

교육과정평가연구  
The Journal of Curriculum and Evaluation  
2026, Vol. 29, No. 3, pp.277~304  
DOI: <https://doi.org/10.29221/jce.2026.29.3.277>

## Sadler의 형성평가론에 비추어 본 과정중심평가의 의미와 시사점<sup>1)</sup>

김미진 (대구대학교 겸임교수)\*

### 요약

과정중심평가는 2015 개정 교육과정 이후 우리나라 평가 정책의 핵심 용어로 자리잡았으나, 평가관과 인식론은 충분히 언어화되지 못한 채 실행되어 왔다. 본 연구는 D. R. Sadler의 형성평가론과 평가적 판단 개념을 렌즈로 삼아 과정중심평가의 의미를 재해석하고자 하였다. 과정중심평가의 정책적 전개와 실행 양상, 쟁점을 통해 교사의 질적 판단의 신뢰성 문제로 귀결됨을 확인하고, 교실평가 이론의 지형 속에서 이 물음을 판단의 인식론 수준에서 가장 일관되게 탐구해 온 Sadler의 논의를 분석하였다. 연구 결과, Sadler의 관점에서 과정중심평가에 대한 시사점은 다음과 같다. 첫째, 평가의 본질은 명시적 기준의 기계적 적용이 아니라 질에 대한 환원 불가능한 판단 행위이며, 평가의 신뢰성은 도구의 정교화가 아닌 교사 판단력의 정련에서 비롯된다. 둘째, 피드백은 결과의 통보가 아니라 학생의 평가적 판단력을 형성하는 교수 행위로 재정의될 필요가 있다. 셋째, 교사의 평가판단역량은 개인적 정련과 함께 외재화된 표준에 정렬하는 공동체적 교정을 통해 길러져야 한다. 넷째, 학생은 평가의 객체가 아니라 자기 작업의 질을 분별하는 평가적 판단의 주체로 재위치될 필요가 있다. 본 연구는 과정중심평가가 암묵적으로 전제하는 평가관을 명료화하고, 평가의 신뢰성을 교사 판단력의 정련과 평가 공동체의 정렬에서 찾는 하나의 관점을 제안하였다는 데 의의가 있다.

주제어 : 과정중심평가, Sadler, 형성평가, 평가판단역량, 피드백, 평가적 판단

1) 이 논문은 2026년 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2026S1A5A8004812)

\* 제1저자, 4007kimm@gmail.com

## I. 서론

과정중심평가는 우리나라 평가 정책의 핵심 용어로 자리잡았지만, 그것이 표방하는 가치가 어떤 평가관과 인식론 위에서 비로소 실현 가능한지는 충분히 논의되지 못하였다. 과정중심평가는 2015 개정 교육과정의 정책적 요구 속에서 결과중심평가에 대응하는 용어로 등장하였고(박정, 2017; 임은영, 2019), 이후 그 의미는 ‘과정의 평가’, ‘과정에서의 평가’, ‘교수·학습과 일체화된 평가’ 등 다양하게 변주되어 왔다. 김유정 외(2019)와 김정민(2018), 이형빈(2023) 등의 연구는 과정중심평가를 형성평가, 수행평가, 진단·총괄평가와의 관계 속에서 정의하고자 시도하였으나, 그 결과는 연구자마다 다른 개념적 지도를 그리는 데 그쳤다. 이는 과정중심평가가 어떤 평가관에 기반하여, 어떤 인식론적 전제 위에서, 어떤 교육적 목적을 향해 작동하는가에 관한 근본적 논의가 충분히 누적되지 못했음을 보여준다.

평가관에 대한 논의가 충분히 누적되지 못한 상황은 두 가지 결과로 이어진다. 첫째, 과정중심평가의 개념적 불명료성이 현장의 실천적 혼란으로 직결된다는 점이다. 과정중심평가가 학습의 ‘과정’을 평가하는 것인지, 수업 ‘중간’에 이루어지는 평가를 의미하는 것인지, 수행평가와 어떻게 구별되는지에 대한 해석이 연구자와 교사에 따라 달라지고(김영실, 2022; 김정민, 2018), 그 결과 같은 정책 용어가 서로 다른 실천을 정당화하는 데 사용된다. 둘째, 평가의 본질에 대한 성찰이 자리잡기 전에 도구의 정교화가 그 자리를 채우면서 과정중심평가의 기술적 환원이 일어난다는 점이다. 평가의 신뢰성과 공정성을 확보하기 위한 방안으로 루브릭의 정교화, 채점 기준의 세분화, 평가 도구의 객관화, 기록의 강화가 강조되며(이형빈, 2023; 이효녕 외, 2017; 홍소영, 2021), 일부 지역에서는 AI를 통한 평가 대체가 시도되고 있다. 이러한 기술적 처방은 과정중심평가가 무엇을 위한 평가이며 어떤 인식론적 행위인지에 대한 성찰이 부재한 자리를 도구의 정교화로 채우려는 시도라 할 수 있다.

과정중심평가가 표방한 가치는 학생의 변화와 성장을 지원하고, 교사의 수업을 개선하며, 교육과정-수업-평가의 일체화를 실현하는 평가이다. 평가의 신뢰성을 도구의 객관화로 확보하려는 시도는 평가의 본질을 ‘기준의 기계적 적용’으로 이해하는 측정 패러다임을 전제한다 볼 수 있다. 그러나 과정중심평가가 강조하는 학생 성장에 대한 다각도의 자료 수집과 맥락적 피드백은 측정 패러다임으로 환원될 수 없는 질적 판단의 실천이다(Sadler, 1989). 과정중심평가는 평가를 질에 대한 판단으로 보는 관점을 암묵적으로 전제하고 있다. 이 전제가 충분히 언어화되지 못한 결과, 과정중심평가는 정책 담론의 수준에서는 측정 패러다임을 넘어서고자 하지만 실천의 수준에서는 다시 측정 패러

다임의 도구들로 회수되는 역설에 놓여 있다.

과정중심평가는 단일한 이론적 구성물이 아니라 ‘학습의 과정도 중요하게 평가하자’는 정책적·실천적 지향에서 출발한 용어이며, 2015 개정 교육과정 총론에도 하나의 개념으로서 등장하지는 않는다. 그 외연 또한 형성평가·수행평가·총괄평가와 교차하거나 포괄할 만큼 넓다. 본 연구에서는 과정중심평가의 외연보다는 내포, 즉 다양한 평가 실천이 ‘과정중심적’이기 위해 공유해야 할 평가관과 인식론에 초점을 둔다. 따라서 연구를 통해 과정중심평가의 범위를 특정 이론가의 작업으로 한정하는 것이 아니라, 그 범위 안의 실천들을 관통하는 인식론적 전제를 드러내고자 한다. 이를 통해 과정중심평가를 해석하는 여러 이론 가운데 Sadler의 이론을 렌즈삼아, 과정중심평가가 전제하는 평가관을 드러내고자 한다.

D. R. Sadler의 형성평가론에 주목한 이유는 다음과 같다. 첫째, Sadler는 50년 가까이 평가의 본성, 형성평가, 피드백의 조건, 학생의 평가 역량, 교사 판단의 사회적 정렬에 관한 일관된 문제의식을 전개해 왔으며, Black & Wiliam(1998b)의 ‘블랙박스 안에서(Inside the Black Box)’ 연구를 통해 국제 형성평가 담론의 이론적 정초로 자리매김하였다. 과정중심평가가 ‘학습을 위한 평가’의 문제의식을 공유한다는 점에서(박정, 2018), 그 담론의 원류에 있는 Sadler의 논의는 우선적으로 검토될 가치가 있다. 둘째, Sadler는 과정중심평가의 실행에서 핵심 난점으로 지적되어 온 교사 판단의 신뢰성과 피드백의 실효성 문제를, 기법의 차원이 아니라 판단의 인식론이라는 근본적 차원에서 가장 일관되게 이론화한 학자이다. 이는 루브릭의 정교화와 기록의 강화로 회수되는 한국적 실천의 역설을 진단하는 데 특히 적합한 이론적 자원이 된다. 셋째, 그의 논의는 이후 평가적 판단(evaluative judgement), 피드백 리터러시(feedback literacy) 등 평가 담론으로 발전되고 있어(Boud et al., 2018; Carless & Boud, 2018; Tai et al., 2018), 과정중심평가를 동시대 담론과 접속시키는 통로가 될 수 있다. 그러나 우리나라에서 Sadler의 이론은 형성평가 논의에서 몇 차례 인용(박정, 2017, 2018)되었을 뿐, 과정중심평가와의 관련 속에서 충분히 검토되지 못하였다. Sadler가 제시한 ‘질에 대한 판단으로서의 평가’, ‘피드백의 기능’, ‘학생의 평가 역량’, ‘교사 판단 역량’ 개념은 과정중심평가가 도구의 정교화나 절차의 표준화로 환원되지 않고, 교사의 전문적 판단과 학생의 평가적 판단, 그리고 공동체적 표준 정렬이라는 세 축 위에서 실천되어야 함을 시사한다.

이에 본 연구는 Sadler의 형성평가론과 평가적 판단 개념을 통해 과정중심평가의 의미를 재해석하고, 그것이 전제하는 평가관을 명료화함으로써 실행을 뒷받침할 이론적 언어를 모색하고자 한다. 이를 위해 과정중심평가의 정책적 전개와 실행 양상, 쟁점을 검토하고, 교실평가 이론의 지형 속에서 Sadler의 논의를 판단, 피드백, 학생의 평가 역량, 교정의 순으로 살펴본다. 이를 바탕으로 과정중심평가에의 함의를 논의함으로써 과

정중심평가를 정책 용어의 수준을 넘어 평가 철학과 이론의 차원에서 재조명하고, 평가의 신뢰성을 도구의 객관화가 아닌 교사 판단력의 정련과 평가 공동체의 정렬에서 찾는 하나의 관점을 제안하고자 한다.

## II. 과정중심평가의 특징과 한계

### 1. 과정중심평가의 전개와 개념

과정중심평가의 정책적 출발점은 2015 개정 교육과정 총론으로, ‘학습의 결과뿐 아니라 학습의 과정을 중시하는 평가’를 강화하고, 학생이 자신의 학습을 성찰하도록 하며, 평가 결과를 교수·학습의 질 개선에 활용할 것을 명시하였다(교육부, 2015). 교육과정 총론에서 과정중심평가를 하나의 개념어로 제시하지 않았지만 방향성은 이후 교육부와 시·도교육청의 정책 문서를 거치며 ‘과정중심평가’로 명명되었고, 교육부와 한국교육과정평가원(2017)의 현장 지원 자료에 이르러 ‘교육과정의 성취기준에 기반한 평가 계획에 따라 교수-학습 과정에서 학생의 변화와 성장에 대한 자료를 다각도로 수집하여 적절한 피드백을 제공하는 평가’(교육부, 한국교육과정평가원, 2017, pp. 6-7)라는 공식적으로 정의하였다. 이 정의에 따르면 과정중심평가는 성취기준 기반성, 수업과 평가의 연계, 과정과 결과의 통합, 다각적 자료 수집, 피드백과 교수·학습 개선이라는 요소를 포함한다. 2022 개정 교육과정은 이 기초를 계승하면서 학생의 개별적 특성과 성장을 고려하는 학생 맞춤형 평가, 자기평가와 동료평가를 통한 학습자의 평가 참여, 평가 결과에 기초한 개별 피드백으로 이를 확장하였다(교육부, 2022). 과정중심평가는 평가의 새로운 유형이기 보다는 우리나라 맥락에서 통용되는 정책 용어에 가까우며, 특정 시점에 완성된 개념이 아니라, 약 10년에 걸쳐 ‘학습 결과에 대한 평가’에서 ‘학습을 위한 평가’와 ‘학습으로서의 평가’로 평가의 기능을 확장해 온(임은영, 2019) 정책적 전개의 산물이다.

개념적 차원에서 과정중심평가는 2015 개정 교육과정과 함께 결과중심평가에 대응하는 개념으로 등장하였으나(박정, 2017; 임은영, 2019), 논의가 진행되면서 학습의 과정과 결과까지 포함하는 의미를 지니게 되었다. 김유정 외(2019)는 과정중심평가와 수행평가, 형성평가를 비교하면서 과정중심평가는 학습의 결과를 포함한다는 점에서 형성평가보다 포괄적인 개념으로 설정한 바 있다. 이형빈(2015) 역시 과정중심평가는 진단평가와 형성평가, 총괄평가를 모두 포함한다고 설명하였다. 임종헌과 최원석(2018)은 자유학기제의 평가 사례 분석을 통해 과정중심평가를 성취기준에 근거하고, 수업과 평가

의 일체화를 추구하며, 결과물의 수월성보다 성장에 관심을 두되, 지필평가와 수행평가 모두를 통해 실행될 수 있는 평가 운영 원리로 체계화하였다. 이러한 논의들은 과정중심 평가가 특정 평가 방법이 아니라 평가의 방향 내지 관점이며, 결과를 배제하는 것이 아니라 과정과 결과를 통합적으로 평가하는 것임을 보여준다.

일각에서는 과정중심평가가 교수-학습에서 평가의 ‘활용’을 강조한다는 점에서 ‘형성 평가의 재등장’으로 해석하기도 한다(박정, 2018). 평가의 세 가지 목적인 책무성, 자격증, 학습 가운데 형성평가는 ‘학습을 위한 평가’에 해당하며(Black et al., 2003: 2), 과정중심평가 역시 교사의 피드백과 학습의 개선을 목적으로 한다는 점에서 형성평가의 문제의식과 맞닿아 있다. 우리나라 맥락에서 진단평가와 총괄평가는 평가의 주체가 교실 외부에 위치하는 경우가 많지만, 형성평가는 수업을 하는 교사가 그 주체가 된다는 점에서 과정중심평가와 상통한다. 이때 ‘형성평가의 재등장’은 특정 시기의 평가로의 회귀가 아니라 학습을 위한 평가로의 확장으로 해석되어야 하는데, 국제적으로 형성평가 담론 자체가 수업 ‘중간’의 평가에 한정되지 않고 총괄평가의 형성적(formative) 활용까지 포괄하는 방향으로 전개되어 왔기 때문이다(Black et al., 2003: 63). 이는 우리나라 과정중심평가가 학습 과정에서의 평가를 강조하며 등장하였다가 학습의 결과까지 포함하게 된 맥락과도 상응한다.

과정중심평가는 교육과정-수업-평가의 일체화를 전제로 작동하며, 학생의 학습을 돕고 교사의 수업을 개선하는 것을 목표로 한다. 이정민(2024)은 과정중심평가는 교육과정 자율화 정책에 기반함을 강조하며, 교사의 교육과정 자율권 확대에 따라 교수-학습 방법이나 교육과정 편성 영역에서는 자율성이 발휘되지만 평가 영역에서는 그렇지 않다는 문제를 지적한다(반재천 외, 2018; 양연동, 2020). 교육과정과 수업, 평가가 분리된 것이 아니라면 교사의 자율적인 평가 실천 역시 중요해진다. 이처럼 과정중심평가는 정책적 전개를 통해 개념적 요소를 갖추어 왔으나, 학교 현장에서 실행되는 양상은 정책이 표방한 방향과 일정한 거리를 보인다(김영실, 2023).

국가 및 시도 수준의 실태 조사와 사례 연구들은 과정중심평가의 실행에서 다음의 공통된 양상을 보고한다. 첫째, 교사들은 과정중심평가를 ‘수업 중간에 하는 평가’, ‘지필평가를 대체하는 평가’로 협소하게 인식하는 경향이 있다(김영실, 2022). 기록하지 않는 평가에 비해 NEIS 기록이 동반되는 평가를 온정적이고 융통성 있게 접근하는 등, 평가의 실행이 기록 여부에 따라 달라지는 양상도 나타난다(김영실, 2022). 둘째, 피드백은 결과물에 간단한 메모를 덧붙이거나 결과를 통보하는 방식으로 이루어지며, 수행의 질에 관한 정보는 빈약하다(김영실, 2022). 셋째, 교사들이 실행에서 가장 우려하는 것은 평가의 객관성과 공정성 확보이며(반재천 외, 2018; 신혜진, 안소연, 김유원, 2017), 이에 대한 현장의 대응은 루브릭과 채점 기준의 세분화, 기록의 강화로 수렴한다. 넷째,

다수 학생을 지속적으로 관찰·기록하고 개별 피드백을 제공해야 하는 부담이 실행의 현실적 제약으로 작용한다. 과정중심평가 관련 연구의 동향에서도 이러한 현실이 확인되는데, 중등학교를 대상으로 한 연구들의 주요 주제는 정책 실행 과정에서의 교사의 인식과 어려움으로 나타난다(최훈원, 2024).

## 2. 과정중심평가의 쟁점과 한계

과정중심평가 관련 정책 문서와 실태 조사 연구(교육부, 한국교육과정평가원, 2017; 반재천 외, 2018; 신혜진 외, 2017), 개념 연구(김유정 외, 2019; 김정민, 2018; 임종현, 최원석, 2018), 실행 연구(김영실, 2022; 양연동, 2020)에서 공통적으로 제기되어 온 문제들을 교사의 평가 전문성, 개념적 합의의 부족, 객관성과 공정성의 세 범주로 요약된다.

먼저, 교사의 평가 전문성 문제이다. 전통적으로 교사의 평가 전문성은 평가 실행에 필요한 지식과 기능 중심이었으나(AFT, NCME, & NEA, 1990; Plake, Impara, & Fager, 1993), 형성평가 논의를 기점으로 최근에는 학습을 위한 평가와 성취기준 평가 등을 반영하고 있으며, 평가 전문성을 향상시키기 위한 프로그램도 평가 지식과 기능보다는 교사 스스로의 지속적인 성찰과 공동체 속 협력을 지향하고 있다(Xu & Brown, 2016). 표준화 검사는 탈맥락적이며, 평가의 주체가 교실의 외부이며, 교사에게 요구되는 것은 공정성과 평가의 결과를 해석하는 능력이다. 반면, 과정중심평가는 수업의 상황을 전제하므로 맥락적이며, 다양한 변인들이 역동적으로 상호작용하며 교사의 판단에 영향을 미친다(우필희 외, 2021). 이에 과정중심평가에서는 평가에 관한 지식뿐 아니라, 교사의 평가 신념과 경험, 상호작용의 맥락, 학생과의 의사소통 역량, 협력적 역량 등을 평가에 필요한 교사의 전문성으로 요구하고 있다(성태제, 임현정, 2014; Adie, 2013; Xu & Brown, 2016). 특히, 김유정 외(2019)는 교사의 과정중심평가 역량 진단도구 개발을 위해 역량의 구인으로 과정중심평가 이해, 과정중심평가 계획, 과정중심평가 실행, 의사소통, 과정중심평가 결과 활용 등 다섯 가지를 제시하기도 하였다. 그러나 교사들은 실제 수행보다 자신의 평가 역량을 더 긍정적으로 인식하는 경향이 있기 때문에 자기보고식 설문은 한계가 있다(김나영, 2023). 또한, 평가 전문성의 기준에 관한 선행연구들은 ‘계획’에만 지나치게 치중하거나, 평가의 단계와 상황을 구체적으로 명시하지 않고 있다는 점(우필희 외, 2021), 과정중심평가에서 요구되는 교사의 평가 전문성은 양적이기보다 질적으로 접근할 필요가 있다는 점에서 여전히 한계가 있다. 요컨대 선행연구들은 교사가 갖추어야 할 역량의 목록을 제시하는 데에는 상세하지만, 그 역량의 핵심에 있는 질적 판단이 실제로 어떻게 이루어지고 길러지는가에 대해서는 충분히 설명하지 못하고 있다.

다음으로, 과정중심평가의 개념에 대한 합의가 부족하다는 점이다. 과정중심평가가 등장하면서 수행평가와는 어떤 차이점이 있는지, 형성평가와 같은 것인지, 교-수-평-기 일체화와는 어떻게 관련되는지, 학습의 결과가 아닌 ‘과정’ 그 자체를 평가하는 것인지, 시기적으로 수업이 진행되는 ‘중간’에 평가하는 것인지 등이 명료하게 제시되지 않아 현장에 혼란을 주고 있다. 이에 대해 김정민(2018)은 평가과제의 수행에 있어 과정을 보다 중점적으로 평가하는 것, 과정중심 수행평가도 가능하다고 하였고, 김유정 외(2019)는 과정중심평가가 형성평가와 수행평가, 지필평가를 모두 포괄하는 개념으로 설명하였다. 과정중심평가에 대한 엇갈린 개념 정의들은 과정중심평가의 실천에도 혼란을 가져온다. 특히, 산출물 중심의 수행평가는 과정중심평가가 아니라는 몇몇 연구들(김유정 외, 2019; 김정민, 2018)은 산출물을 학습의 결과로 볼 수 있고, 결과도 포함하는 것이 아니냐는 의문이 생기게 만든다. 교사의 실천과 변화를 가져오기 위해서는 과정중심평가가 무엇인지에 대한 ‘이해’가 선행되어야 한다는 점에서(김유정 외, 2019) 과정중심평가의 명료한 정의와 이에 대한 교육계의 합의가 우선될 필요가 있다. 한편, 정정철(2020)은 과정중심평가가 과정과 결과, 성장 지원과 선발, 학습과 평가라는 상충하는 기능들 사이의 긴장을 안고 있으며, 과정에 대한 관찰이 참여도·성실성 평가로 변질될 가능성을 철학적으로 분석한 바 있다. 과정중심평가의 어려움은 단순히 용어 정의의 문제가 아니라, 서로 다른 평가 기능을 하나의 실천 안에서 교육적으로 조정해야 하는 문제이기도 하다.

마지막으로, 과정중심평가는 탈맥락적인 양적 평가이기보다는 교사가 수업이라는 맥락에서 실시하는 양적·질적 평가라는 점에서 객관성과 공정성의 문제가 제기된다. 과정중심평가 실태 조사 연구(반재천 외, 2018; 신혜진, 안소연, 김유원, 2017)에 따르면 현장에서 가장 우려되는 요소로 평가의 객관성과 공정성 확보 문제를 들고 있다. 과정중심평가에서 활용하는 수행평가, 포트폴리오 평가, 자기평가, 교사 관찰 등과 같은 평가 방법들이 활용하는 체크리스트는 결국 학생의 학습의 정도나 수준을 객관적으로 ‘측정’하는 것이며, 수업의 ‘과정’에서 작은 단위의 총괄평가를 자주 실시하여 평가 결과의 객관성을 확보하고자 하는 시도 역시 측정의 관점에 기반한다(박정, 2018). 이러한 현실은 평가에 대한 관점을 전환하려는 잦은 시도에도 불구하고 여전히 가시적인 결과에 대해 기준을 적용하는 데에서 진일보하지 못하고 있음을 보여준다.

종합하면, 과정중심평가는 약 10년의 정책적 전개를 통해 성취기준 기반, 과정과 결과의 통합, 피드백과 성장 지원이라는 개념적 요소를 갖추어 왔으나, 실행의 수준에서는 협소한 인식, 통보형 피드백, 객관성 확보를 위한 도구의 정교화로 회수되는 양상을 보인다. 교사 평가 전문성의 성격, 개념적 합의의 부족, 객관성과 공정성과 같은 쟁점들은 결국 ‘수업이라는 맥락 속에서 이루어지는 교사의 질적 판단은 어떻게 형성되고, 정당화

되며, 신뢰될 수 있는가'라는 질문과 연결된다. 정책 자료와 선행연구는 교사가 '무엇을 해야 하는가'를 제시하는 데에는 상세하지만, 교사가 학생 수행의 질을 어떻게 알아보고 판단하는가라는 이 물음에 대해서는 충분히 답하지 못하고 있다. Sadler는 이 질문에 대해 평가 기법이 아니라 판단의 인식론이라는 수준에서 일관되게 탐구해 왔다. 이에 다음 장에서는 교실평가 이론의 지형 속에서 Sadler의 위치를 확인한 뒤, 그의 논의를 검토하고자 한다.

### III. Sadler의 형성평가론과 평가적 판단

D. Sadler는 교육평가 분야의 이론가로, Queensland 대학교 및 Griffith 대학교에서 50년 가까이 일관된 문제의식을 가지고 평가의 본성, 형성평가, 학습자의 평가 역량 등을 연구해 왔다. 본 장에서는 교실평가 이론의 지형 속에서 Sadler의 위치를 확인한 후, 그의 사유가 전개된 논리를 따라 평가의 본질과 피드백의 조건, 학생의 평가 역량, 교사 판단의 사회적 정렬 순으로 검토하고자 한다. 분석의 대상은 네 개념의 형성과 전개를 보여주는 1985년부터 2013년까지의 주요 논문 8편이다. 이들은 그의 저작 중 평가 담론에서 가장 널리 인용되는 논문들로, Black과 Wiliam(1998a)의 문헌 종합과 이후 평가적 판단 연구(Boud 외, 2018; Tai 외, 2018)가 공통적으로 참조하는 문헌들이며, 초기(기준과 표준, 1985; 1987), 중기(형성평가와 피드백, 1989; 1998), 후기(판단의 비결정성과 교정, 2005; 2009; 2010; 2013)에 걸쳐 그의 문제의식의 연속과 전개를 확인할 수 있도록 선정하였다. 분석의 주요 대상이 된 연구물의 목록은 아래의 표와 같다.

<표 1> Sadler의 주요 연구 목록

| 연도   | 제목                                                                    | 주요 내용                                                                                                                                                                                                                                                       |
|------|-----------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1985 | 평가 기준의 본질과 기능<br>The Origins and Functions of Evaluative Criteria     | <ul style="list-style-type: none"> <li>• 기준(criteria)과 표준(standards), 절대적 판단(absolute judgement)과 상대적 판단(relative judgement)을 구분하였음</li> <li>• 전체적(global) 판단이 평가의 1차적 행위이며, 기준은 그 판단을 사후적으로 정당화하기 위해 도출됨</li> </ul>                                        |
| 1987 | 성취 표준을 명시하고 공표하기<br>Specifying and promulgating achievement standards | <ul style="list-style-type: none"> <li>• 기준(criterion)과 표준(standard)의 엄밀한 구별</li> <li>• 표준 참조 평가(standards-referenced assessment)의 제안</li> <li>• 표준의 명시화(specifying)와 공표(promulgating) 문제</li> <li>• 참조 수준의 외재화(externalising a reference level)</li> </ul> |
| 1989 | 형성평가와 교수 구조 설계                                                        | <ul style="list-style-type: none"> <li>• 형성평가 이론의 적용 범위: 복수 기준(multiple criteria)을 활용하는 질적 학습 결과에 적용함</li> </ul>                                                                                                                                            |

| 연도   | 제목                                                                                                            | 주요 내용                                                                                                                                                                                                                                                                                                          |
|------|---------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|      | Formative Assessment and the Design of Instructional Systems                                                  | <ul style="list-style-type: none"> <li>피드백의 세 가지 조건을 다음과 같이 제시함: ① 양질의 작업에 대한 인식(concept of quality), ② 자신의 작업을 비교·점검할 수 있는 평가적 기술(evaluative skill), ③ 작업을 수정할 수 있는 전략(store of tactics)</li> <li>학생이 스스로 작업의 질을 모니터할 역량을 길러야 학습이 가능함</li> </ul>                                                              |
| 1998 | 형성평가: 그 영역을 다시 살펴보다<br>Formative Assessment: Revisiting the Territory                                         | <ul style="list-style-type: none"> <li>피드백의 양이 아니라 질(quality)이 형성평가의 핵심임</li> <li>교실 평가 사건의 3단계로 제안함: ① 학생 산출물에 교사의 주의 환기 → ② 루브릭에 비추어 평가 → ③ 판단·행동 결정</li> <li>자기(ego-involving) 피드백과 과제(task-involving) 피드백을 구분함</li> </ul>                                                                                  |
| 2005 | 고등교육에서 기준 기반 평가와 채점에 대한 검토<br>Interpretations of Criteria-based Assessment and Grading in Higher Education    | <ul style="list-style-type: none"> <li>기준 기반 채점(criteria-based grading) 정책들에 대한 비판적 검토</li> <li>criteria와 standards가 자주 혼동되며, 교사의 근본적 판단은 여전히 주관적이고 학생에게 보이지 않은 채 남음</li> <li>기준 기반(criteria-referenced)에서 표준 참조(standards-referenced)으로의 이동을 제안하였음</li> </ul>                                                 |
| 2009 | 총체적 평가와 채점에서 복합적 학습으로의 전환<br>Transforming Holistic Assessment and Grading into a Vehicle for Complex Learning | <ul style="list-style-type: none"> <li>총체적(holistic) 평가와 분석적(analytic) 평가의 구별 및 변증법적 관계를 분석하였음</li> <li>전문가는 두 방식을 끊임없이 오가며 판단을 형성함. 전경(criterion-bound)과 배경(holistic)의 교차</li> <li>명시적(explicit), 잠재적(latent), 메타 기준(meta-criteria)의 분류 체계를 제안하였음</li> <li>감식안(connoisseurship): 전문적 판단의 역할을 제시하였음</li> </ul> |
| 2010 | 피드백을 넘어: 복합적 평가 역량 발달시키기<br>Beyond Feedback: Developing Student Capability in Complex Appraisal               | <ul style="list-style-type: none"> <li>피드백을 받는 것만으로는 학습 향상이 일어나지 않음.</li> <li>학생이 스스로 전유해야 할 세 가지 근본 개념을 제시하였음: ① 과제 준수(task compliance), ② 질(quality), ③ 기준(criteria)</li> <li>명시적 루브릭 뿐 아니라, 평가의 암묵적 지식(tacit knowledge) 축적 필요</li> <li>복합적 평가 역량(complex appraisal capability) 개념을 제시함</li> </ul>           |
| 2013 | 학문적 성취 기준: 조정에서 교정으로<br>Assuring Academic Achievement Standards: From Moderation to Calibration               | <ul style="list-style-type: none"> <li>사회적 조정(moderation)에서 교정(calibration)으로의 전환을 제안</li> <li>기준 기술문(standards statements)과 사전 명시된 루브릭만으로는 일관된 채점을 보장할 수 없음을 강조함</li> <li>교정된 예시(calibrated exemplars)와 해설(commentaries)을 통해 교사 판단의 공유된 참조 틀을 구축할 것을 제안함</li> </ul>                                           |

## 1. 질적 평가 이론의 흐름과 Sadler의 위치

측정 중심의 평가관을 넘어서려는 이론적 시도는 있어 왔다. 1990년대 이후 교실평가 분야에서는 평가를 학습과 분리된 사후 측정이 아니라 교수·학습을 구성하고 조절하는 교육적 실천으로 재개념화하려는 흐름이 형성되었고, 이들은 평가 정보가 교수와 학습의 개선에 활용되어야 하고, 학생이 평가의 수동적 대상이 아니라 참여자이며, 교사의 해석과 전문적 판단이 중요하다는 인식을 공유한다.

Black과 Wiliam(1998a; 2009)는 ‘학습의 증거를 어떻게 수집하고 활용하여 다음 교수·학습을 개선할 것인가’에 대하여 방대한 문헌 검토를 통해 형성평가의 학업성취 효과를 종합하고, 대규모 평가 중심의 책무성 정책이 교수·학습 과정을 등한시해 왔다는 문제의식 아래 「블랙박스 안에서(Inside the Black Box)」(Black & Wiliam, 1998b)를 통해 국제적인 형성평가 담론을 촉발하였다. 이후 형성과 총괄을 평가의 시기가 아니라 기능으로 구분하고 총괄평가의 형성적 활용까지 포괄하는 방향으로 논의를 확장하였는데(Black et al., 2003), 이는 형성평가 중심으로 등장했다가 결과까지 포함하게 된 우리나라 과정중심평가의 전개와 상응한다. Shepard(2000)는 ‘새로운 학습이론에 부합하는 평가문화는 어떠해야 하는가’에 관하여 행동주의·심리측정 전통의 평가가 사회구성주의적 학습관과 양립할 수 없음을 논증하고, 평가를 학급의 학습문화를 형성하는 활동으로 재정의하였다. Allal(2013; 2020)는 ‘평가를 통해 학습이 어떻게 조절되는가’에 대하여, 형성평가를 상호작용적·소급적·선제적 조절의 과정으로 설명하고, 교사의 평가 판단을 인지적 행위인 동시에 사회적으로 위치지어진 실천으로 개념화하였다. Gipps(1994, 1999)는 ‘평가는 어떤 사회문화적 조건 속에서 공정하고 타당할 수 있는가’에 대하여, 심리측정 패러다임의 신뢰도·객관성 기준이 수행평가의 교육적 가치를 포착하지 못함을 지적하고, 평가의 타당성을 학습 기회와 형평성의 문제와 연결하였다.

Sadler는 이들 모두가 전제하면서도 직접적으로 다루지는 않은 판단의 인식론 질문을 던졌다. 교사는 학생 수행의 질을 어떻게 알아보며, 그 판단은 어떻게 정당화되고 신뢰될 수 있는가. Black과 Wiliam은 교사가 학습의 증거를 해석하여 수업을 조정해야 한다고 말하지만, 무엇이 의미 있는 증거인지를 교사가 어떻게 알아보는지는 이론화하지 않았다. 실제로 이들의 ‘블랙박스’ 연구가 명시적으로 이론적 토대로 삼은 것이 Sadler(1989)의 형성평가론이라는 점은, Sadler의 논의가 이 지형에서 병렬적 대안이 아니라 기저 이론의 위치에 있음을 보여준다. Shepard는 평가가 학습문화 속에서 학생의 사고를 해석하는 행위라고 보았지만, 복합적 수행의 질에 대한 지각과 판별이 어떻게 이루어지는지는 구체화하지 않았다. Allal은 교사의 전문적 판단에 가장 근접하였으나, 그의 관심은 판단이 학습의 조절로 이어지는 과정에 있으며, 판단 자체의 인식론적 구조와 판단의 신뢰성을 확보하는 제도적 장치는 그의 이론의 중심 주제가 아니었다. Gipps는 측정 패러다임 비판에서 Sadler와 가장 가까우나, 그 비판을 평가의 사회문화적 조건과 형평성의 방향으로 전개하였고, 교사 판단의 질을 어떻게 정련하고 정렬할 것인가라는 문제에는 별도의 답을 제시하지 않았다.

앞서 보았듯이 과정중심평가의 세 가지 쟁점은 ‘수업 맥락에서 이루어지는 교사의 질적 판단은 어떻게 신뢰될 수 있는가’라는 물음에 대한 해답을 요구하며, 현장은 루브릭의 세분화와 기록의 강화, AI 도구의 사용이라는 도구적 처방으로 응답해 왔다. Black과

William의 실천 전략은 형성평가를 ‘어떻게 할 것인가’에는 풍부한 답을 주지만, 교사의 판단이 신뢰될 수 있는가라는 의문을 해소하는 논리는 제공하지 않는다. Shepard의 학습문화론과 Gipps의 사회문화적 관점은 왜 측정 패러다임을 넘어서야 하는지를 설득하지만, 판단의 신뢰성을 무엇으로 확보할 것인지에 대해서는 논의가 부족하다. 이에 비해 Sadler는 이 물음에 대한 진단과 해법을 모두 제시한다. 진단의 차원에서 그는 아무리 세분화된 기준도 그 적용에는 판단이 요구되므로 판단의 제거는 원리적으로 불가능하고 (Sadler, 1985, 2009), 해법의 차원에서 판단의 신뢰성을 도구의 정교화가 아니라 교사 판단의 정련과 외재화된 표준에의 교정(calibration)에서 찾는 경로를 제안한다(Sadler, 2013). 과정중심평가가 부딪힌 난점, 즉 질적 판단을 표방하면서 그 신뢰성에 대한 불안 때문에 측정의 문법으로 회귀하는 역설을 이론의 수준에서 해부하고 대안적 경로를 제시하였다는 점에서, 그의 이론은 과정중심평가의 철학과 실천을 설명하는 데 상대적으로 적합하다고 보았다.

## 2. 평가의 본질: 질(quality)에 대한 판단(judgement)

Sadler(1985)에게 평가의 본질은 명시적 기준의 기계적 적용이 아니라, 질에 대한 환원 불가능한 판단 행위이다. ‘사전에 정의된 기준에 학생의 작업을 비추어 판정하는 작업’이라는 통상적인 평가 개념은 평가의 실제 작동 방식과 부합하지 않으며, 평가에 대한 보다 근본적인 인식론적 검토가 요청된다고 주장한다.

‘명시적 기준의 기계적 적용이 아니라’는 주장은 인식의 선행성과 기준의 불완전성에 의해 뒷받침된다. Sadler(1985)는 평가의 역사를 거슬러 올라가면서, 인간이 어떤 것의 좋음을 알아보는 능력이 그것을 평가하는 기준을 정식화하는 능력보다 시간적, 논리적으로 앞선다고 논증한다. 기준은 이미 이루어진 판단을 사후적으로 정당화하는 자원이자, 판단의 출발점이 아니라는 것이다. Sadler(1989)는 또한 평가에 사용되는 대부분의 기준이 애매하다(fuzzy)는 점을 지적한다. 예를 들어 독창성, 일관성, 명료함 같은 기준은 맞고 틀림처럼 이분법적으로 적용될 수 없고 정도(degree)의 문제이므로 그 정도를 판정하려면 결국 또 다른 판단이 필요하게 된다. 따라서 기준의 기계적 적용으로 평가가 완결된다는 가정 자체가 성립하지 않는다는 것이다. 설령 기준들이 모두 명시되었다고 하더라도, ‘어떤’ 기준을 ‘언제’, ‘어떻게’ 적용할 것인가라는 상위 판단이 또 필요하게 된다. Sadler(1989)는 이를 메타 기준이라 부르고, 전문가의 평가 판단력은 본질적으로 이 메타 기준을 다루는 능력이라고 본다.

‘질에 대한 환원 불가능한 판단 행위’라는 주장은 질의 성격과 전문적 감식안(connoisseurship)으로 뒷받침된다. Sadler(1985, 2010)에게 질은 부분들의 합산이 아니

라 전체로서의 통합적 성질을 지닌다. 어떤 글이나 음악이 좋은지는 문법, 구성, 표현, 내용 같은 요소들의 점수의 총합으로 결정되기 보다는 전체적인 탁월함으로 판단된다. 질은 이런 형상적 통일체로서만 파악될 수 있으며, 따라서 그 판단은 형상적 인식의 성격을 띤다. Sadler(1989, 2010)는 Polanyi의 인식론을 끌어와, 평가 전문성이 언어로 완전히 정식화될 수 없는 암묵적 지식(tacit knowledge)에 본질적으로 기반한다고 주장한다. 경험 많은 교사는 학생 글을 보면서 ‘이 글이 왜 좋은가’를 즉각 알아보지만, 그 판단의 기반을 모두 명시적 규칙으로 풀어내지는 못한다. 이는 전문가의 결함이 아니라 평가 판단의 본성이 그러하기 때문이다. Sadler(2009)는 평가에 관한 경험적 연구를 인용하면서, 숙련된 평가자들이 작업의 전체적 인상을 먼저 형성한 뒤 그 인상을 기준에 비추어 설명한다는 것을 논증한다. 그들은 분석적 채점과 전체적 채점을 끊임없이 오가며, 어느 한쪽으로 환원되지 않는 판단을 만들어 낸다. 이는 평가의 본성이 분석적 절차가 아니라 통합적 판단임을 경험적으로 뒷받침하는 것이다.

### 3. 형성평가와 피드백의 조건

Sadler의 1989년 「형성평가와 교수 구조 설계」는 그의 핵심적인 연구물이다. 이 논문은 전 세계적으로 평가 개선 운동이 확산되게 한 Black & William(1998b)의 「블랙박스 안에서」 논문에 이론적 토대를 제공하면서 그에게 ‘형성평가의 현대적 정초자’라는 타이틀을 주었다. 이어진 1998년 「형성평가: 그 영역을 다시 살펴보다」는 Black & William의 논문에 대한 응답으로, 피드백에 대해 보다 정교한 아이디어를 담고 있다.

Sadler(1989, 1998)는 형성평가를 평가가 이루어지는 시점이 아니라 목적과 기능으로 정의한다. 형성평가는 학생 수행의 질에 관한 판단을 통해 시행착오 학습의 비효율과 무작위성을 건너뛰(short-circuit)으로써 역량을 개선하는 데 관심을 두는 반면, 총괄평가는 성취 상태를 요약, 보고하는 데 목적이 있어 학습에 즉각적 영향을 주지 않는다. 따라서 형성과 총괄을 가르는 기준은 평가의 시기가 아니라 교사의 판단이 어떻게 쓰이는 가이며, 같은 자료라도 격차를 줄이는 데 사용되면 형성적이고 기록, 보고에 그치면 그렇지 않다(Sadler, 1989). 수업 중간에 수집한 정보라도 학생이 이를 자기 작업의 수정에 활용하지 못하면 형성적이지 않다. 형성평가의 효과는 피드백의 양이 아니라 질과 학습자의 활용 가능성에 달려 있으며(Sadler, 1998), 학생의 평가 역량은 다음 절에서 다룬다.

Sadler(1989)는 Ramaprasad(1983)를 인용하여 피드백을 ‘실제 수준과 기준 수준 사이의 격차에 관한 정보로서, 그 격차를 줄이는 데 사용되는 것’으로 규정한다. 피드백이 격차를 바꿀 주체가 아닌 제3자에게 전달되거나 지나치게 부호화, 요약되어 행동으로

이어지지 못하면, 그것은 피드백이 아니라 ‘허공에 매달린 자료(dangling data)’에 불과하다는 것이다(Sadler, 1989). 이처럼 성적이나 점수가 학생에게 일방향 암호가 되면 형성적으로 오히려 역효과를 낼 수 있다.

그는 피드백이 형성적으로 기능하기 위한 세 가지 조건을 제시한다. 첫째, 질에 대한 인식(concept of quality)이 있어야 하고, 둘째, 현재 수행과 표준을 비교할 능력(evaluative skill)을 갖추어야 하며, 셋째, 격차를 줄이는 전략(store of tactics)이 있어야 한다. 이 세 조건은 순차적이 아니라 동시에 충족되어야 한다(Sadler, 1989).

Sadler는 칭찬이 형성평가에 미치는 역설적 효과를 줄기 위해 자아 관련 피드백(ego-involving)과 과제 관련 피드백(task-involving)을 구분한다.

<표 2> 자아 관련 피드백과 과제 관련 피드백의 비교

| 구분      | 자아 관련(ego-involving) 피드백                                                                                                           | 과제 관련(task-involving) 피드백                                                                                                                         |
|---------|------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| 겨냥하는 대상 | 학생의 자아·정체성·능력                                                                                                                      | 학생이 수행한 작업의 질                                                                                                                                     |
| 평가의 초점  | 너는 어떤 사람인가                                                                                                                         | 이 작업은 어떠한가                                                                                                                                        |
| 예시      | <ul style="list-style-type: none"> <li>• 너 참 똑똑하구나, 잘 썼어.</li> <li>• 역시 너는 글쓰기에 재능이 있어.</li> <li>• 이번에도 열심히 노력했네, 대견하다.</li> </ul> | <ul style="list-style-type: none"> <li>• 이 글의 두 번째 문단은 첫 문단의 주장을 충분히 뒷받침하지 못하고 있어.</li> <li>• 도입부의 문제 제기가 명료해서 독자가 글의 방향을 쉽게 따라갈 수 있어.</li> </ul> |

자아 관련 피드백은 학생의 능력, 재능, 노력, 성격, 정체성에 대해 판단함으로써 학생이라는 사람을 평가한다. 반면, 과제 관련 피드백은 학생이 만들어낸 작업 자체에 대해 판단하고 있다. Black & Wiliam(1998a)에 따르면, 교사들은 학생을 격려하기 위해 질이 낮은 작업에도 칭찬을 덧붙이는 경향이 있다. 자아 관련 피드백은 학생의 자존감을 보호한다는 좋은 의도에서 나왔지만, 두 가지 문제가 있다는 점을 지적하였다. 첫째, 학생이 자신의 작업이 실제로 어디서 부족한지를 알지 못한 채 칭찬을 받게 된다. 둘째, 학생이 자아 평가에 매몰되어, 과제에 대한 객관적 평가 능력을 기르지 못하게 된다. 평가의 초점이 자기 자신에서 작업으로 이동해야 학생이 스스로의 작업을 판단할 수 있게 되는데, 자아 관련 피드백은 이 이동을 방해한다는 것이다. Sadler는 형성평가가 효과적으로 작동하려면 피드백이 자아가 아니라 과제를 겨냥해야 한다고 주장한다.

피드백의 세 가지 조건은 모두 피드백을 제공하는 교사가 아니라 그것을 사용하는 학생 안에서 충족되어야 하는 조건들이다. 때문에 Sadler의 피드백 이론은 학생의 평가 역량이라는 주제로 필연적으로 나아가게 된다.

#### 4. 학생의 평가 역량과 평가의 사회적 차원

Sadler의 평가에 대한 사유는 학생의 평가 역량으로 귀결된다. 형성평가의 성패는 교사가 무엇을 제공하는가가 아니라 학생이 무엇을 할 수 있는가에 달려 있기 때문이다. 학교 교육에서 제공하는 모든 교수와 평가가 궁극적으로 학생의 학습과 성장을 지향한다는 점에서(홍후조, 2026), 평가론의 자연스러운 종착점이기도 하다.

Sadler(1989)는 이를 ‘길드 지식(guild knowledge)의 이양’이라는 개념으로 표현한다. 질을 알아보는 지식은 오랜 판단 경험을 통해 교사 집단 안에 축적된 일종의 장인적 지식인데, 형성평가의 목적은 이 지식을 교사가 독점한 채 판정 결과만 전달하는 것이 아니라, 학생이 점진적으로 공유하도록 이양하는 데 있다는 것이다. 학생의 평가 전문성(evaluative expertise)은 이러한 맥락에서 제안된 개념으로, 단순한 자기평가 기법이 아니라 작업의 질을 분별하는 본질적인 능력을 뜻한다(Sadler, 1989).

그러나 20여 년의 실행 경험을 통해 피드백을 받는 것만으로 학습 향상이 일어나지 않는다는 사실을 반복적으로 확인하였고, Sadler(2010)는 「피드백을 넘어: 복합적 평가 역량 발달시키기」에서 이 문제를 정면으로 재검토한다. 그는 피드백을 학습으로 전환하는 능력을 ‘복합적 평가 역량(complex appraisal capability)’으로 개념화하는데, 여기에는 피드백의 의미를 해독하는 능력, 자기 작업의 어느 부분이 피드백과 관련되는지 식별하는 능력, 세 핵심 개념(과제 준수, 질, 기준)을 내면화·작동화·구체화하는 능력이 포함된다. 1989년의 평가 전문성이 작업 중의 자기 모니터링에 초점을 두었다면, 2010년의 복합적 평가 역량은 외부에서 주어진 피드백 정보를 해석하고 자신의 학습에 연결하는 능력까지 포괄한다.

<표 3> 학생 평가 역량으로서 평가 전문성과 복합적 평가 역량 비교

| 구분       | 평가 전문성<br>(evaluative expertise, 1989) | 복합적 평가 역량<br>(complex appraisal capability, 2010) |
|----------|----------------------------------------|---------------------------------------------------|
| 맥락       | 형성평가 이론의 정초                            | 피드백 이론의 재검토                                       |
| 일차적 기능   | 작업 중 자기 모니터링                           | 피드백의 해독과 활용                                       |
| 핵심 구성 요소 | 질의 개념 / 평가 기술 / 전략 저장고                 | 과제 준수 / 질 / 기준의 전유                                |
| 인식론      | 암묵적 지식·감식안의 체화                         | 핵심 개념의 내면화·작동화                                    |
| 함의       | 전문성 — 깊이·체화                            | 역량 — 보편성·발달 가능성                                   |
| 학생의 위치   | 형성평가의 주체                               | 피드백 사이클의 주체                                       |

Sadler의 논의는 이후 학생의 평가 역량 담론의 출발점이 되었다. 평가 전문성과 복합적 평가 역량 개념은 Tai 외(2018), Boud 외(2018)의 ‘평가적 판단(evaluative judgement)’ 개념과, Carless와 Boud(2018)의 ‘피드백 리터러시(feedback literacy)’ 개념으로 발전하

였다. 질에 대한 판단이 오랜 참여를 통해 형성되는 암묵적·체화적 능력이라면, 그것을 모든 학생에게 기대할 수 있는가에 대해서는 여전히 의문이 남는다. Sadler 역시 이 역량의 발달에 상당한 시간과 구조화된 경험이 필요함을 인정하며, 바로 이 때문에 그는 개별 피드백의 개선이 아니라 학생이 판단 경험을 축적할 수 있는 학습 구조의 설계가 필요함을 주장한다(Sadler, 2010).

## 5. 교사 판단의 사회적 정렬: 조정에서 교정으로

학생의 평가적 판단은 교사의 판단과 피드백에 의해 매개되므로, 학생 평가 역량의 문제는 필연적으로 그 원천인 교사 판단의 질과 일관성 문제로 되돌아온다. 판단이 암묵적 지식에 기반한 개인적 행위라면, 여러 교사가 같은 과목을 가르치고 평가할 때 교사마다 판단이 달라지는 문제를 피할 수 없다. Sadler의 후기 저작은 이 문제에 대한 응답이라 할 수 있다.

호주, 영국, 뉴질랜드와 같은 나라들은 이 문제에 대응하기 위해 교사들이 모여 평가에 대해 토론하고 합의하는 사회적 조정(social moderation)을 오랜 기간 발달시켜 왔다. Sadler(2013)는 사회적 조정의 의의를 인정하면서도 그 한계를 지적한다. 합의에 도달했다는 사실 자체가 그 합의의 타당성을 보증하지 않으며, 조정이 참조하는 표준이 집단 내부에 암묵적으로 머무는 한, 합의는 ‘우리끼리의 기준’으로 수렴할 수 있다는 것이다. 이는 1989년에 그가 교사와 학생 사이에서 제기한 “암묵적 지식은 공유되지 않으면 독점된다”는 문제를 교사 공동체의 수준에서 다시 제기한 것이기도 하다. 학생에게 길드 지식의 이양이 필요했듯이, 교사 공동체에는 판단의 참조점을 외부화하는 작업이 필요함을 주장한다.

Sadler(2013)는 사회적 조정에서 교정(calibration)으로의 전환을 제안한다. 교정이라는 용어는 저울의 영점을 맞추는 것처럼 측정 도구를 표준에 맞추는 작업을 뜻하며, 교사 개개인이 자신의 판단을 외재화된 표준(externalised standards)에 맞추어 정렬하는 과정이 교사 판단의 교정이다. 이는 교정된 예시(calibrated exemplars)와 해설(commentaries)로 구현될 수 있다. 교정된 예시는 합의되고 승인된 예시들의 모음으로, 하나의 학교나 교사 집단이 아니라 광범위한 전문가 검토를 거쳐 표준으로 인정된 예시들을 뜻한다. 해설은 예시들의 평가 결과에 대한 논거를 상세히 풀어쓴 것으로, 해설을 통해 표준이 단순한 결과가 아니라 추론 과정으로 드러나게 된다. Sadler는 교정을 통해 판단의 제어가 아니라 판단의 사회적 훈련을 지향한다. 루브릭의 세분화가 판단을 규칙 적용으로 대체하려는 시도라면, 교정은 판단을 판단인 채로 남겨두되 그것이 정련될 수 있는 공동의 참조 구조를 만드는 시도이다.

Sadler의 이론을 종합하면, 판단은 기준에 선행하고(1985), 형성평가는 그 판단 능력의 학생에의 이양이며(1989), 판단의 신뢰성은 도구가 아니라 공동체적 정렬에서 온다(2013). 암묵적 판단을 어떻게 신뢰할 수 있는가에 대한 그의 답은, 판단을 명시적 규칙으로 대체하는 것이 아니라 판단이 정련되는 개인적·사회적 조건을 만드는 것이라 볼 수 있다.

#### IV. 과정중심평가에의 함의

과정중심평가에 대한 함의를 논하기에 앞서, Sadler 이론의 한계를 확인할 필요가 있다. 첫째, Sadler 이론 자체의 한계이다. 질(quality)에 대한 판단이라는 평가의 정의와 그 기준은 여전히 모호하고, 결국 교사의 역량에 크게 의존해야 한다. 학생의 평가 역량 또한 모든 학생에게 기대하기에는 높은 수준이며, 무엇보다 학교 풍토나 평가의 문화가 우리나라의 맥락과는 차이가 있다. 둘째, 렌즈로서의 한계이다. Sadler의 관점은 과정중심평가의 모든 차원을 포착하지 못한다. 형성평가의 구체적 수업 전략, 평가의 학습문화적·사회문화적 조건, 학습 조절의 역동은 다른 이론가들의 몫으로 남으며, 특히 우리나라 맥락에서 성장 지원과 선발 기능 사이의 긴장, 그리고 대학입시 체제라는 제도적 조건(정정철, 2020)은 그의 이론으로 모두 설명하지 못한다. 이러한 한계들은 본 연구에서 다루지 않은 것이지, 과정중심평가에서 중요하지 않은 것이 아니다. 그럼에도 불구하고, 판단의 인식론의 차원에서 Sadler의 평가 개념들은 과정중심평가의 개념과 실행에 다음과 같은 시사점을 준다.

##### 1. 측정에서 판단으로

과정중심평가에서는 평가의 신뢰성과 공정성 확보에 대해 우려가 제기되고 있고(김유정 외, 2019; 박정, 2018) 교사에게는 높은 수준의 평가 전문성이 요구된다(우필희 외, 2021; 이정민, 2024). 과정중심평가의 실천에 있어 신뢰성을 확보하기 위해 루브릭의 정교화와 객관화가 강조되고 있다(이형빈, 2023; 이효녕, 이유정, 이현동, 2017; 홍소영, 2021).

Sadler에 따르면, 평가는 명시적 기준의 기계적 적용이 아니라, 질에 대한 전문가적 판단 행위이다(Sadler, 1985, 1989). 채점을 위한 루브릭은 판단의 도구이지, 판단을 대체하는 자동 장치가 아니다. 이형빈(2023)은 학교 현장에서 루브릭이 ‘채점 기준표’로만

인식되는 한계를 지적하며, 단순히 루브릭을 세분화하는 것이 아니라, 학생의 학습 과정과 성찰을 안내하는 도구로 사용해야 함을 강조한 바 있다. 이처럼 과정중심평가에서 루브릭의 정교화와 객관화에 대한 지나친 강조는 오히려 교사의 판단 행위를 위축시키고 평가의 본질을 놓치게 할 우려가 있다(Sadler, 1989). Sadler(1985, 1989)에게 평가의 본질은 학습의 질(quality), 즉 학습이 전체로서 의도한 목적을 달성하는 정도에 대한 인식적 판단이다. 평가의 본질에 대한 이러한 인식은 과정중심평가가 루브릭의 세분화나 잦은 기록으로 환원되는 현상에 대해 시사점을 준다.

Sadler의 연구물들에서 판단은 다양한 맥락으로 등장하며, 강조점이 조금씩 다르다. 첫째, 인식(recognition)의 행위로서의 판단은 질을 알아보는 행위이지, 규칙을 적용하는 행위가 아니다(Sadler, 1985). 인식은 추론이 아니며, 일차적인 평가적 행위이고, 어떤 기준에도 선행한다. 우리가 어떤 그림을 보고 ‘아름답다’고 알아볼 때, 그 알아봄은 기준을 적용한 결과가 아니라 기준에 선행하는 일차적 행위이다. 평가에서 일어나는 일이 본질적으로 이와 같다는 것이 Sadler의 판단에 대한 일차적 개념이다. 둘째, 전체적 행위로서의 판단(configurational judgement)은 부분들의 합이 아닌 형상적 통일체로 파악하는 행위이다(Sadler, 1989). 이는 게슈탈트(Gestalt) 심리학의 통찰을 평가 이론에 적용한 것으로, 질은 각 기준에 대한 판단을 합산해서 도출되는 것이 아니라, 전체를 하나의 통합된 형상으로 파악하는 행위이다(Sadler, 2009). 셋째, 암묵적 지식(tacit knowledge)으로 체화된 행위로서의 판단은 명시적 규칙이나 절차로 완전히 풀어낼 수 없는 차원으로(Sadler, 1989), 감식안을 갖춘 사람의 지도 아래 평가 활동에 오랜 기간 참여하는 귀납적 과정을 통해 발달한다. Sadler에 따르면 경험 많은 교사는 학업의 질을 즉각적으로 알아보지만, 그 판단의 모든 근거를 언어로 풀어내지는 못한다. 이는 판단의 결함이 아니라 판단의 본성이다. 마지막으로 사회적 행위로서의 판단은 3절에서 다룬다.

과정중심평가는 표준화와 측정 중심의 평가와 대비되는 입장을 취하며 교사의 질적 판단을 강조하는 듯하지만, 실제로는 루브릭의 정교화, 잦은 기록 등 기법적인 측면으로 환원되며, 일부 시·도에서는 AI로 평가를 대체하기도 한다. Sadler의 판단 개념에 비추어 보면, 아무리 세분화된 루브릭도 그 적용은 결국 교사의 판단을 요구하며, 그 판단의 질에 따라 같은 루브릭이 다른 결과를 낳는다. 따라서 평가의 신뢰성은 도구의 정교화가 아니라 교사 판단력의 정련에서 비롯되는 것으로 재해석될 수 있다. 또한, 판단이 이루어지더라도 판단의 근거를 모두 언어로 풀어내는 것이 어렵다는 점은 잦은 기록이 판단의 본질에 비추어 바람직한 실천 대안인지 의문을 품게 한다.

과정중심평가가 강조하는 교육과정-수업-평가의 일체화는 평가가 학습으로 이어져야 한다는 점을 강조한다(김영실, 2022). 평가가 학습에 도움이 되기 위해서는 학습의 주체인 학생이 ‘학습의 질’에 대한 개념을 형성해야 하는데, 이 개념은 단순히 성취기준을

아는 것이 아니라 질에 대한 판단의 경험을 통해 형성된다. 교사가 자신의 평가 행위를 판단의 행위로 이해하지 못하면, 학생에게 전달되는 것은 루브릭에 맞추는 법이지 질을 분별하는 안목이 아니다. 이러한 질에 대한 판단은 교사 전문성의 본질에 대해 다시 생각하게 한다. Sadler의 관점에서 보면, 과정중심평가가 교사에게 요구하는 높은 평가 전문성은 도구를 정확히 적용하는 기능적 능력이 아니라, 질을 분별하고 그 판단을 학생의 학습으로 전환하는 통합적 전문성으로 이해될 수 있다.

## 2. 피드백의 재정의

학교 현장에서 피드백은 결과물에 간단한 메모를 덧붙이거나 결과를 통보하는 방식으로 이루어지며, 수행의 질에 관한 정보는 빈약하다(김영실, 2022). 이는 과정중심평가 이후 점수의 단순 안내를 넘어 코멘트 제공이 주요 평가 활동으로 자리잡았음에도(김영실, 2022), 피드백이 여전히 전달 수준에 머물러 있음을 보여준다.

Sadler의 관점에서 평가가 학습으로 이어지는가를 가르는 것은 평가의 시기가 아니라 판단이 어떻게 쓰이는가이며, 같은 자료라도 격차를 줄이는 데 사용되면 형성적(formative)이고, 단지 기록과 보고에 그치면 그렇지 않다(Sadler, 1989). 이에 비추어 보면, 수업 ‘중간’에 평가를 실시하였더라도 피드백이 기록과 통보로 사용되면 평가가 형성적 기능을 충분히 수행하지 못하고 있는 것이다.

Sadler는 그의 2010년 논문 「피드백을 넘어: 복합적 평가 역량 발달시키기」에서 말해주기(telling)식의 피드백을 비판하고, 피드백을 ‘결과에 대한 정보 전달’이 아니라 ‘학생의 평가적 판단력 형성을 위한 교수 행위’로 재정의한다(Sadler, 2010). 피드백은 결과에 대한 정보 그 자체가 아니라 격차를 변화시키는 데(gap-closing) 사용되는 정보일 때만 피드백이다(Sadler, 1989). 전달(telling)만 하는 피드백은 아무리 정교하더라도 학생이 의미를 이해하지 못하거나, 수정할 부분을 식별할 배경지식을 갖추지 못하면 작동하지 않기 때문이다. 따라서 피드백이 제대로 작동하기 위해서는 학생의 평가 역량이 요구된다. III장에서 언급하였듯이, 피드백에 의해 학습이 일어나기 위해서는 세 가지 조건이 동시에 작동하여야 한다. 세 가지 조건은 도달하려는 기준에 대한 개념, 현재 수행을 기준과 비교, 격차를 줄이는 행동에 착수하는 것이다. 이처럼 Sadler는 평가에 학생을 주체로 끌어들인다. 피드백의 성립 여부는 교사의 제공 행위가 아니라 학습자의 활용 여부에 달려 있기 때문이다.

과정중심평가에서 피드백이 단순 메모나 통보로 실행되는 것이 학습이나 성장으로 이어질 것인지는 의문으로 남는다. 학습의 의지와 역량이 있는 어떤 학생은 교사가 남긴 메모를 읽고 다음 학습에 착수하겠지만 그렇지 않은 대다수의 학생은 평가 결과를 학습

으로 연결짓지 못하기 때문이다. 따라서 Sadler는 피드백을 학습의 구조(system)로 설계해야 함을 강조한다(Sadler, 2010). 형성적 학습 구조에서는 학생이 스스로의 학습을 점검할 수 있는 능력을 형성하는 데에 보다 초점을 둔다. 과정중심평가에서는 교사의 평가 전문성은 강조되지만(김유정 외, 2019; 박정, 2018; 우필희 외, 2021; 이정민, 2024) 학생은 여전히 수동적 존재로 위치지워진다. 국내에서도 학생을 평가의 주체로 보아야 한다는 논의가 제기되어 왔다. 박정(2019)은 자기평가를 학습 후의 단순한 반성이 아니라 학생이 자신의 수행 수준을 판단하고 목표와의 차이를 확인하는 형성적 학습 과정으로 보고, 자기평가 능력 자체가 길러야 할 역량임을 강조한 바 있다. 과정중심평가가 평가 행위를 질적인 판단과 학습과의 연결로 재개념화하기 위해서는 피드백의 목적과 기능을 재검토함과 동시에 학생을 평가 주체로 인식할 필요가 있다.

Sadler의 이러한 문제의식은 이후 학생 평가 역량 담론에서 학생 피드백 리터러시 개념으로 발전되었다. Carless와 Boud(2018)는 피드백의 효과가 교사가 제공하는 정보의 양과 질이 아니라 학생이 피드백 과정에 참여하는 방식에 달려 있다고 보고, 피드백의 가치 인식, 수행의 질에 대한 판단, 정서 관리, 실제 수정 행동의 네 요소로 이를 개념화하였다. 국내에서도 학생용 피드백 리터러시 척도가 개발되었고(박민애, 손원숙, 2019), 교사가 제공하는 피드백의 질과 교실의 피드백 환경이 학생의 피드백 리터러시 수준과 관련됨이 확인되는 등(박민애, 노현중, 2022) 관련 연구가 축적되고 있다. 이처럼 Sadler의 논의는 국제·국내 담론으로 이어지고 있으며, 피드백 리터러시가 학생 개인의 성향이 아니라 교사의 피드백 실천과 교실 환경 속에서 발달하는 역량이라는 국내 연구의 결과는, 피드백을 개별 정보의 전달이 아니라 학습 구조의 설계로 보아야 한다는 Sadler(2010)의 주장을 경험적으로 뒷받침한다.

### 3. 교사 공동체와 표준 정렬

과정중심평가는 수업 맥락에서 개별 교사의 판단에 의존하므로, 같은 결과물이라도 교사나 학교마다 다른 평가가 나올 수 있다는 신뢰성 문제가 제기된다. 이에 대해 학교 현장에서는 루브릭을 표준화, 정교화하거나 학년·교과협의회, 평가위원회를 통해 학교 내에서 합의를 거치는 과정으로 대응한다. 학교 단위의 합의는 그 학교 안에서만 통용되는 ‘내부 규약’에 머무르기 쉽다는 점에서 한계를 지닌다.

Sadler(2013)는 호주·영국·뉴질랜드에서 발달해 온 사회적 조정(social moderation)의 의미를 인정하면서도, 합의에 도달했다는 사실 자체가 그 합의의 타당성을 보증하지 않는다는 점을 지적하였다. 학교의 학년협의회나 평가위원회 운영도 이러한 성격을 띠기 쉽다는 점에서, Sadler의 비판은 우리나라 맥락에도 일정한 함의를 지닌다.

Sadler(2013)는 사회적 조정에서 교정(calibration)으로의 전환을 제안한다. 교정은 저울의 영점을 맞추는 것처럼, 교사 개개인의 판단을 외재화된 표준(externalised standards)에 정렬시키는 작업이다. 여기서 표준은 어느 한 학교나 교사 집단의 합의가 아니라, 광범위한 전문가 검토를 거쳐 공적으로 인정된 참조점을 뜻한다. 교정은 두 가지 도구를 통해 구체화된다. 먼저, 교정된 예시(calibrated exemplars)는 합의되고 승인된 학생 수행 사례들의 모음이다. 교정된 예시는 단순히 좋은 답을 보여주는 것이 아니라 ‘이 정도가 어떤 수준인지’의 참조점을 제공하는 데 목적이 있다. 해설(commentaries)은 예시가 왜 그 수준으로 판정되었는지의 논거를 상세히 풀어쓴 자료이다. 해설을 통해 표준은 결과가 아니라 추론 과정으로 드러나며, 교사가 자신의 판단을 그 추론 과정에 비추어 점검할 수 있게 된다.

Sadler의 교정 개념을 우리나라 맥락에 접목하면, 과정중심평가의 신뢰성을 확보할 수 있는 방안을 탐색할 수 있다. 첫째, 학교 내 합의를 넘어 시도 또는 전국 수준에서 교사들이 학생 수행을 함께 검토하고 정렬해 가는 평가 공동체의 구축이 필요하다. 호주의 외부 조정(external moderation) 사례나 일본의 단위종합평가 연구회 사례를 참고하여 교정된 예시와 해설집의 개발이 시도 단위로 이루어질 필요가 있다. 둘째, 교정된 예시와 해설의 개발은 현행 제도와 접속될 수 있다. 한국교육과정평가원이 성취평가제 지원을 위해 개발해 온 성취기준·성취수준 자료에 학생 수행의 실제 예시와 그 판단의 논거를 담은 해설을 결합한다면, 성취수준 기술문(예: A~E)이 추상적 진술을 넘어 ‘이 수준이 실제로 어떤 수행인지’를 보여주는 참조점으로 기능할 수 있다. 아울러 학교생활기록부의 교과학습발달상황 기재가 수행의 나열이 아니라 질에 대한 판단의 논거를 담는 방향으로 개선된다면, 기록 자체가 해설(commentary)의 성격을 지니게 되어 잦은 기록의 부담을 판단의 정련 기회로 전환하는 계기가 될 수 있다. 이는 ‘평가 도구의 정교화’와는 다른 방향의 작업이며, 교사의 판단을 대체하는 것이 아니라 판단의 참조점을 제공하는 것이다. 셋째, 활발히 운영되고 있는 교사 전문적 학습공동체의 기능을 조정할 필요가 있다. 현재 교사 전문적 학습공동체는 수업 나눔이나 교육과정 재구성에 초점을 두는 경우가 많은데, 평가 영역에서 교정된 예시를 함께 검토하고 자신의 판단을 정렬해 가는 작업이 핵심 활동으로 포함될 필요가 있다. Sadler(2013)가 말한 교정은 일회적 사건이 아니라 지속적 정렬 과정이기 때문이다. 다만 이러한 제안이 현행 제도 안에서 실현되기 위해서는 대학입시 중심의 평가 체제, NEIS 기록 구조와의 정합성 등 추가적인 검토가 필요하며, 여기서는 방향만을 제시하는 한계가 있다.

## V. 결론 및 제언

본 연구는 Sadler의 형성평가론과 평가적 판단 개념을 바탕으로 과정중심평가에 대한 시사점을 탐색하고자 하였다. Sadler의 이론이 하나의 이론적 렌즈로, 과정중심평가에 대한 유일한 해석을 주장하는 것이 아님을 전제하면서, 연구의 과정과 결과를 통해 다음과 같은 결론에 도달하였다.

첫째, 과정중심평가의 개념은 질적 판단의 행위로 재해석될 수 있다. 과정중심평가는 그간 정책 용어로 통용되어 왔으나 그것이 전제하는 평가관은 암묵적인 채로 남아 있었다. 본 연구는 Sadler의 이론에 비추어 과정중심평가의 본질이 사전에 명시된 기준의 기계적 적용이 아니라, 학생 수행의 질에 대한 환원 불가능한 판단 행위로 이해될 수 있음을 논의하였다. 이는 과정중심평가의 신뢰성이 루브릭의 정교화나 평가 도구의 객관화에서 오는 것이 아니라, 교사의 판단력의 정련에서 비롯됨을 시사한다. Sadler의 관점에서 보면, 과정중심평가의 개념은 ‘과정에서 이루어지는 평가’ 또는 ‘과정을 평가하는 평가’의 시점·대상 차원을 넘어, 질에 대한 전문적 판단의 행위로 재해석될 여지가 있다.

둘째, 피드백은 결과 전달이 아닌 학습 구조의 설계로 재정의될 필요가 있다. 현재 과정중심평가에서 피드백은 결과의 통보나 기록에 머무르는 경향이 있다(김영실, 2022). Sadler(1989, 2010)에 따르면 피드백은 실제 수준과 기준 수준 사이의 격차를 줄이는데 사용될 때만 비로소 피드백이며, 단순한 전달은 ‘허공에 매달린 자료(dangling data)’에 불과하다. 이에 비추어 보면, 과정중심평가의 피드백은 자아가 아닌 과제를 겨냥해야 하며, 질에 대한 인식·평가적 기술·수정 전략의 세 조건이 동시에 작동하는 학습 구조로 설계될 때 형성적 기능을 온전히 수행할 수 있다. 이는 피드백을 정보 제공의 차원에서 교수 행위의 차원으로 끌어올리는 작업이다.

셋째, 교사 전문성은 측정 기능이 아닌 평가판단역량으로 재개념화될 수 있다. 지금까지 교사의 평가 전문성은 평가 도구의 개발과 시행에 필요한 지식·기능을 중심으로 이해되어 왔다(AFT, NCME, & NEA, 1990; Plake et al., 1993). 그러나 과정중심평가는 수업 맥락의 다양한 변인이 상호작용하는 가운데 교사가 학생 수행의 질을 분별하고, 그 판단을 학생의 학습으로 전환할 것을 요구한다. 이때 요청되는 것은 측정의 기능이 아니라 질을 알아보는 감식안과 그것을 학생의 학습으로 매개하는 평가판단역량이다. 다만 이러한 판단은 개인의 직관에 머무를 경우 ‘우리끼리의 기준’으로 수렴할 위험이 있다(Sadler, 2013). 따라서 교사의 평가판단역량은 개인적 정련의 차원과 함께, 교사 공동체가 외재화된 표준에 자신의 판단을 정렬해 가는 교정(calibration)의 차원에서 함께 길러

져야 한다. 교정된 예시와 해설을 매개로 한 공동체적 정렬은 평가판단역량의 사회적 조건이자, Sadler가 제시하는 평가의 신뢰성 확보의 본질적 경로이다.

넷째, 학생은 평가의 객체가 아닌 평가적 판단의 주체로 재위치될 필요가 있다. 과정중심평가에서 교사의 평가 전문성은 강조되어 왔으나, 학생은 여전히 평가의 객체로 위치지어져 왔다. Sadler(1989, 2010)가 강조하듯, 교사의 피드백이 아무리 정교하게 제공되더라도 학생이 그것을 해독하고 자기 작업에 연결하지 못하면 학습은 일어나지 않는다. 학생은 질의 개념을 형성하고, 자신의 수행을 기준에 비추어 점검하며, 격차를 줄이는 전략을 사용할 수 있어야 하며, 이러한 복합적 평가 역량을 갖춘 평가적 판단의 주체로 재위치될 필요가 있다. 이는 과정중심평가가 교육과정-수업-평가의 일체화를 넘어 ‘학습-평가의 일체화’로 심화되는 길이기도 하다.

이러한 논의를 바탕으로 후속 연구와 정책, 교사교육의 방향에 관하여 다음과 같이 제언하고자 한다. 첫째, 과정중심평가에 관한 논의를 이론적으로 확장할 필요가 있다. 본 연구는 Sadler의 논의를 중심 자료로 삼았으나, 이는 판단의 인식론이라는 하나의 축을 따라간 독해이며, 과정중심평가의 철학적·이론적 토대를 구축하기 위해서는 보다 다양한 평가 이론과의 대화가 필요하다. Shepard(2000)의 학습문화론, Allal(2013, 2020)의 학습 조절론, Gipps(1994, 1999)의 사회문화적 평가론에 관한 연구, 평가 조정에서 교사 판단의 중심성을 분석한 Wyatt-Smith 외(2010)의 연구를 검토하는 작업이 후속 논의에서 이루어질 필요가 있다. 아울러 학생의 피드백 리터러시가 교사의 피드백 실천 및 교실 환경과 상호작용하며 발달한다는 점(박민애, 노현중, 2022)을 고려할 때, 교사의 평가판단역량과 학생의 피드백 리터러시가 교실에서 어떻게 함께 형성되는지를 탐구하는 연구도 요청된다.

둘째, 과정중심평가에 대한 경험적 탐구가 보강될 필요가 있다. 본 연구는 과정중심평가의 의미를 재해석한 이론적 연구로, 경험적·실증적 검증이나 현장의 사례를 다루지 못하였다. 따라서 본 연구에서 제안한 교사의 평가판단역량 개념을 경험적으로 탐색하는 후속 연구가 요청된다. 자기보고식 설문이 교사의 실제 수행을 정확히 반영하지 못한다는 선행연구의 지적(김나영, 2023)을 고려할 때, 교사가 학생 수행의 질을 판단하는 실제 과정을 사례 연구, 인지 면담, 평가 사례 분석 등의 질적 방법으로 탐구하는 작업이 필요하다.

셋째, 과정중심평가 실천을 위한 정책적 지원 방안을 탐색할 필요가 있다. 과정중심평가가 결국 교사의 평가 전문성을 기반으로 한 평가 자율성을 요청한다는 점에서, 교사의 자율적인 평가 실천을 제도적으로 보장하는 방향의 정책 설계가 요청된다. 교사가 교육과정과 수업에서 자율권을 행사하면서도 평가에서는 표준화된 절차에 종속되는 현 상황(반재천 외, 2018; 양연동, 2020)은 교육과정-수업-평가 일체화의 취지와 부합하지 않

는다. 특히 성장 지원과 선발 기능의 긴장(정정철, 2020)은 본 연구에서는 본격적으로 다루지 않았으나 실행의 성패를 좌우하는 조건이므로, 이 긴장을 정면으로 다루는 연구가 필요하다. 평가의 자율성과 평가의 신뢰성을 양립시키는 정책적 상상력이 필요하며, Sadler의 교정 개념은 그러한 상상력의 하나의 이론적 자원이 될 수 있으리라 기대한다.

## 참고문헌

- 교육부(2015). 초·중등학교 교육과정 총론. 교육부 고시 제2015-74호 [별책 1].
- 교육부(2022). 초·중등학교 교육과정 총론. 교육부 고시 제2022-33호 [별책 1].
- 교육부, 한국교육과정평가원(2017). **과정을 중심하는 수행평가 어떻게 할까요?** ORM 2017-19-2. 한국교육과정평가원.
- 김나영. (2023). 교사의 학생평가 역량 탐색: 교사 자기평가와 포트폴리오 평가 방식 간 차이를 중심으로. **교육학연구**, 61(7), 25-53.
- 김영실(2022). 초등학교 교사들의 과정중심평가 실천 방식과 수행 유형 분석. **교육과정평가연구**, 25(2), 139-166.
- 김유정, 이정건, 장원형, 홍훈기(2019). 교사의 과정 중심 평가 역량(T-PEC) 진단도구 개발 및 타당화. **한국교원교육연구**, 36(3), 99-121.
- 김정민(2018). 과정중심평가의 개념과 교육적 의의 탐색. **학습자중심교과교육연구**, 18(20), 839-859.
- 박민애, 노현중(2022). 교사가 제공하는 피드백 유형과 교실 피드백 환경 조성이 초등학생의 피드백 리터러시에 미치는 영향: 다층 잠재프로파일 분석의 적용. **교사교육연구**, 61(3), 293-310.
- 박민애, 손원숙(2019). 학생용 피드백 리터러시 척도(FLSS) 개발 및 타당화. **교육평가연구**, 32(3), 473-495.
- 박정(2017). 수업에서 학생 평가 의미 탐색. **교육평가연구**, 30(3), 397-413.
- 박정(2018). 형성평가와 평가의 객관성. **교육평가연구**, 31(3), 483-499.
- 박정(2019). 과정중심평가를 위한 수업과 학생 자기평가 의미 탐색. **교육평가연구**, 32(3), 421-440.
- 반재천, 김선, 박정, 김희경(2018). 교사별 과정 중심 평가에 대한 교사의 인식. **교육과정평가연구**, 21(3), 105-130.
- 성태제, 임현정(2014). 형성평가의 재인식에 따른 교사와 학교교육의 변화를 위한 제언. **교육평가연구**, 27(3), 597-615.
- 신혜진, 안소연, 김유원(2017). 과정 중심 평가 활용의 정책적 분석. **교육과정평가연구**, 20(2), 135-162.
- 양연동(2020). 초등교사의 교육평가 경험에 대한 자문화기술지. **교육사회학연구**, 30(1), 111-142.

- 우필희, 송명섭, 송주범(2021). 국내 교사 평가 전문성 기준과 측정도구 분석. **교육과정 평가연구**, 24(1), 173-201.
- 이정민(2024). 교사들의 자율권 실천에 관한 협력적 실행연구: 학생평가를 중심으로. **한국 교원교육연구**, 41(4), 251-280.
- 이형빈(2023). 성장중심평가의 취지에 따른 평가 루브릭 개발 가능성 탐구. **교육과정평가 연구**, 26(2), 259-278.
- 이효녕, 이유정, 이현동(2017). 과학 탐구 과정 평가를 위한 루브릭 개발 및 적용. **중등교육 연구**, 65(1), 145-171.
- 임은영(2019). 교사의 학생평가 역량 연구. **교육과정평가연구**, 22(1), 101-124.
- 임종헌, 최원석(2018). ‘과정 중심 평가’의 특징과 의미에 관한 연구: 자유학기제 ‘과정 중심 평가’를 중심으로. **한국교육**, 45(3), 31-59.
- 정정철(2020). ‘과정 중심 평가’의 교육적 의미와 난점 분석을 통해 본 교육평가의 지향점 탐구. **교육철학연구**, 42(4), 217-238.
- 최훈원(2024). 키워드 네트워크 분석을 활용한 과정중심평가의 연구동향 분석. **교육과정 평가연구**, 27(2), 251-280.
- 홍소영(2021). 교육과정-수업-평가의 일체화를 위한 통합교과 평가 루브릭 개발. **통합 교육과정연구**, 15(2), 113-141.
- 홍후조(2026). **알기 쉬운 교육과정**. 제3판. 학지사.
- Allal, L. (2013). Teachers’ professional judgement in assessment: A cognitive act and a socially situated practice. *Assessment in Education: Principles, Policy & Practice*, 20(1), 20-34.
- Allal, L. (2020). Assessment and the co-regulation of learning in the classroom. *Assessment in Education: Principles, Policy & Practice*, 27(4), 332-349.
- American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). Standards for teacher competence in educational assessment of students. *Educational Measurement: Issues and Practice*, 9(4), 30-32
- Black, P. & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in education*, 5(1), 7-74.

- Black, P. & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–148.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D.(2003). *Assessment for learning: Putting it into practice*. Open University Press.
- Boud, D., Ajjawi, R., Dawson, P., & Tai, J. (Eds.). (2018). *Developing evaluative judgement in higher education: Assessment for knowing and producing quality work*. Routledge.
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325.
- Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. Falmer Press.
- Gipps, C. V. (1999). Socio-cultural aspects of assessment. *Review of Research in Education*, 24, 355–392.
- Plake, B., Impara, J., & Fager, J.(1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice*, 12(4), 10–39.
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28(1), 4–13.
- Sadler, D. R. (1985). The origins and functions of evaluative criteria. *Educational Theory*, 35(3), 285–297.
- Sadler, D. R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education*, 13(2), 191–209.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144.
- Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy & Practice*, 5(1), 77–84.
- Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education*, 30(2), 175–194.

- Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education*, 34(2), 159-179.
- Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35(5), 535-550.
- Sadler, D. R. (2013). The futility of attempting to codify academic achievement standards. *Higher Education*, 67(3), 273-288.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.
- Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education*, 76(3), 467-481.
- Wyatt-Smith, C., Klenowski, V., & Gunn, S. (2010). The centrality of teachers' judgement practice in assessment: A study of standards in moderation. *Assessment in Education: Principles, Policy & Practice*, 17(1), 59-75.
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149-162.

논문접수 : 2026.7.2. / 수정본접수 : 2026.7.30. / 게재승인 : 2026.8.11.

ABSTRACT

## Reinterpreting Process-Centered Assessment Through Sadler's Theory of Formative Assessment: Meaning and Implications

Mijin Kim

Adjunct Professor, Daegu University

Process-Focused Assessment has become a central term in Korean assessment policy, yet the view of assessment and the epistemological assumptions underlying it have remained largely implicit. This study offers an interpretive reading of Process-Focused Assessment through the lens of D. R. Sadler's theory. With the literature review, this study tries to answer this question: how can teachers' qualitative judgement be trusted? From Sadler's perspective, Process-Focused Assessment can be reinterpreted as follows. First, assessment is an irreducible act of judgement about quality, and its trustworthiness derives from the refinement of teacher judgement rather than the elaboration of instruments. Second, feedback needs to be redefined as a pedagogical act that cultivates students' evaluative judgement rather than the transmission of results. Third, teachers' evaluative judgement competency should be developed through communal calibration against externalized standards as well as individual refinement. Fourth, students need to be repositioned as agents of evaluative judgement rather than objects of assessment. This study contributes by articulating the view of assessment implicitly presupposed by Process-Focused Assessment and by proposing one perspective that grounds its trustworthiness in the refinement of teacher judgement and the alignment of assessment communities.

*Key Words:* Process-Focused Assessment, Sadler, formative assessment, evaluative judgement competency, feedback, evaluative judgement