

## 올바른 평가 데이터의 수집

모든 교사들은 이미 오래 전부터 비공식적으로나마 형성 평가를 수행해 왔습니다. 형성 평가는 사람과 사물에 대한 판단을 형성하고자 하는 인간의 본성과 맞닿아 있습니다. 하지만 이러한 판단의 대부분은 무의식적으로 이루어지며, 많은 경우 잘못된 신념과 오해를 낳기도 합니다. 형성 평가를 통해 수집되는 데이터가 타당성(Validity)을 가지려면, 평가하고자 하는 대상이 명확해야 하고, 신뢰성이 있어야 하며, 수집된 정보는 재사용이 가능해야 합니다.

타당성 있는 평가는 능력, 전략, 지식 등의 평가 대상을 명확하게 규정하고 있습니다. 수학에서 문제 해결에 관련한 선다형(multiple-choice) 문제를 출제하는 것만으로는, 학생이 얼마나 뛰어난 수학 문제 해결 능력을 갖고 있는지 파악할 수 없습니다. 이런 문제를 맞힌 학생들은 문제 해결 전략을 잘 암기했거나 어렵짐작하는 능력이 뛰어나 지는 몰라도, 실제로 문제 해결이 필요한 상황에서 뛰어난 수행 능력을 보일 것이라 기대하기는 어렵습니다. 이러한 선다형 문제는 21세기 핵심 역량에 대해서는 타당성을 갖기 어렵습니다.

Stiggins(2004)는 이렇게 경고합니다. "우리는 평가의 정확성을 보장하는데 많은 투자를 하지 않았다. 이로 인해 모든 레벨에서 부정확한 평가, 그리고 비효율적인 판단의 가능성이 눈에 띄게 증가하고 있다"(p. 25). 교사가 제한적인 정보만을 가지고 학생의 지식과 능력을 성급하게 판단하는 경우, 교사의 판단은 학생의 성장을 독려하는 것이 아니라 오히려 방해할 수 있습니다.

Araison(2001)은 타당성에 대한 위협 요소를 이렇게 정의하고 있습니다:

1. 고정관념화(stereotyping): 개인적인 인상이나 기존의 편견을 바탕으로 결론을 도출
2. 논리적 오류(logical error): 주제와 무관한 특성을 바탕으로 학생의 능력을 평가(예: 옷차림, 다른 형태의 성취도를 기준으로 학생을 평가) (이러한 문제는 일반적으로 무의식적으로 나타나기 때문에 교사 자신이 인식하지 못하는 경우가 많습니다.)
3. 부적절한 샘플링(inadequate sampling): 한 가지 관찰 결과 또는 정보만을 바탕으로 평가를 수행
4. 일반화(generalizing): 학생이 특정 상황에서 행동한 결과를 바탕으로 다른 상황에서도 똑같이 행동할 것이라고 판단하는 것

학생의 학습 수행에 관련하여 수집되는 데이터는 또한 신뢰성(reliability)이 있어야 합니다. 신뢰성 있는 정보는 일관성(consistency) 있고 전형적(typical)이어야 합니다. 한 예로, 방학이 시작하기 바로 전날 수집한 학생의 사고 능력 평가 정보는, 학생의 행동이 전형적이지 않기(atypical) 때문에 신뢰성이 떨어질 가능성이 높습니다.

평가 데이터가 유용한 결론을 도출하는데 도움이 되려면, 데이터에 타당성, 신뢰성이 있어야 하고, 데이터가 중요하면서 일반적인 무엇인가를 증명하고 있어야만 합니다. 학자들은 데이터로부터 결론을 도출하기 위해 사용되는 프로세스를 위해 "삼각측정법(triangulation)"이라는 개념을 적용하고 있습니다. 신문 기자들이 범죄에 대한 취재 기사를 인쇄하기 전에 반드시 확인 과정을 거치는 것처럼, 교사들은 학생의 능력에 대한 결론을 내리기 전에 반드시 두 가지 이상의 정보를 확보해야 합니다. 또 설사 정보가 두 가지 이상 모였다 하더라도, 결론은 잠정적이어야 하며 이에 반대되는 데이터를 개방적으로 수용할 수 있어야 합니다. 한 가지 예를 들어 보겠습니다. 어떤 교사가 그룹 프로젝트와 학습 일지를 통해 어떤 학생의 일반화 능력이 부족한 것으로 판단합니다. 하지만 나중에, 이 학생은 다른 주제 영역에 대해서는 일반화를 효과적으로 수행함을 증명할 수 있습니다. 교사는 학생의 일반화 능력이 부족한 것처럼 보였던 것이 실제로는 (논리적 사고 능력이 부족한 때문이 아니라) 주제 영역에 대한 지식이 불충분했던 때문이라고 잠정적인 결론을 내릴 수 있습니다.

대부분의 교사들은 항상 기민한 태도로 학생들의 행동에 지속적인 관심을 보입니다. 교사들은 학생들이 어떤 행동을 하는지, 또 어떤 말을 하는지 주의 깊게 관찰하고 있습니다. 안타까운 사실은, 교사들 대부분이 이러한 비공식적 관찰이 형성 평가의 일부가 될 수 있음을 고려하지 않고 있으며, 관찰 결과를 체계적인 방법으로 기록해 두지 않는다는 사실입니다. 용의 주도한 분석이 동반되지 않은 상태에서 이러한 유형의 관찰이 수행되는 경우, 충분한 데이터가 고려되기 어렵기 때문에 왜곡된 관점과 잘못된 결론을 낼 수 있습니다. 체계적이지 않은 주먹구구 방식으로 수집된 데이터를 기반으로 수행되는 교습은 학생의 발전을 오히려 가로막을 수 있습니다. 형성 평가를 통해 학생의 데이터를 용의 주도하게 수집하고 검토하는 과정은 많은 시간과 준비 과정을 필요로 합니다. 하지만 형성 평가가 학생들에게 미치는 학습, 동기 유발 효과는 그러한 노력을 상쇄하고도 남음이 있습니다.