하기
- 교사의 채점을 전면적으로 재채점하기 위해 나머지 학생 과제를 요청하기

IB가 모든 조정 실패를 해결할 수 있도록 모든 학생의 내부 평가 과제를 시험 기수 종료일까지 반드시 보관해야 합니다.

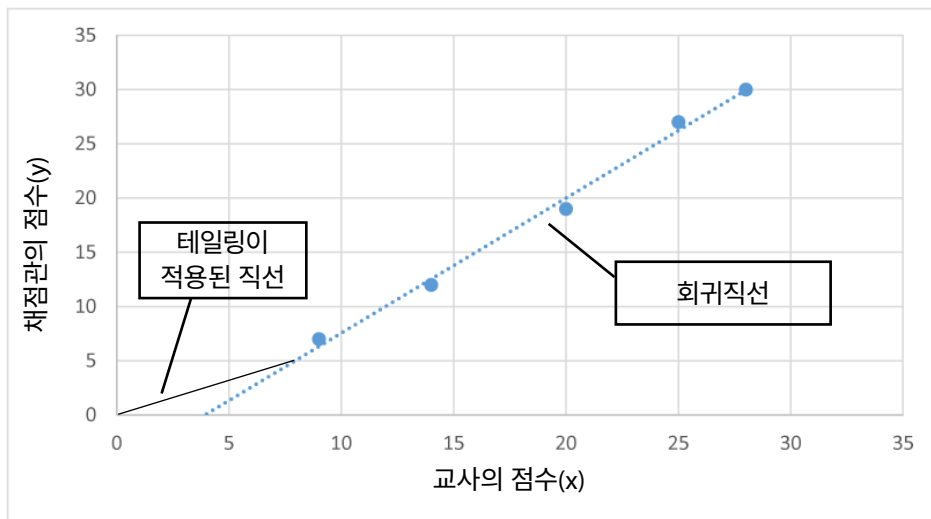
선형 모델의 수정

조정에 사용되는 직선 모델은 ‘테일링(tailing)’을 사용해 어느 정도 수정됩니다. 직선형 조정은 가능한 점수 범위의 양 끝단에서 부적절한 영향을 미쳐, 어떤 학생에게도 0점을 부여할 수 없게 만들 수 있습니다. 실제로 0점에 해당하는 과제에 소수의 점수가 부여되거나, 반대로 과제가 미흡하더라도 몇 점은 받을 자격이 있는 학생이 조정을 통해 0점을 받게 되는 경우가 생길 수 있습니다.

이러한 문제를 해결하기 위해 하위 20%에 속하는 점수에 ‘테일링’을 적용할 수 있습니다. 이 극단적인 구간에서는 계산된 회귀 직선을 수정하여, 기존 회귀직선에서 최소 좌표로 연결되는 새로운 ‘테일링’ 직선으로 대체됩니다. 회귀직선 상위 구간에는 테일링을 적용하지 않는데, 이는 교사가 최고 점수를 받을 자격이 없는 학생에게도 만점을 부여하는 사례가 종종 있기 때문입니다.

[그림 51]의 그래프는 학생의 점수가 최소값으로 조정되는 경우를 방지하기 위해 회귀직선에 테일링을 적용한 사례를 보여줍니다.

[그림 51] 테일링이 적용된 회귀직선



테일링은 조정된 0점이 원래 0점에서만 도출되도록 보장합니다. 이를 통해 약간의 점수라도 받을 자격이 있는 과제가 0점이 되는 것을 방지하고, 동시에 0점을 받아야 하는 과제에 점수가 부여되는 것 또한 방지할 수 있습니다. 이는 교사의 채점이 자동 조정 과정을 거친다는 것을 전제로 합니다. 만약 교사의 채점이 조정에 실패하여 자동으로 조정될 수 없는 경우에는 테일링이 적용되지 않습니다.

참고 문헌

- Ananiadou, K., & Claro, M. (2009). 21st century skills and competences for new millennium learners in OECD countries. *OECD Education Working Papers* (No. 41). OECD.
- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Baird, J. A., Cresswell, M., & Newton, P. (2000). Would the real gold standard please step forward? *Research Papers in Education*, 15(2), 213–229. <https://doi.org/10.1080/026715200402506>
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32(3), 347–364. <https://www.jstor.org/stable/3448076>
- Black, P. (1999). Assessment, learning theories and testing systems. In P. Murphy (Ed.), *Learners, learning & assessment* (pp. 118–134). Paul Chapman Publishing in association with The Open University.
- Bloom, B. S. (Ed.), Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: Handbook 1: Cognitive domain*. Longman.
- Broadfoot, P. M. (1996). *Education, assessment and society: A sociological analysis*. Open University Press.
- Brown, R. (2002). Cultural dimensions of national and international educational assessment. In M. Hayden, J. Thompson, & G. Walker (Eds.), *International education in practice: Dimensions for schools and international schools*. Routledge.
- Chamberlain, S. (2010). *AQA: Public perceptions of reliability*. In Ofqual's *Reliability compendium* (Chapter 18). The Office of Qualifications and Examinations Regulation.
- Coe, R., Searle, J., Barmby, P., Jones, K., & Higgins, S. (2008). *Relative difficulty of examinations in different subjects*. Report for SCORE (Science Community Supporting Education). CEM Centre, Durham University.
- Cresswell, M. J. (1986). Examination grades: How many should there be? *British Educational Research Journal*, 12(1), 37–54.
- Cresswell, M. J. (1996). Defining, setting and maintaining standards in curriculum-embedded examinations: Judgemental and statistical approaches. In H. Goldstein & T. Lewis (Eds.), *Assessment: Problems, developments and statistical issues* (pp. 57–84). Wiley.
- Cresswell, M. J. (2000). The role of public examinations in defining and monitoring standards. *Proceedings of the British Academy*, 102, 69–120.
- Crooks, T. J., Kane, M. T., & Cohen, A. S. (1996). Threats to the valid use of assessments. *Assessment in Education: Principles, Policy & Practice*, 3(3), 265–286. <https://doi.org/10.1080/0969594960030302>
- Dolan, R. P., Burling, K., Harms, M., Strain-Seymour, E., Way, W., & Rose, D. (2013). *A universal design for learning-based framework for designing accessible technology-enhanced assessments* (Research Report). Pearson. <https://eric.ed.gov/?id=ED576691>
- Frith, D. S., & Macintosh, H. G. (1984). *A teacher's guide to assessment*. Stanley Thornes.

- Gibbs, G. (1992). *Teaching more students: Assessing more students*. The Oxford Centre for Staff Learning and Development, Oxford Brookes University.
- Gipps, C., & Murphy, P. (1994). *A fair test? Assessment, achievement and equity*. Open University Press.
- Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist*, 18(8), 519–521. <https://doi.org/10.1037/h0049294>
- Goldstein, H. (1996). Group differences and bias in assessment. In H. Goldstein, & T. Lewis (Eds.), *Assessment: Problems, developments and statistical issues* (pp. 85–93). John Wiley & Sons Ltd.
- Good, F. J., & Cresswell, M. J. (1988). *Grading the GCSE*. Secondary Examinations Council.
- He, Q., Opposs, D., & Boyle, A. (2010). A quantitative investigation into public perceptions of reliability in examination results in England. In Ofqual's *Reliability compendium* (Chapter 19, p. 68). The Office of Qualifications and Examinations Regulation.
- Hieronimus, A. N., & Hoover, H. D. (1986). *Iowa tests of basic skills: Manual for school administrators*. Riverside Publishing Company.
- Hughes, D. C., Keeling, B., & Tuck, B. F. (1983). Effects of achievement expectations and handwriting quality on scoring essays. *Journal of Educational Measurement*, 20(1), 65–70. <https://www.jstor.org/stable/1434941>
- Humphreys, L. G. (1986). An analysis and evaluation of test and item bias in the prediction context. *Journal of Applied Psychology*, 71(2), 327–333. <https://doi.org/10.1037/0021-9010.71.2.327>
- 국제 바칼로레아(2019) 「IB 교육이란 무엇인가?」 국제 바칼로레아 기구.
- 국제 바칼로레아(2024). IB 교육 목표. 국제 바칼로레아 기구.
- Lambert, D., & Lines, D. (2000). *Understanding assessment: Purposes, perceptions, practice*. Routledge Falmer.
- Linn, M. C. (1992). Gender differences in educational achievement. In J. Pfliegerer (Ed.), *Sex equity in educational opportunity, achievement, and testing*. Educational Testing Service.
- Llewellyn, D. (2014). *Inquire within: Implementing inquiry- and argument-based science standards in grades 3–8* (3rd ed.). Corwin.
- Meyer, A., Rose, D. H., & Gordon, D. (2014). *Universal design for learning: Theory and practice*. CAST Professional Publishing.
- Mitra, S. (2011, October 13–16). *Responsibility, leadership and education* [Keynote address]. Heads of Schools Conference, Singapore.
- Murphy, P. (1999). *Learners, learning & assessment*. Paul Chapman Publishing in association with The Open University.
- National Research Council. (2011). *Assessing 21st century skills: Summary of a workshop*. The National Academies Press.
- Newton, P. (2011). A level pass rates and the enduring myth of norm-referencing. *Research Matters*. <https://doi.org/10.17863/CAM.100441>
- Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, 14(2), 149–170. <https://doi.org/10.1080/09695940701478321>
- Newton, P. E. (2012). *We need to talk about validity* [Paper presentation]. The National Council for Measurement in Education Annual Meeting, Vancouver, Canada.

- Newton, P. E., & Shaw, S. D. (2014). *Validity in educational and psychological assessment*. London: SAGE Publications.
- Peterson, A. D. C. (1971). *New techniques for assessment of pupils' work*. Council of Europe.
- Peterson, A. D. C. (2003). *Schools across frontiers: The story of the International Baccalaureate and the United World Colleges* (2nd ed.). Open Court Publishing Company.
- Popham, W. J. (1978). *Criterion-referenced measurement*. Prentice Hall.
- RAND Corporation. (2012, April). *Teaching and learning 21st century skills: Lessons from the learning sciences*. Asia Society. Retrieved January 1, 2017, from http://www.rand.org/pubs/external_publications/EP51105.html
- Rao, K., Currie-Rubin, R., & Logli, C. (2016). *UDL and inclusive practices in IB schools worldwide*. CAST Professional Learning.
- Schleicher, A. (2016). *The case for 21st-century learning*. OECD. Retrieved April 4, 2025, from <https://web-archiv.e.oecd.org/2012-06-14/61660-the-case-for-21st-century-learning.htm>
- Shepard, L. A. (1992). Commentary: What policy makers who mandate tests should know about the new psychology of intellectual ability and learning. In B. R. Gifford, & M. C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement and instruction* (pp. 301–328). Kluwer Academic Publishers.
- Sireci, S. G. (2007). On validity theory and test validation. *Educational Researcher*, 36(8), 477–481. <https://www.jstor.org/stable/4621103>
- Smith, D. J., & Tomlinson, S. (1989). *The school effect: A study of multi-racial comprehensives*. Policy Studies Institute.
- Snyder, B. R. (1971). *The hidden curriculum*. MIT Press.
- Surgenor, P. (2010). *Teaching toolkit: Effect of assessment on learning*. UCD Dublin.
- Vygotsky, L. S. (1962). *Thought and language*. MIT Press.
- William, D. (1993). Validity, dependability and reliability in National Curriculum assessment. *The Curriculum Journal*, 4(3), 335–350. <https://doi.org/10.1080/0958517930040303>
- Winkley, J., & Cresswell, M. (2011). Introduction to the concept of reliability. In Ofqual's *Reliability compendium* (Chapter 1). The Office of Qualifications and Examinations Regulation.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- Wood, D. (1998). *How children think and learn: The social contexts of cognitive development* (2nd ed.). Blackwell Publishing.
- Zanga, G., & De Gioannis, E. (2023). Discrimination in grading: A scoping review of studies on teachers' discrimination in school. *Studies in Educational Evaluation*, 78. <https://doi.org/10.1016/j.stueduc.2023.101284>

용어 해설

IB는 쉽게 이해할 수 있는 언어를 사용하고 전문 용어 사용을 최소화하기 위해 노력합니다. 이 노력의 일환으로 이 용어 해설은 본 간행물과 기타 평가 관련 자료에 등장하는 주요 단어와 개념들을 정의합니다. 이러한 용어 중 일부는 IB의 문서에도 사용될 수 있습니다.

용어	정의
학문적 진실성	학습, 교수, 평가에서 개인의 진실성과 바람직한 관행을 장려하는 일련의 가치와 기능을 의미합니다.
학업적 부정행위	하나 이상의 평가 항목에서 해당 학생 또는 다른 학생이 불공정한 이익을 얻는 결과를 초래하거나 초래할 수 있는(의도적이거나 의도치 않은) 행위입니다. 다른 학생에 불이익을 줄 수 있는 행위 또한 학업적 부정행위로 간주됩니다.
성취 단계	평가를 위해 제출된 과제가 상응하는 기술어를 반영했을 때 주어지는 단계입니다. 성취 단계는 평가 기준표의 왼쪽 열에 표시됩니다.
성취 단계 기술어	각 성취 단계에서 요구되는 평가 제출 과제의 특징을 기술한 설명입니다.
관리자 콘솔	MYP 온 스크린 시험에 사용되는 보안 웹사이트입니다. 이를 통해 MYP 코디네이터 및/또는 지정된 수석 감독관이 기기 등록 및 시험 패키지 접근과 같은 주요 관리 업무를 수행할 수 있습니다. 관리자 콘솔은 IBIS(IB 정보 시스템)의 링크를 통해 접속합니다. 참고: DP/CP 디지털 시험의 경우 디지털 시험 시스템 을 참고하십시오.
불리한 상황	학생이 통제할 수 없는 자신의 평가 성과에 악영향을 미칠 수 있는 상황입니다. 여기에는 심각한 스트레스, 이례적으로 어려운 가정 상황, 가족 및 친지의 사망, 또는 학생의 건강이나 안전을 위협할 수 있는 사건 등이 포함됩니다. 이와 같은 상황이 학교 내 특정 학생 집단 또는 모든 학생에게 영향을 미칠 수 있습니다. 다음은 불리한 상황에 해당하지 않습니다. <ul style="list-style-type: none"> • 학생이 등록된 학교의 과실로 인해 발생한 상황 • 승인된 포용적인 접근성 조치가 제공되었음에도 학생이 성과를 내지 못한 경우
종합	점수와 성적을 합산하여 최종 결과를 산출하는 과정입니다.
정렬	학습에 관한 공통된 가치와 지향점(문서화된 교육과정), 교사의 지도 방식(지도된 교육과정) 그리고 학생이 배우는 내용(평가된 교육과정)에 대해 원칙과 실천 방안을 일치시키는 것입니다.
분석적 채점 기준표	올바른 답안과 점수가 부여되어야 하는 지점을 안내하는 채점 기준표입니다.
평가	학생이 자신의 능력을 증명하기 위해 수행하는 모든 과제를 의미합니다. 이러한 과제는 학교에서 자체적으로 개발하여 시행하거나 IB에 제출하는 형태일 수 있으며, 학습과 교수에 관한 판단의 근거가 되는 증거 모음입니다.
평가 항목	시험, 과제 포트폴리오, 프로젝트 또는 연구 과제와 같이 전체 평가의 일부를 구성하기 위해 함께 수집되는 하나 이상의 과제를 의미합니다.

용어	정의
평가 내용 개발자	평가에 사용될 시험 문제와 관련 채점 기준표를 작성 및 검토하는 전문가입니다. 평가 내용 개발자는 거의 대부분 선임 채점관이 맞지만, 이 역할에는 별도의 전문성이 요구됩니다.
평가 기준	학생의 성과를 평가할 때 척도가 되는 기준입니다.
평가 주기	시험을 포함한 평가를 제작, 수행, 채점하는 일련의 단계입니다. IB가 각 시험 기수의 경험을 토대로 향후 기수를 개선해 나가는 순환 구조이므로 ‘주기’라고 부릅니다.
평가 답안	학생이 평가 과제에 대한 답안으로 생성한 모든 자료를 지칭하는 용어입니다.
평가 전략	학생의 학습에 관한 정보를 수집할 때 사용되는 방법이나 접근법입니다(예: 관찰, 개방형 과제, 선택된 답안).
평가 과제	평가가 이루어질 수 있도록 학생이 참여하는 단일 활동 또는 일련의 활동입니다.
평가 도구	학생의 성과 및 이해에 대한 정보를 수집하는 방법입니다.
성취 기준 평가(약한 절대 평가)	사전에 정의된 성취 수준 설명(기준)과 이전 전체 학생 집단의 성과에 비추어 학생의 성취를 비교하는 방식입니다. IB는 기준을 유지하기 위해 이 접근 방법을 활용합니다.
이례적인 답안	일반적으로 제출되는 답안과 상당히 다른 답안입니다. 이례적인 답안의 예시에는 미완성 과제, 규정을 준수하지 않은 과제, 예상치 못한 답안, 문제가 있는 과제 또는 부정행위가 포함됩니다.
진위성 확인	제출된 과제가 학생 본인에 의해 수행되었음을 입증하는 증거 및 검증 절차입니다. 예시로는 학생의 과제가 본인이 수행한 것임을 확인하는 교사와 학생의 서명이 있습니다.
자동 채점	객관적으로 정답 또는 오답이 명확한 사전 정의된 채점 기준표에 따라 테크놀로지를 활용하여 학생의 과제를 평가하는 과정입니다.
역류 효과	과정의 후반부가 해당 과정의 초반부에 미치는 영향입니다. 교육적 맥락에서 ‘역류’란 일반적으로 평가가 학습과 교수에 영향을 미치는 방식을 의미합니다.
대역폭 검사 도구	DP/CP 학교 행정 직원이 사용하는 온라인 도구로, 다수의 학생이 동시에 DP/CP 디지털 시험을 치를 수 있을 만큼 학교 인터넷 대역폭이 충분한지 확인합니다. 최적의 설정을 위해 시험 기수 전에 대역폭 검사 도구를 미리 실행할 것을 권장합니다.
편향	특정 집단이 평가 대상인 기능이나 지식의 능력 이외의 다른 요인으로 인해, 특정 질문이나 과제에서 평균과 다른 성과를 내는 현상을 의미합니다.
개인 기기 지침	학생이 학교 과제 수행을 위해 개인 기기를 사용하는 관행입니다. 학교는 학생이 학교에서 사용할 적절한 기기를 구매할 수 있도록 학부모 및 법적 보호자에게 기기 사양 지침을 제공합니다.
수석 채점관	한 교과군 내에서 학문 간 기준이 시간이 지나도 유지되도록 보장하는 책임을 맡은 가장 선임인 채점관을 의미합니다.

용어	정의
수석 감독관	<p>DP/CP: 수석 감독관은 디지털 시험 시스템에서의 행정 업무를 수행하고 시험 기간과 그 이전에 모든 기술적 이슈를 해결하는 등 학교에서 디지털 시험을 감독해야 할 책임이 있습니다. 수석 감독관은 학교 교직원이어야 하며, 프로그램 코디네이터가 My IB에서 이 역할을 배정합니다.</p> <p>이 역할은 프로그램 코디네이터와 구별되어야 하지만 시험이 원활하게 진행되도록 코디네이터와 긴밀히 협력해야 합니다. 경우에 따라 프로그램 코디네이터가 수석 감독관 역할을 겸할 수도 있습니다.</p> <p>MYP: 수석 감독관의 역할은 선택 사항입니다. 프로그램 코디네이터는 온 스크린 시험과 관련된 행정 업무를 지원할 사람을 지정할 수 있습니다.</p>
지시어	평가 대상이 되는 평가 목표를 설명하는 문제 속 단어입니다.
비교 가능성	특정한 결과를 다른 결과와 같다고 간주할 수 있는 정도입니다. 이 용어는 일반적으로 학년 간, 또는 과목 간에 사용됩니다.
호환성 검사 도구	MYP 온 스크린 및 DP/CP 디지털 시험에서 기기가 디지털 시험을 실행하기 위한 기술적 요건을 충족하는지 확인하기 위해 사용되는 독립형 애플리케이션입니다. 평가에 사용될 모든 기기에서 반드시 검사 도구를 실행해야 합니다. 검사를 통과하지 못한 기기에서는 시험을 치를 수 없습니다.
구성개념 적합성	평가가 의도된 기능과 지식을 실제로 평가하는 정도를 의미합니다.
과정	정해진 학습 기간 동안 규정된 수업 횟수, 수업 또는 교수 시간을 의미합니다. 학교는 학문적, 통합교과 과정을 통해 과목의 학습 및 교수를 구성합니다.
과정 결과	<p>IB의 주요 평가 결과 문서입니다. 학생이 이수한 각 과목과 성취한 성적(1~7점 또는 해당하는 경우 A~E)이 표시됩니다. 또한 핵심 과정에서 성취한 성적 등 관련 평가 결과도 함께 표시됩니다.</p> <p>마지막으로 이 문서에는 학생의 이름, 학생 코드, 시험 기수 번호, 수여가 이루어진 기수, 발행일, 학생이 등록된 학교의 이름(및 해당하는 경우 대체 학교의 이름)이 기록됩니다.</p>
절대 평가(성취 기준 평가)	성적 부여를 위해 사전에 정의된 성취도에 대한 설명(기준)과 학생의 성취도를 비교하는 것입니다.
준거 기반 평가	사전에 합의된 기준과 비교하여 성취 단계를 결정하는 평가 과정입니다. 따라서 기준은 전체 학생 집단의 성취도에 따라 변동되지 않고 일정하게 유지됩니다.
확정적 점수	<p>책임 채점관이 평가를 위해 제출된 특정 과제에 대해 부여한 점수입니다. 이는 다른 모든 채점관이 채점 시 기준으로 삼아야 하는 점수입니다. (품질 모델을 참고하십시오.)</p>
기기	<p>IB 디지털 평가에 사용되는 노트북 또는 데스크톱 컴퓨터입니다.</p> <p>IB는 향후 태블릿 및 기타 플랫폼과의 호환성을 계속해서 탐구하고 있으며, 이를 반영하기 위해 ‘컴퓨터’ 대신 ‘기기’라는 용어를 사용합니다.</p>
(평가적 의미의) 차별화	서로 다른 수준의 역량을 보이는 학생들을 구별해내는 것을 의미합니다.
(학습 및 교수의) 차별화	다양한 학습자의 요구를 충족하기 위해 학습 내용, 과정, 결과물을 다양화하여 교수 전략을 조정하는 것을 의미합니다.

용어	정의
디지털 시험 답안	IB 디지털 평가에서 학생이 완료한 과제는 프로그램에 따라 다르게 지칭됩니다. MYP에서는 이를 ‘학생 답안 파일’, DP/CP에서는 이를 ‘학생 답안’이라고 합니다. 두 용어 모두 시험 중 생성되어 평가를 위해 제출된 최종 디지털 산출물을 의미합니다. 서면 기반 시험에서는 이를 ‘답안지’라고 합니다.
디지털 시험 시스템	학교가 DP/CP 디지털 시험을 시행하기 위해 사용하는 온라인 시스템입니다.
디지털 시험	MYP는 10년 이상 온 스크린 시험을 시행해 왔습니다. 이는 디지털 환경에 맞게 특별히 설계되었으며, 보안이 유지되는 시험 환경에서 수행되는 상호작용적이고 다양한 매체를 활용한 과제를 포함합니다. DP/CP에서 디지털 시험은 보안이 유지되는 시험 환경에서 학생이 기기를 사용하여 마지막 학년에 수행하는 공식 평가입니다. 이는 단계적으로 도입되어 기존의 서면 기반 시험을 디지털 방식으로 재현하게 됩니다.
학문	학습의 한 분야 또는 학술적 연구 영역을 의미하며, 교수 목적에 따라 지식을 분류하는 방식입니다.(실질적인 평가 목적으로 MYP와 DP에서는 일반적으로 ‘과목’으로 통칭됩니다).
역동적 표본 추출	품질 평가를 더욱 효율적으로 활용할 수 있도록 조정 과정을 개선한 방식입니다.
컴퓨터 기반 평가	MYP 평가는 온 스크린 평가, e-포트폴리오, 개인 프로젝트의 세 가지 항목으로 구성됩니다.
e-채점	채점관이 기기 화면에서 직접 시험 자료를 채점하는 과정을 의미합니다.
성적에 대한 문의	학교의 요청에 따라 점수를 재검토하는 절차입니다.
e-포트폴리오	학교가 MYP 학생의 내부 평가 시험 또는 과제를 업로드하여 IB의 외부 조정을 받도록 하는 시스템 또는 과정을 의미합니다.
시험	시험은 평가의 한 형태로, 다양한 유형의 하나 이상의 과제(단답형, 확장형, 문제 해결 또는 분석형 질문, 때로는 실습 또는 구술 과제)를 포함하며, 학생들은 엄격히 통제된 조건과 정해진 시간 내에 답해야 하는 평가 방식입니다.
시험 감독관	시험 환경을 감독하고 시험 보안 유지를 돕는 사람입니다.
시험지	시험에서 학생이 수행해야 하는 일련의 과제와 질문을 의미합니다. 이는 컴퓨터 기기 화면으로 보는 시험이나 펜과 종이로 수행하는 시험을 지칭할 수 있습니다.
시험 기수	시험이 실시되고 채점되는 특정 기간입니다. IB는 매년 5월과 11월, 두 차례의 시험 기수가 있습니다.
채점관	학생의 외부 평가에 점수를 부여하는 사람입니다.
채점관 재채점	채점관의 채점이 일관되지 않거나 요구되는 기준에서 크게 벗어난 경우, 해당 채점관이 배정받은 학생 답안(평가 구조에 따라 전체 답안 또는 문제 문항 그룹)을 재채점하는 과정입니다. 이는 보통 조정이 실패했을 때 발생합니다.
예외적인 상황	평가 접근성 관련 지원이 필요한 다른 학생들이 일반적으로 겪지 않는 상황을 의미합니다. IB는 어떤 상황이 ‘예외적’으로 간주되는지 결정할 권리가 있으며, 이에 따라 특정한 포용적인 접근성 조치를 결정합니다.
외부 평가	학생의 교사가 아닌 IB에서 출제하고 채점하는 평가입니다.
외부 조정	조정을 참고하십시오.

용어	정의
외부적으로 평가된	과제의 평가가 전적으로 IB에 의해 이루어지는 것을 의미합니다.
숙지	학생들이 콘텐츠가 포함되지 않은 연습용 버전의 시스템을 사용하여 IB 디지털 시험 환경과 도구를 탐색하는 법을 익히는 과정입니다. 학교는 학생들이 디지털 시험 전에 숙지 도구를 주기적으로 사용할 수 있도록 보장할 책임이 있습니다. 숙지 환경은 프로그램 자료실에서 MYP 및 DP/CP용으로 모두 제공됩니다.
숙지 도구	콘텐츠가 없는 디지털 시험 환경 시뮬레이션입니다. 이를 통해 학생들은 과목별 내용 없이 도구를 탐색하고 시험 레이아웃을 이해하며 플랫폼을 탐색하는 법을 연습할 수 있습니다.
최종 평가	학습의 마무리 단계에서 수행하는 학생 과제에 대한 총괄 평가입니다.
형성 평가	교수를 지도하고 학생 성과를 높이기 위해 정보를 제공하는 것을 목표로 하는 지속적인 평가입니다.
성적	학생의 성취도에 대한 설명입니다. 평가를 위해 제출된 과제는 1점(최하점)부터 7점(최고점)까지의 최종 성적이 부여됩니다. 성적은 학생이 보여준 전반적인 역량에 대한 IB의 판단을 나타내며 연차 및 과목 간에 일관성을 유지합니다. 성적은 학생의 성과 수준을 나타내며 모든 시험, 연도, 과목에서 동일한 의미를 지녀야 합니다. 이는 기수 간 항목들의 상대적 난이도와 기준의 변경 사항을 모두 고려한 것입니다.
성적 부여	점수를 성적으로 변환하는 방식을 결정하는 과정입니다. 이는 어느 시험 기수에서 시험을 보든 성적이 동일하다는 것을 의미합니다.
성적 급간	학생의 성취도가 한 성적에서 다른 성적으로 바뀌는 지점입니다. 이는 보통 특정 성적에 상응하는 평가 기준 총점의 최저점 또는 최고점을 나타내는 데 사용됩니다.
성적 기준표	각 성적에서 학생이 달성할 것으로 기대되는 역량에 대한 설명입니다. 성적 기준표는 특정 과목 또는 교과군에 해당하거나 전체 프로그램 전반에서 일반적으로 적용될 수 있습니다. 각각의 경우에서 성적 기준표는 같은 특징을 설명해야 합니다. 보다 구체적인 예시는 해당 설명이 특정 과목의 맥락에서 무엇을 의미인지를 보여주기 위함입니다.
전체론적 기준	과제의 개별 요소(예: 의사소통, 교과 지식, 주장의 설득력)를 따로 평가하는 것이 아니라 하나의 결과물로 간주하여 평가하는 접근 방법입니다.
IB 정보 시스템(IBIS)	프로그램 코디네이터가 행정 절차를 완수하고, 인증과정이 필요한 웹 서버를 통해 IB의 최신 소식과 정보를 제공하는 시스템입니다.
포용적인 접근성	모든 학생의 요구를 고려하여 학생이 교과에서 자신의 역량을 공정하게 발휘할 수 있도록 설계된 평가입니다.
포용적인 접근성 조치	평가 접근성 관련 지원이 필요한 학생을 위해 평가 과정에서 변경되거나 추가되는 조건을 의미합니다. 이를 통해 학생은 자신의 성취 수준을 보다 공정하게 드러낼 수 있습니다.
교과 통합 평가	두 가지 이상의 학습 또는 학업 분야를 하나의 평가에 통합하거나 연관 짓는 것을 의미합니다. DP에서 통합 교과 과목이란 하나의 과목으로 두 가지 교과군의 요구 사항을 충족하는 것을 의미합니다. MYP에서 통합교과적 연구는 교과군 내, 교과군 간 모두에서 개발될 수 있습니다. MYP 외부 교과 통합 평가는 항상 여러 교과군을 포함합니다.
내부 평가(IA)	학교에서 교사가 수행하는 평가입니다.

용어	정의
내부 표준화	학교 내 특정 과목의 모든 교사가 동일한 기준으로 평가하고 있음을 보장하는 과정입니다.
내부적으로 평가된	학생의 교사에 의해 평가되는 과제입니다. IB가 조정을 위해 내부적으로 평가된 자료 중 샘플을 선정합니다.
결과 발표	학생이 수행한 평가 결과를 바탕으로 IB로부터 성적을 받는 과정입니다.
항목	평가를 적절히 세분화했을 때의 가장 작은 단위를 의미합니다. 이는 ‘교수요목 개요’ 또는 교과 가이드의 ‘외부 평가 세부 사항’에 개요가 명시되고 전달되며 기대되는 하나의 독립적이면서도 완전한 평가 요소입니다.
판단	개별 평가 기준에 비추어 학생의 과제를 고려하는 과정입니다.
잠금 모드	디지털 시험 중 학생이 다른 애플리케이션, 웹사이트 또는 파일에 접근할 수 없도록 보안이 유지되는 시험 환경을 의미합니다. MYP: 잠금 기능은 시험 패키지에 내장되어 평가 시작 시 자동으로 활성화됩니다. DP/CP: 잠금 모드는 Windows 및 Mac용 안전한 시험 브라우저(Safe Exam Browser) 및 학교 관리형 Chromebooks을 위한 Chrome Kiosk 모드를 통해 시행됩니다. 학교는 학생 기기를 설정할 책임이 있습니다.
행정 실책	IB 월드 스쿨이 IB 규칙과 규정을 위반하여 IB 시험과 평가의 공정성을 해칠 가능성이 있는 행위를 의미합니다. 학교의 행정 실책은 평가나 시험이 완료되기 이전이나 시험 도중, 혹은 그 이후에 발생할 수 있습니다.
부정행위	학문적 진실성 원칙을 준수하지 않는 모든 행위(예: 표절)를 의미합니다.
관리 가능성	평가 및 개별 과제가 학생이나 학교에 주는 부담의 정도를 의미합니다. 관리 가능성의 예시에는 평가의 길이, 평가를 수행하기 위해 필요한 장비 또는 자료, 자격 취득에 요구되는 평가의 수 등이 포함됩니다.
점수	특정한 과제에 대한 학생 답안의 수준을 수치로 나타낸 값입니다. 학생이 문제에 얼마나 올바르게 답했는지를 반영하기 위해 기준, 점수 구간표 또는 채점 기준표에 따라 점수가 부여됩니다. 점수 배정 방식은 각 문제와 시험마다 다릅니다.
점수 구간표	특정 수준이나 특징을 보이는 학생의 답안에 부여해야 하는 구체적인 점수 범위입니다.
채점 기준표	주어진 과제에 대해 평가 기준을 적용하고 점수를 부여하는 데 사용되는 지침입니다.
분실 성적 절차	MYP 학생에게 성적을 부여하기 위한 메커니즘으로, 학생이 완료한 과제를 바탕으로 IB가 정확하거나 공정한 성적을 산출할 수 없을 때 사용합니다. 증거가 부족한 이유가 IB 또는 제3자(학교 제외)에 의해 초래된 경우, 그리고 학생에게 다른 시기에 평가를 완료하도록 요구하는 것이 합리적이지 않은 경우 이러한 절차를 사용하는 것이 적절합니다.
분실 답안 채점 절차	DP/CP 학생에게 점수를 부여하기 위한 메커니즘으로, 학생이 완료한 과제를 바탕으로 IB가 정확하거나 공정한 점수를 산출할 수 없을 때 사용합니다. 증거가 부족한 이유가 IB 또는 제3자(학교 제외)에 의해 초래된 경우, 그리고 학생에게 다른 시기에 평가를 완료하도록 요구하는 것이 합리적이지 않은 경우 이러한 절차를 사용하는 것이 적절합니다.

용어	정의
조정	공통된 평가 기준이 달성되도록 하는 과정입니다. 이 과정에는 평가를 위해 제출된 샘플 과제를 검토하고 필요한 경우 평가자의 점수를 조정하는 것이 포함됩니다.
조정 인자	확정적 기준에 맞추기 위해 교사의 점수(기준 단계에서)에 적용되는 산술적 조정입니다.
조정용 샘플	채점이 요구되는 기준에 부합하는지 확인하기 위해 IB에 제출되는 샘플 과제입니다.
특수 변형 시험지	특정한 요구가 있는 학생이 특정한 요구가 없는 학생과 동등한 조건에서 평가를 수행할 수 있도록 변경 사항이 적용된 평가를 의미합니다(예: 글꼴 모양이나 스타일 변경). 이러한 수정은 제시되는 질문의 본질을 바꾸어서는 안 됩니다.
선다형 문제	학생이 제시된 선택지 목록에서 정답을 선택해야 하는 문제를 의미합니다.
MYP 온 스크린 시험	MYP를 위해 특별히 설계된 공식적이고, 시간 제한이 있으며, 미디어를 다채롭게 활용하는 디지털 평가입니다. 이 시험은 보안이 유지되는 디지털 환경에서 수행되며 학생이 기기를 사용하여 답하는 과목별 과제로 구성됩니다. 시험을 시행하는 데 인터넷 연결이 필수는 아니지만, 온라인 방식으로 시행하는 것이 강력히 권장됩니다. 이렇게 하면 학생 답안이 자동으로 저장되고 행정적 업무 부담이 줄어듭니다. 이러한 이유로 대부분의 학교는 온라인 방식으로 평가를 수행하고 있습니다.
상대 평가	평가가 설계된 전체 학생 집단의 성과와 학생의 성과를 비교하여 성취도를 결정하는 평가 방식입니다.
목표	평가될 기능, 지식과 이해를 설명하는 일련의 진술문 중 하나입니다.
시험지	시험지 를 참고하십시오.
시범 과목	평가가 진행 중인 과목입니다. 평가가 성공적으로 완료되면, 해당 시범 과목은 정식으로 도입됩니다.
표절	고의성의 여부와 관계없이 다른 사람의 아이디어, 말 또는 결과물을 적절하고 명확한 출처 표기 없이 사용하는 것을 의미합니다.
연습 답안지	표준화 과정에서 식별되고 채점된 평가를 위해 제출된 과제 예시로, 채점관이 따라야 할 채점 기준을 보여주기 위해 제공됩니다.
예측 가능성	무엇이 발생할지 또는 언제 발생할지를 가능할 수 있는 정도를 의미합니다. 평가의 맥락에서 ‘예측 가능성’이란 학교가 시험에 출제될 문제와 그 시점을 미리 예측할 수 있는 능력을 의미합니다.
책임 채점관	평가 항목의 평가를 주도하는 책임자를 말합니다. 책임 채점관은 평가의 기준을 설정하며, 일반적으로 출제자 중 한 명이기도 합니다. MYP에서 책임 채점관의 역할은 다른 시험 체계와는 조금 다릅니다. MYP 책임 채점관은 특정 학문의 최고 선임으로서 평가 설계 팀을 이끌고 기준을 수립 및 유지하며, 채점관 팀 리더를 멘토링할 책임이 있습니다.
채점 자격 취득용 답안지	책임 채점관이 선정한 평가를 위해 제출된 과제 예시로, 채점관이 실제 학생 답안지를 채점하기 전에 요구되는 채점 기준을 이해했는지 확인하는 데 사용됩니다.

용어	정의
품질 모델	학생이 올바른 평가 결과를 받을 수 있도록 보장하기 위해 IB가 취하는 접근 방법입니다. 책임 채점관이 문제별 답안 기준을 세우면, 모든 채점관은 이를 그대로 재현해야 합니다. 외부 채점 평가의 경우, 표준화를 통해 채점관에게 지침을 제공하고, 채점 자격 취득용 답안지로 기준을 이해했는지 확인한 다음 시드 답안지를 통해 채점을 주기적으로 모니터링함으로써 이를 수행합니다.
문제(문항)	학생이 특정 과목에서 자신의 역량을 보일 수 있도록 하는 과제 또는 활동입니다.
문제 은행	관련 주제에 대한 정보와 예상 난이도를 포함한 문제 모음입니다. 문제 은행의 정보는 시험지 제작에 활용될 수 있습니다. IB는 현재 문제 은행을 사용하지 않습니다.
문제 문항 그룹	시험지 내에서 하나의 그룹으로 취급되는 하나 이상의 연관된 문제를 말합니다. 채점관은 전체 답안지가 아닌 개별 문항 그룹 단위로 채점하며, 이러한 접근 방법은 전체 답안지를 채점하는 방식보다 더욱 신뢰할 수 있는 채점을 가능하게 합니다.
신뢰성	학생의 과제가 평가될 때마다 동일한 결과를 얻게 되는 정도입니다. 이는 채점관 간의 신뢰성 또는 단일 채점관의 신뢰성을 의미할 수 있습니다.
시험 언어	학생이 평가에 답할 때 사용하는 언어입니다.
재응시	MYP 수료증 또는 IB 디플로마를 취득하거나, 이미 취득한 이수증의 총점을 높이기 위해 하나 이상의 시험에 두 차례 이상 응시하는 것을 의미합니다.
답안지	서면 기반 시험에서 학생이 작성한 답안입니다. 평가를 위해 제출된 모든 과제를 의미할 수도 있습니다.
시드 답안지	책임 채점관이 채점한 답안지로, 채점관에게 배정된 일련의 답안지들 사이에 무작위로 추가된 답안지를 의미합니다. 채점관이 시드 답안지에 부여한 점수는 책임 채점관의 점수와 비교되어, 채점관이 설정된 기준에 따라 채점하고 있는지 확인합니다. 역동적 표본 추출 조정 시드는 조정 과정에서 동일한 방식으로 사용됩니다.
선임 채점관	책임 채점관을 지원하는 경험이 풍부한 채점관입니다.
기수	시험 기수 를 참고하십시오.
견본 시험지	과목별로 구성된 IB 시험의 전체 예시입니다. 견본 시험지에는 실제 평가 내용이 포함되어 있으며, 최종 평가의 형식과 수준을 그대로 반영합니다. 일반적으로 학생이 실제 시험과 유사한 조건에서 연습하도록 돕는 데 사용됩니다. MYP: 견본 시험지 패키지는 프로그램 자료실에서 확인할 수 있습니다. DP/CP: 견본 시험지는 프로그램 자료실을 통해 제공될 예정입니다. 디지털 시험의 경우, 활성화된 디지털 시험 시스템을 통해 추가 접근 권한이 제공될 수 있습니다.
기준	특정한 점수, 성적 또는 결과에 상응하는 성과 수준을 의미합니다.
표준화	조정자 또는 채점관 간에 공통된 평가 기준을 확립하기 위한 협력적 과정입니다.
표준화 회의	채점에 요구되는 기준을 설명하고 시드 답안지를 설정하기 위해 책임 채점관이 주관하는 회의입니다.
학생	IB 과정 또는 IB 교육 프로그램 참여자입니다.
학생 등록	프로그램 코디네이터가 학생을 IB 평가에 등록하기 위해 수행하는 과정입니다.

용어	정의
학생 답안 파일	MYP 온 스크린 시험에서 ‘학생 답안’이란 학생이 시험 중 완료한 과제를 의미하며, 멀티미디어가 포함될 수 있습니다. 특히 답안을 작성해 나가는 과정도 평가 대상의 일부입니다. 학생 답안은 전자 형식의 파일로 제출되며, 일반적으로 ‘학생 답안 파일’이라고 합니다.
제출	학생(또는 학생을 대신한 학교)이 최종 과제를 IB에 제출하는 과정입니다.
총괄 평가	일반적으로 학습 과정 또는 단원이 종료되는 시점에 학생의 역량이나 성취 단계를 판단하기 위해 실시되는 평가입니다.
시스템 요구 사항	IB 디지털 시험을 안전하고 원활하게 실행하기 위해 기기가 갖춰야 할 최소한의 기술 사양입니다. 이 요구 사항은 학생이 평가 중 하드웨어의 문제로 불이익을 받지 않도록 하기 위해 도입되었습니다. MYP와 DP/CP 모두에 요구 사항이 공지되어 있으며, 학교는 모든 학생 기기가 필수 사양을 충족하는지 확인해야 할 책임이 있습니다.
시스템 요구 사항 검사 도구	기기가 DP/CP 디지털 시험 실행에 필요한 기술적 요구 사항을 충족하는지 확인하는 데 사용되는 도구입니다. 이를 통해 학교는 사전에 기기를 점검하고 시험이 안전하고 원활하게 진행될 수 있는지 확인할 수 있습니다. 이 검사 도구는 시험 당일이 아닌 시험 준비 단계에서 사용되며, 공지된 시스템 요구 사항을 보완합니다.
교사용 보조 자료	교사가 IB 과정의 요구 사항을 더 잘 이해할 수 있도록 돕는 추가 정보입니다. 교사용 보조 자료는 실질적인 조언을 제공하고 교과 가이드에 제시된 이론의 이해와 시행을 돕기 위한 것입니다.
팀 리더	채점관들로 구성된 팀을 이끄는 채점관입니다.
테스트(시험)	‘테스트’는 일반적으로 시험을 지칭하는 용어로 사용되지만, 평가 관련 학술 논문에서는 특정한 의미로 사용되기도 합니다. 때로는 학생에 대한 전체 평가가 서로 다른 시기에 수행되는 여러 가지 개별 평가로 나누어지기도 합니다. 이는 ‘평가 항목’ 또는 ‘구성요소’로 불립니다. 구성요소의 예로는 구술 시험, 모든 개별 시험 또는 내부 평가를 들 수 있습니다.
허용 오차	책임 채점관의 확정적 점수와 소폭 차이가 있으나, IB는 이 정도 차이는 채점관이 여전히 올바른 기준으로 채점하고 있음을 보여줄 만큼 충분히 근접하다고 판단합니다. 채점은 판단에 기반하기 때문에 허용 오차는 필수적입니다. 숙련된 채점관이라도 동일한 과제를 재채점하면 점수가 약간 달라질 수 있기 때문입니다. 허용 오차는 점수, 문제 유형, 과목에 따라 달라집니다.
보편적 평가 설계	일부 학생을 다르게 취급하기보다는 모든 평가가 학생이 가지고 있을 수 있는 다양한 요구 사항을 이해하여 개발되어야 한다는 개념입니다. 이는 보편적 학습설계를 위한 IB의 노력의 일환입니다.
보편적 학습설계	모든 학생을 중심에 두는 학습을 위한 방향성과 프레임워크입니다. 이는 참여, 정보 제시, 표현의 다양한 수단을 제공하여 포용적인 환경을 만드는 것을 목적으로 합니다.
타당성	평가 또는 평가 결과가 목적에 맞게 사용되는지를 설명하는 용어입니다.
타당성 주장	평가가 목적에 부합함을 정당화하기 위해 평가 설계 과정에서 내린 결정에 대한 근거와 설명입니다.

용어	정의
완화된 절대 평가 (성취 기준 평가)	사전에 정의된 성취 수준 설명(기준)과 이전 학생 집단의 성과에 비추어 학생의 성취를 비교하는 방식입니다. IB는 기준을 유지하기 위해 이 접근 방법을 활용합니다.
공용어	IB가 이해관계자들과 소통할 때 사용하는 언어이자 프로그램의 시행을 위해 다양한 서비스를 제공하는 데 사용하는 언어입니다. 현재 IB의 공용어는 영어, 프랑스어, 스페인어입니다.

인쇄 가능한 자료

본 간행물 전반에는 인쇄 가능한 자료에 대한 참조가 포함되어 있습니다. 편의를 위해 인쇄할 수 있는 자료의 전체 목록이 아래에 나열되어 있습니다.

- ‘평가 주기 내 주요 역할의 책임과 의무’(PDF)
- ‘블룸(Bloom)의 교육목표 분류학’(PDF)
- ‘윤리적 사고방식 기르기’(PDF)
- ‘포용적인 접근성 조치의 시행’(PDF)
- ‘내부 평가 채점하기: 조정 과정에서 교사에게 기대되는 사항’(PDF)
- ‘역동적 표본 추출을 활용하는 조정’(PDF)
- ‘예측 성적: 교사를 위한 가이드’(PDF)
- ‘필수 품질 모델 내부 평가 답안지 범위’(PDF)
- ‘총괄 평가 및 형성 평가’(PDF)
- ‘IB의 평가 원칙’(PDF)
- ‘타당성 사슬’(PDF)