



8장. 평가기준의 설정 및 활용

Contents

- I** 평가기준에 대한 이해
- II** 평가기준의 표현 형태
- III** 보편적 평가기준
- IV** 평가기준의 설정 방법

I. 평가기준에 대한 이해

평가 기준

평가 준거에 대해 목표 달성 수준, 기대하는 변화의 정도, 의도하는 특성이 나타나는 정도, 변화 또는 성취를 나타내는 점수, 의도한 장점이 표현되는 맥락이나 상황의 수준, 원하는 가치가 산출되는 정도나 범위 등

- 특정 검사 점수나 성취도 점수 / 변화를 나타내는 인정할만한 업적이나 성취면에서의 구체적인 정도나 수준 / 질적 변화를 증거하는 특정 수준의 수치 / 만족도를 나타내는 특정 지표 등

I. 평가기준에 대한 이해

1 평가 기준과 평가 준거 간 관계

- 평가 기준은 그 자체로서는 특정한 의미가 없음
→ 평가 준거 및 평가 지표와의 관련에 의해서만 그 의미가 부여됨
- 평가 기준 설정 시,
이미 설정된 준거를 중시하고 어떤 준거에 대한 기준을 어떤 자료 및
지표를 이용하여 설정할 것인지 고려
- 하나의 평가 준거에 하나의 평가 기준이 있을 수 있지만,
평가 준거 하나에 여러 평가 기준이 있을 수 있으며,
여러 종류의 평가 준거에 대해 하나의 평가 기준만을 설정하여 활용할
수도 있음
→ 평가 준거의 성격과 내용 및 평가 자료에 따라 적절한 판단 필요

I. 평가기준에 대한 이해

2 평가 기준과 평가 지표 간 관계

- 평가 기준은 각 준거에 관련된 평가 자료를 중심으로 그 준거의 속성과 내용, 또는 그로 인한 산출 및 결과에 관하여 어느 정도 / 얼마 만큼이 바람직한 수준, 점수, 기준점이 되는가를 나타내는 것으로서 일반적으로 점수나 수준 및 정도 등으로 표현됨
- 평가 준거마다 하나 이상의 **평가 지표**가 수집되고 그에 따라서 **평가 기준**이 설정되는 바, 평가 기준은 평가 자료를 통해 표현되기 때문에 **평가자료를 구하기 전에 그에 대한 기준을 설정하기가 어려움**
- 평가 기준의 설정은 다른 평가 활동과 무관한 상태에서 별도로 이루어지는 일이 아니고 여러 관련 활동과 상호 관련 하에서만 이루어질 수 있음

I. 평가기준에 대한 이해

3 평가 기준의 특성

평가 준거와 평가 기준 간의 차이점

평가 준거	평가 기준
<목적적인 관점> '무엇'을 평가할 것인가 규명하며 평가 활동의 내용과 범위, 또는 방향을 제시	<수단적인 관점> 평가 준거를 '어떻게' 평가할 것인지를 의미하며 그 준거의 특정한 속성과 내용 및 측면에 관한 바람직한 범위, 수준, 정도 등을 규정

- 평가 기준은 '평가 준거의', '평가 준거에 근거하여', '평가 준거에 관하여' 바람직한 수준이나 정도, 상태, 질 등을 나타냄

II. 평가기준의 표현 형태

평가 기준

평가 준거의 성격이나 수집된 자료의 특성에 따라, 활용 목적과 상황에 따라, 활용 방법과 절차에 따라, 표현되는 형식에 따라 여러 형태로 표현되고 활용될 수 있음

- 1) 보편적 기준과 상황적 기준 (universal-situational standard)
- 2) 공통적 기준과 개별적 기준 (common-individual standard)
- 3) 절대적 기준과 상대적 기준 (absolute-relative standard)
- 4) 명시적 기준과 암시적 기준 (explicit-implicit standard)
- 5) 질적 기준과 양적 기준 (qualitative-quantitative standard)
- 6) 과정적 기준과 결과적 기준 (process-product standard)
- 7) 주관적 기준과 객관적 기준 (subjective-objective standard)
- 8) 내재적 기준과 외재적 기준 (intrinsic-extrinsic standard)
- 9) 최소한의 기준과 최대한의 기준 (minimum-maximum standard)
- 10) 이상적 기준과 현실적 기준 (ideal-practical standard)
- 11) 사전 설정 기준과 사후 설정 기준 (predetermined-postdetermined standard)

II. 평가기준의 표현 형태

1 보편적 기준과 상황적 기준

▪ 보편적 기준

- 일반적으로 인정할 수 있는 보편성을 중시하여 시간적, 공간적 차이에도 불구하고 적용될 수 있는 면을 규정하여 제시한 기준을 의미

▪ 상황적 기준

- 비교적 일시적인 상황을 중시하여 각 상황마다 강조되는 속성과 추구되는 가치를 순간적으로 정하여 그 상황에서만 적용하기 위하여 설정한 기준.
- 매상황마다 변화할 가능성이 많고 매우 시한적이고 순간적인 동시에 특수한 상황마다 달라질 수 있어 그 기준의 적용 범위가 매우 제한되어 있음

II. 평가기준의 표현 형태

1 보편적 기준과 상황적 기준

	보편적 기준	상황적 기준
장점	적용 범위가 광범위하고 시간적으로도 오랫동안 사용할 수 있음	각 상황의 조건만을 중시하여 순간적으로 설정하여 활용할 수 있음
단점	일반성, 보편성, 항구성, 공통성 등을 중시하여야 하기 때문에 설정 자체가 매우 어려움	매상황마다 새롭게 설정해야 할 가능성이 많기 때문에 그 적용 시한이 매우 짧고 적용 범위가 매우 좁음

II. 평가기준의 표현 형태

2

공통적 기준과 개별적 기준

▪ 공통적 기준

- 다양한 프로그램 평가 상황이나 여러 가지 평가 준거에 공통적으로 적용될 수 있는 기준으로서 프로그램의 목적/내용/대상자가 유사한 경우, 평가 준거의 내용이 유사한 경우에 공통적으로 사용할 수 있는 평가 기준

▪ 개별적 기준

- 각 평가준거의 성격에 따라, 각 평가 상황에 따라 각기 다르게 설정하여 각 상황이나 준거에 따라 각기 다르게 활용하는 경우의 기준을 의미

II. 평가기준의 표현 형태

3

절대적 기준과 상대적 기준

■ 절대적 기준

- 시대적 상황이나 사회 문화적 조건을 막론하고 보편적으로 추구하는 절대적인 선이나 가치를 프로그램이 추구하는 경우에는 바람직한 수준 및 정도를 절대적 기준이라 할 수 있음

예) 대중의 생명과 재산문제를 다루는 직종의 자격 부여 여부를 판단하기 위하여 설정한 수준 및 정도

■ 상대적 기준

- 특정 프로그램 목표나 준거에 있어서 추구하고 달성해야 할 수준이나 정도를 프로그램이 적용되는 상황이나 조건, 또는 대상자들의 능력 및 태도 등을 고려하여 비교적 유동적으로 결정하여 활용하는 평가 기준

예) 학업성취도를 평가하고자 할 때 규준집단의 평균적인 성취나 도달 정도에 따라서 그 기준이 달라짐

II. 평가기준의 표현 형태

4

명시적 기준과 암시적 기준

▪ 명시적 기준

- 특정 영역이나 준거에서의 구체적이고 가시적인 결과를 명료하게 수치화하고 행동적 용어로 진술하여 제시하거나 표현한 성취 수준 또는 정도를 의미
- 구체적인 결과를 제시하거나 행동 용어로 제시한 특정 점수나 두드러진 성취 결과를 말하는 바, 비교적 구체적이고 가시적인 측정 및 관찰 가능한 성취 및 행동 결과를 이용하여 특정 준거에 관한 성취하고 달성해야 할 바람직한 수준이나 정도를 나타냄

▪ 암시적 기준

- 표현 용어가 비교적 일반적이고 추상적인 경우를 말함
- 프로그램의 여러 목표 중에서 비교적 포괄적이고 함축적인 용어로 표현된 목표로부터 평가 준거를 추출하고 그에 대한 평가기준을 설정할 경우에 활용됨

II. 평가기준의 표현 형태

4

명시적 기준과 암시적 기준

	명시적 기준	암시적 기준
장점	비교적 측정 및 관찰하기가 용이하고, 이를 이용하여 판단하기가 용이하므로 결과의 신뢰성이나 객관성 을 인정받을 수 있음	암시적 기준을 사용한 판단 이 비교적 융통성 있게 이루어질 수 있음
단점	기준이 제한적이고 행동적인 측면만을 표현하므로 사용 한계가 좁을 수 있음	측정 및 관찰이 어렵고 입증 자료 수집이 곤란하며, 결과에 대한 객관성이나 신뢰성을 보장받기 어려움

II. 평가기준의 표현 형태

5

질적 기준과 양적 기준

▪ 질적 기준

- 특정 영역에서 달성 및 성취해야 할 수준이나 정도를 기술 및 묘사한 형태의 기준을 의미

▪ 양적 기준

- 특정 영역에서 달성 및 성취 수준을 수량화하여 표현한 경우를 의미
- 평가 기준을 행동적인 용어로 표현하거나 관찰 및 측정이 용이한 경우에는 대체로 양적인 기준을 사용하는 경우가 많으며, 측정 및 관찰 결과를 수량화할 수 있는 경우에 한해서만 사용 가능함

II. 평가기준의 표현 형태

5 질적 기준과 양적 기준

	질적 기준	양적 기준
장점	상세하게 여러 관점을 표현할 수 있어 보다 타당한 평가적 판단 이 가능함	객관성 및 신뢰성 을 보장받기 용이하며, 사용이 간편함
단점	객관성 이나 신뢰성 을 보장받기가 어려움	기준이 의미하는 바가 매우 제한적 이고, 그 기능이 매우 단순함

II. 평가기준의 표현 형태

6

과정적 기준과 결과적 기준

■ 과정적 기준

- 특정 프로그램 목표를 추구하는 절차 과정이나 그 수행 과정의 바람직한 상태나 수준을 나타낸 것으로서, 프로그램을 개발하거나 개선하기 위한 목적으로 평가하는 경우이나 형성 평가의 기능 강조하는 상황에 사용함

예) 교육평가의 경우, 수학문제를 푼 결과로서 정답여부보다는 푸는 과정의 적절성을 중시한다든가 실험하는 과정 및 계열상의 적합성을 강조한 관점에서 과정적 기준이 필요함

■ 결과적 기준

- 프로그램 종료 후에 실시 효과나 그 목적 달성 정도를 평가하기 위한 총합평가 기능을 강조하는 상황에서 주로 사용함
- 프로그램 종료 후 각 준거별로 어느 정도로 목적을 달성하였고 어떤 효과를 얻게 되었는가에 초점을 두고 프로그램의 산출 결과면에서의 특정 수준과 정도를 제시하거나 표현

II. 평가기준의 표현 형태

7

주관적 기준과 객관적 기준

■ 주관적 기준

- 평가적 판단 과정에서 평가자의 주관적 판단에 중점을 두고 특정 범위, 수준, 정도 등을 확인하거나 인식하여 그를 인정하고 수용함으로써 그 기준이 제 기능을 하게 되는 경우의 기준을 말함
- 특정 영역과 속성에 있어서 표현하고 기술하기 어렵거나 관찰 및 측정하기가 곤란한 속성의 정도, 범위, 수준 등을 판단자와 평정자의 주관적인 인식과 수용을 통해서만 기준으로 설정하고 활용할 수 있는 경우의 기준을 뜻함
- 주로 예술적인 표현 및 수행 영역, 복합적이고 난해한 정의적 영역, 순간적인 동작 및 수행 영역 등에서의 특정 수준이나 정도를 기준으로 설정하는 경우에 활용됨

II. 평가기준의 표현 형태

7

주관적 기준과 객관적 기준

■ 객관적 기준

- 프로그램 목표 영역 또는 실천 과정에서의 특정 수준이나 정도를 비교적 객관적인 형태로 표현하고 제시하며 다수가 용이하게 동의할 수 있는 경우의 기준을 말함
- 일반적으로 객관적인 기준은 수량화하거나 다수가 인식하고 실천하기 용이한 명시적이며 행동적인 용어로 표현하므로 양적인 기준 형식, 명시적인 기준 형식, 외재적인 기준 형식을 취함

Ⅱ. 평가기준의 표현 형태

8

내재적 기준과 외재적 기준

▪ 내재적 기준

- 프로그램의 목표 및 내용을 중심으로 **프로그램 자체**에 관련된 특정 수준이나 정도를 규정한 기준
- 프로그램의 질적 개선 및 프로그램의 개발을 목적으로 한 경우에 사용됨

▪ 외재적 기준

- 프로그램으로 인한 효과나 영향에 관한 측면과 프로그램이 적용되는 사회 문화적 측면에 관한 특정 수준과 범위 등을 규정한 기준
- 프로그램 관련인의 요구와 기대를 그들의 가치와 의미를 중시하여 프로그램의 결과 및 효과에 비추어 규정함

II. 평가기준의 표현 형태

9

최소한의 기준과 최대한의 기준

■ 최소한의 기준

- 특정 수준의 성취나 능력 수준이 사회에서 요구하는 사회화의 필수 요건이 되는 경우나 대중의 인권, 재산, 생명 등에 관련된 중차대한 내용과 관련된 경우 사용됨

예) 전문직종의 자격 취득을 인정하는 기준 점수, 특정 제품이나 상품의 품질을 관리하기 위한 목적으로 설정한 실수율이나 폐기율 등

■ 최대한의 기준

- 특정 영역이나 내용에 관한 성취 또는 산출에 관한 최대한의 수준이나 정도를 표현한 기준. 일반인들이 이상시하는 상태나 매우 탁월하여 모두가 귀감으로 삼을만한 점을 강조하고 모두가 지향해야 할 점을 부각시키기 위해 사용함

예) 생산 현장에서의 생산량 증대를 위해 지향하는 특정 생산량 정도나 성취 수준, 특정영역에서 일반인들로부터 존경 받고 그 업적을 기리며 교육적 귀감으로 삼을 만한 인물을 결정하기 위한 경우 등

Ⅱ. 평가기준의 표현 형태

10 이상적 기준과 현실적 기준

■ 이상적 기준

- 실제로는 실현하거나 달성하기가 어려운 수준이나 상태를 제시하지만, 특정 프로그램을 개발하거나 개선하기 위해 이상적 기준을 하나의 지향점이나 이상적으로 삼아 보다 질적으로 우수하게 만들기 위해 사용됨
- 특히, 교육 현장에서 학생들에게 교육적인 가치를 중시하여 이상적인 학생상을 강조하기 위해 이상적 기준을 활용함

■ 현실적 기준

- 프로그램 개발 당시, 의도하는 목표 수준에 비해 성취된 목표 수준이 만족스럽지 못할 경우, 실제로 획득하고 성취한 목표 수준을 중시하고 프로그램 실천과 작동 과정상의 상황을 참작하여 현실적 기준을 설정함
- 프로그램 실시 과정 중의 관련 상황을 고려한 현실성을 중시하여 실질적으로 실현 가능하고 성취 가능한 수준이나 정도를 평가 기준으로 설정함

II. 평가기준의 표현 형태

11 사전 설정 기준과 사후 설정 기준

- 평가기준 설정 시기를 평가활동이 있기 전이냐 그 이후이냐를 중심으로 사전/사후 설정 기준을 구분할 수 있음
- **사전 설정 기준**
 - 실제 평가기준을 평가를 계획하고 준비하는 단계에서 결정해 놓고, 본격적인 평가자료 수집활동을 임할 때 미리 설정해 놓은 기준을 뜻함
- **사후 설정 기준**
 - 평가계획에 의거하여 종합처리 한 후, 적절한 수준의 기준을 설정하고 평가적 판단에 이를 때, 상황에 따라 적절한 기준을 설정하는 것을 뜻함

Ⅲ. 보편적 평가기준 : 평가수행에 관한 규범

평가 목적을 중시하여 교육프로그램을 평가하는 경우,
이상적이고 보편적인 기준 형식으로 표현하여 제시하고 있는 두 종류의
기준이 있음

- ▶ '교육 평가 기준 합동위원회'의 30개 기준

- ▶ '평가 기준 연구회'의 55개 기준

Ⅲ. 보편적 평가기준 : 평가수행에 관한 규범

1 합동위회의 4개 분야 30개 기준

1) 유용성 기준

- 이 부류에 해당되는 기준은 프로그램 평가를 통하여 필요한 정보를 제공하며 평가가 시기적절하고 영향력이 있는 것이 되도록 설정한 것임
- 평가자로 하여금 평가 관련자 및 수혜자를 확인하고 그들과 친숙하게 되며 그들이 요구하는 정보가 무엇인지를 파악하며, 그 정보에 대한 요구에 부응하고 나아가서 적절한 정보를 적시에 명료하게 보고할 수 있도록 요구하는 내용들을 규정하고 있음
- 유용성 기준은 평가가 보다 유익하고, 시의 적절하고 영향력이 있도록 안내하는 역할을 수행함

Ⅲ. 보편적 평가기준 : 평가수행에 관한 규범

1 합동위회의 4개 분야 30개 기준

1) 유용성 기준

- **U1 이해관계자의 확인(Stakeholder Identification)**

평가에 관여하거나 영향 받는 사람들을 확인해야만 그들의 요구를 제대로 파악할 수 있음

- **U2 평가자의 신임도(Credibility)**

평가를 수행하는 사람들은 평가를 수행하는데 있어 신뢰할 수 있어야 하고 유능해야 함. 그래야만 신뢰할 수 있는 평가자에 의한 평가 결과의 신임도와 수용도를 최대화 시킬 수 있음

- **U3 정보의 범위와 선택(Information Scope and Selection)**

수집된 정보는 프로그램에 관하여 적절한 문제들을 다룰 수 있도록 폭넓게 선택되어야 하고, 평가의뢰인과 지정된 이해관계자들의 이익과 요구를 반영해야 함

Ⅲ. 보편적 평가기준 : 평가수행에 관한 규범

1 합동위회의 4개 분야 30개 기준

1) 유용성 기준

- **U4 가치의 확인(Values Identification)**

결과를 해석하는 데 사용된 관점, 절차, 논리적 근거가 주의 깊게 기술되어 질 때, 가치 판단에 대한 근거가 명백해짐

- **U5 보고서의 명료성(Report Clarity)**

평가 보고서는 문맥, 목적, 절차, 평가의 결과를 포함해서 평가 대상 프로그램을 명료하게 기술해야 함. 그렇게 함으로써 핵심적인 정보가 제공되고 쉽게 이해할 수 있음

- **U6 보고서의 적시성 및 배포(Report Timeliness and Dissemination)**

유의미한 잠정적 결과와 평가 보고서가 예상되는 평가 관련자들에게 적시에 배포되어야만 그 결과 보고서를 유용하게 활용할 수 있음

- **U7 평가의 영향(Evaluation Impact)**

평가는 이해관계자들의 입장에서 결과를 활용하는 데 중점을 두고 계획되고 수행되며 보고되어야 함

Ⅲ. 보편적 평가기준 : 평가수행에 관한 규범

1

합동위원회의 4개 분야 30개 기준

2) 실행가능성 기준

- 평가자로 하여금 실험실과 같은 조작적인 분위기가 아닌 자연스러운 상황에서 귀중한 자원과 시간을 절약하며 평가를 실시해야만 보다 현실적인 평가가 될 수 있다는 관점에 따라 설정된 것

Ⅲ. 보편적 평가기준 : 평가수행에 관한 규범

1

합동위회의 4개 분야 30개 기준

2) 실행가능성 기준

- **F1 실제적 절차(Practical Procedures)**

필요한 정보를 얻는 동안, 혼란을 최소화하기 위하여 평가절차는 실제적이어야 함

- **F2 정치적 실행가능성(Political Viability)**

평가는 여러 이해 집단의 서로 상이한 입장을 예상하여 계획되고 관리되어야 함. 그러면 그들의 협력을 이루어낼 수 있을 것이며, 또한 이해관계자 집단 중 누군가에 의해 가능한 도전들을 피할 수 있게 됨

- **F3 비용 효과(Cost Effectiveness)**

평가는 효율적이고 충분한 가치가 있는 정보를 산출하여 자원의 사용을 정당화해야 함

Ⅲ. 보편적 평가기준 : 평가수행에 관한 규범

1 합동위원회의 4개 분야 30개 기준

3) 정당성 기준

- 평가활동이 평가 관련자와 평가 수혜자의 인권을 무시하고 불법적이고 부주의하며 비윤리적으로 이루어지지 않도록 방지하기 위해 설정된 기준
- 평가자들의 불법적, 비양심적, 비윤리적인 행위에 대한 경고와 민감성을 키우기 위한 것이며, 평가자들이 사생활, 정보의 사용권, 인권 보호와 같은 문제에 대한 법을 이해하도록 함

Ⅲ. 보편적 평가기준 : 평가수행에 관한 규범

1

합동위회의 4개 분야 30개 기준

3) 정당성 기준

- P1 서비스 지향(Service Orientation)
- P2 공식적 합의(Formal Agreement)
- P3 피험자의 권리(Rights of Human Subjects)
- P4 인간적 상호작용(Human Interactions)
- P5 완전하고 공정한 사정(Complete and Fair Assessment)
- P6 결과의 발표(Disclosure of Findings)
- P7 이해관계의 갈등(Conflict of Interest)
- P8 경제적, 재정적 책임(Fiscal Responsibility)

Ⅲ. 보편적 평가기준 : 평가수행에 관한 규범

1 합동위회의 4개 분야 30개 기준

4) 정확성 기준

- 평가자로 하여금 프로그램의 장점과 가치를 판단하기 위하여 필요한 보다 타당하고 신뢰로운 정보와 자료를 수집하도록 요구하는 내용에 관한 것
- 평가 결과로서 보다 정확하고 적절한 정보를 산출하기 위한 목적을 달성하기 위하여 평가자의 전문성을 중시한 체계적인 평가 절차와 과정에 초점을 두고 있음
 - A1 프로그램 문서화(Program Documentation)
 - A2 상황 분석(Context Analysis)
 - A3 기술된 목적 및 절차(Described Purposes and Procedures)
 - A4 변론 정보 출처(Defensible Information Source)
 - A5 타당한 정보(Valid Information)

Ⅲ. 보편적 평가기준 : 평가수행에 관한 규범

1

합동위회의 4개 분야 30개 기준

4) 정확성 기준

- A6 신뢰성 있는 정보(Reliable Information)
- A7 체계적 정보(Systematic Information)
- A8 양적 정보 분석(Analysis of Quantitative Information)
- A9 질적 정보 분석(Analysis of Qualitative Information)
- A10 정당한 결론(Justified Conclusion)
- A11 공명 정대한 보고(Impartial Reporting)
- A12 메타평가(Metaevaluation)

Ⅲ. 보편적 평가기준 : 평가수행에 관한 규범

2

평가연구회 6개 분야 55개 기준

- 1) 문제 설정과 협상 (12개 기준)
- 2) 구조와 설계 (6개 기준)
- 3) 자료 수집과 준비 (12개 기준)
- 4) 자료 분석과 해석 (9개 기준)
- 5) 의사소통과 발표 (10개 기준)
- 6) 결과의 활용 (6개 기준)

Ⅲ. 보편적 평가기준 : 평가수행에 관한 규범

3 기타 평가기준

1) AEA의 평가자를 위한 규준적 원칙

- AEA(American Evaluation Association)에서 평가수행에 바람직한 방향을 알려줄 규준적 원칙을 개발하였음
- 5개 영역 23개의 항목으로 구성되어있음
 - **체계적 탐구 영역** (3개 항목) : 평가자는 평가대상의 성격막론하고 근거자료에 입각하여 수행해야 한다는 원칙
 - **적합한 능력함양 영역** (4개 항목) : 평가자는 평가이해관련자들에게 자신있는 평가수행을 제공해야 한다는 원칙
 - **통합/정직성 영역** (7개 항목) : 평가자가 평가 전반에 걸친 과정을 정직하고 통합적으로 수행해야 한다는 원칙
 - **인간존중 영역** (6개 항목) : 평가자가 상호작용하는 반응자들의 안전, 존엄성, 가치를 존중해야한다는 원칙
 - **대중복지를 위한 책무성 영역**(5개 항목) : 평가자가 일반 대중과 관련된 관심과 가치의 다양성을 파악하고 그를 참조해야한다는 원칙

Ⅲ. 보편적 평가기준 : 평가수행에 관한 규범

3 기타 평가기준

2) 정부 심사 기준

- 미국 정부의 회계감사원이 주도하여 개발한 감사 또는 심사를 위한 기준
- 프로그램 평가를 포함하여 재무 심사 등에 적용되는,
일반 기준, 현장 기준, 보고 기준으로 구분되고 있음
 - **일반 기준** : 독립성, 전문적 판단, 능력, 질 관리 및 보장에 관한 사항 포함
 - **현장 기준** : 계획수립, 감독, 근거제시, 심사 기록 등을 규정
 - **보고 기준** : 보고형식, 보고내용, 보고의 질적 요인들, 보고서 배포 등을 규정

평가기준의 설정 방법 개요

1 평가기준 설정

1) 평가의 실제에서 논란적인 활동요소

2) 평가적 판단(가치판단)을 위한 평가기준 요청

- 기준설정+판단과 연결
- 평가활동의 가장 핵심적인 부분 차지

3) 프로그램의 효과나 가치 판단 결과 좌우: 기준 설정방법에 따름

4) 학교평가

- 진급여부 좌우
- 프로그램 존폐 좌우
- 학생 개인, 프로그램 기관 의사결정 직간접적 영향
- 평가적 판단에서 결정적인 역할

평가기준의 설정 방법 개요

2 교육평가 및 성적 산출 활동

1) 서열화된 성취기준에 따른 수준 분류의 중요한 역할 수행

- 숙련자, 미숙련자, 탁월, 우수, 기초, 기초이하 수준 분류

2) 성취기준 활용의 교육적 유의미성 부여

- 시험점수 향상: 한 학급 학생들의 수학점수 2점 상승
- 비교: 우수한(proficient: 명확한 정의) 수준의 학생들이 30%에서 35%로 증가

3) 점수 산출 보고위한 성취기준 활용

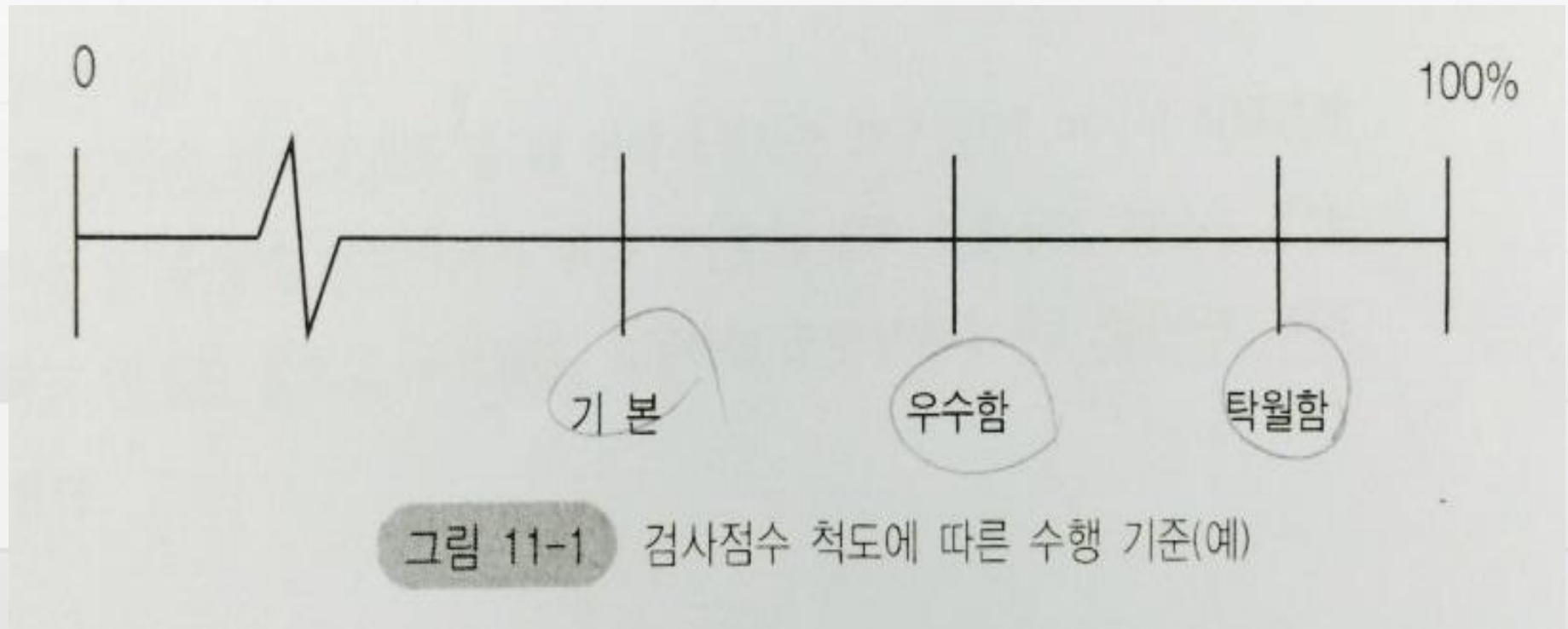
- 평가의 목적: 합격-불합격 vs 수준에 따라 4개의 서열화된 그룹 분류

4) 학교평가

- 진급여부 좌우

평가기준의 설정 방법 개요

3개의 성취 기준



평가기준의 설정 방법 개요

3

교육평가 활동

1) 피험자의 내용이나 기술 측면의 충분한 성취 결정

- 검사점수 척도 중 하나의 수행 기준 요구
- 두 개의 범주: 숙련과 미숙련, 합격과 불합격, 자격있음과 자격없음으로 분류

2) 기준설정과정의 설계, 점검하는 검사기관의 효력있고 가치있는 성취기준 제작의 유용성

평가기준의 설정 방법 개요

4 수행기준관련 선행연구 분석시 3가지 유의점

1) 내용기준과 성취기준의 명확한 구분 중요

- 내용기준: 교육과정에서 피험자 알아야 할 것, 그들이 할 수 있는 것
- 성취기준: 내용기준 바탕으로 기초수준, 우수함, 탁월함 수준의 구분
- 성취기준? 내용기준과 관련하여 피험자가 얼마나 잘 수행하였는가 나타내는 것

2) 모든 기준설정 과정은 판단과정 포함함

- 패널수와 구성방식 결정, 피드백 제공여부, 제공법, 기준설정과정의 통합적 요소임

3) 객관식 문항활용에 대한 교육적 사정활동에서 성취기준 설정법 바른 개발과 실행단계 명확해야 함

가. 일반적인 수행기준 설정 절차(Hambleton(2001))

성취수행 기준 설정 과정 및 절차 11단계(491쪽)

제1단계) 패널 구성

- 교육적 사정활동을 통한 의사결정 시 교사, 교육과정 전문가, 학교 행정가로 구성(일반 시민 대표)

제2단계) 기준설정 방법 선택/훈련자료 준비/회의안건 완성

- 성취기준설정 도구 다양함
- 패널의 과업특성
 - 평가자료와 문항 검토하여 판단
 - 문서선정, 서류분류, 분석적 판단 등 피험자의 성과에 대한 판단
 - 시험점수 분석표에 대한 판단
 - 후보자에 대한 판단
- 기준설정 기법선택 시 평가문항의 혼합여부, 기준설정 위한 시간제한, 방법활용의 선행경험, 도구의 타당도에 대한 인식과 근거 기초, 사전 현장검증 거친 훈련자료 활용

가. 일반적인 수행기준 설정 절차(Hambleton(2001))



성취수행 기준 설정 과정 및 절차 11단계

제3단계) 성취범위에 대한 기재문 작성

- 타당도 높은 성취기준 제작위해 성취수준 기재문의 중요성 높아짐
- 시험자의 지식, 기술, 능력이 구체적으로 진술됨

제4단계) 선택된 방법을 패널이 활용할 수 있도록 훈련

- 효용성과 타당성 높은 성취기준 설정위해 패널에게 효과적인 훈련과 연습필요
- 효과적인 패널 훈련 요소
 - 기준설정에 따른 단계 설명과 모형화(경계선 사람 산정과 정해진 범주의 시험자 시험지 각각 분류)
 - 점수부여의 요점과 방식 보여주고, 확실히 이해시키기
 - 활용하기 쉬운 등급 산출방식 만들기

가. 일반적인 수행기준 설정 절차(Hambleton(2001))



성취수행 기준 설정 과정 및 절차 11단계

제4단계) 선택된 방법을 패널이 활용할 수 있도록 훈련(계속)

- 등급을 산출하는데 익숙하도록 하기
- 과정에서 활용할 규범적 자료에 대해 설명하기
- 패널들이 평가내용에 익숙하도록 하기
- 성취수준의 경계부분에 대해 정의하기
- 기준 하 또는 기준조건과 유사한 상황에서 검사 실시하기
- 성취기준이 설정될 문항 풀 검토하기

■패널의 역할

- 피험자의 성취에 영향을 미칠 요인 알고 있어야 하며 기준설정 과정에서 요인 고려
- 성취기준 전달할 위원회 혹은 주무기관과의 관계 올바르게 이해하고 있어야 함

가. 일반적인 수행기준 설정 절차(Hambleton(2001))



성취수행 기준 설정 과정 및 절차 11단계

제5단계) 패널로부터 문항등급의 자료 수집

- 훈련이 잘 이루어졌다면 간단하게 이루어짐
- 유창한 피험자에게 최소한 기대되는 성취기준 결정 요청했을 경우 패널들 간 중간 값이나 최소/최대 범위 산출→다음 단계 패널토의 시 제공

제6단계) 패널 토의 실시

- 패널 간 토의 성립
- 토의목적: 최초 등급설정에 대한 재고의 기회 제공과 오류, 잘못된 인식, 오해 찾기
- 피드백의 내용: 평균 성취정도와 피험자 점수분포, 평가의 과제, 패널들의 등급부여에 대한 기술통계가 일반적임

가. 일반적인 수행기준 설정 절차(Hambleton(2001))



성취수행 기준 설정 과정 및 절차 11단계

제7단계) 토의를 통한 2차 문항등급 결과 수집

- 상호작용적인 과정으로 실시되나 필수단계 아님
- 두 단계 등급부여 과정: 등급을 제공하고 토의 이루어진 뒤 → 패널, 등급부여 작업함
- 등급변경이 필수는 아니나 한 번의 기회 더 주어짐의 의미

제8단계) 패널이 부여한 등급결과 수집 및 성취기준 수립

- 패널들이 부여한 등급결과가 성취기준으로 설정
- 작은 표집단위나 비대칭분포일 경우 중간 값 선호
- 하위패널과 패널 사이 성취기준 다양성에 대해 정리
- 오차는 성취기준 결과의 변동성 결정에 활용
 - 오차 클 경우: 기준 신뢰도 낮추고
 - 오차 작을 경우: 기준 신뢰도 높임
 - 오차: 성취기준 조정에 활용

가. 일반적인 수행기준 설정 절차(Hambleton(2001))



성취수행 기준 설정 과정 및 절차 11단계

제9단계) 패널에게 결과자료 제시

- 앞선 단계의 선정 기준이 패널의 경험과 합리적 기준에 부합하지 않는 경우 수정기회 제공
- 예) 80%이상의 피험자가 '기초이하'수준 분류된 경우: 비합리적, 피험자의 다른 자료와 비일치 → '기초'수준의 피험자 기준 변경 → '기초'수준의 피험자 늘어나고 '기초이하'수준 감소
- 정보, 보고시간, 표준 보고형식: 패널에게 너무 이르게 또는 늦지 않게 전달 적절한 선정

제10단계: 필요시 수정을 통한 성취기준 완성/결과물에 대한 패널평가 실시

제11단계: 타당한 근거 수집과 기술 내용의 문서화

- 기준설정 수행 시: 패널 선정을 위한 특성분석, 훈련, 도구활용, 최종결과 모두 문서화

나. 전통적인 기준 설정 방법

1) 전문가 견해 이용법

- 프로그램 관련 전문가의 판단에 의거하여 프로그램 실시 결과로 예상되는 성취 수준이나 산출 정도 결정하는 방법
- 특정교과 전문가의 견해 중요한 기준 설정의 근원이 되며 경험과 전문성에 입각하여 기준 설정하며 교육성취도 평가에서 주로 활용

2) 선행 업적(과거 성취 및 수행 결과) 이용법

- 사전에 프로그램에 참여집단의 성취 결과나 업적에 근거한 기준설정법

3) 비교법

- 특정 집단의 성취 결과를 여러 다른 집단의 성취 결과와 일시에 비교하거나 한 집단의 성취 결과들과 여러 차례 비교, 여러집단의 여러 차례 성취 결과 종합적 비교법을 통한 평가 기준 설정

나. 전통적인 기준 설정 방법

4) 필수 기능 수행 결과 활용법

- 특정 프로그램의 성취에 필수적으로 요청되는 기초적이고 기본적인 기능중시, 수행능력이나 성취수준을 기준으로 삼는 방법

5) 장기 목표 활용법

- 장기적으로 추구하는 기능이나 학습과제인 경우 1년 후, 또는 한 학기 후에 어느 정도 수준에 도달하거나 어떤 성취수준 보일지 중시하여 점진적 향상과 성취를 전제로 체계적 기준을 설정하여 활용하는 법
- 출발점상태의 학습능력 및 수준파악→기초하여 한 학기나 1년간 성취할 가능성 바탕으로 달성할 만한 수준을 계산하여 기준선으로 삼는 방안

나. 전통적인 기준 설정 방법

6) 동료 및 유사 집단 이용법

- 수용가능한 성취수준인 기준을 동년배 집단, 동일 학년 집단, 프로그램 대상집단과 유사한 집단의 평균성취결과를 바탕으로 설정하는 방법

7) 규준집단 이용법

- 측정이론적 배경을 바탕으로 측정 대상 속성에 관련한 규준집단을 구성하여 그 규준집단의 성취결과를 바탕으로 통계적인 기준 설정과정을 거쳐 평가기준을 설정하는 방법
- 표준화 학력검사를 이용한 프로그램 효과평가

8) 문항반응결과 활용법

- 최저능력수준을 설정하기 위한 경우 주로 사용하는 방법으로 학생들의 성취 결과를 일정한 방식으로 분석하여 그로부터 얻은 점수를 기준으로 활용

나. 전통적인 기준 설정 방법

8) 문항반응결과 활용법: Nedelsky 모형(502쪽)

- 2~3인의 교사가 능력검사 또는 학력검사 문항을 일일이 고찰하고 문항마다 가상적으로 가장 낮은 수준의 학생의 오답 답지 골라 제외하여 나머지 답지 대상으로 정답을 짐작하여 맞출 확률을 중심으로 최저 통과 점수를 계산하고 기준설정하는 법

나. 전통적인 기준 설정 방법

8) 문항반응결과 활용법: Ebel 모형(502쪽 하단)

- 학력검사 문항 각각에 대하여 3인의 교사가 각자 주관적으로 적절한 성과곤란도를 평정하고 그를 바탕으로 계산하는 방법
 - 문항의 적절성(교사): 필수적임-중요함-수용할만함-의문시됨 네 분류
 - 곤란도: 쉽다-중간이다-어렵다 세 분류 종합하여 유관표 제작 문항통과율(성공율) 계산(503쪽 표 11-3)

표 11-3 Ebel 모형에 의한 문항 통과율 계산

문항분류		문항수	예상 성공률	문항수 성공률
필수적임				
	쉽다		100	
	중간이다	21	75	1,575
	어렵다	53	50	2,650
중요함				
	쉽다	24	90	2,160
	중간이다	49	65	3,185
	어렵다	38	40	1,520
수용할만함				
	쉽다	24	75	1,800
	중간이다	59	45	2,655
	어렵다	19	30	570
의문시됨				
	쉽다	4	50	200
	중간이다	7	25	175
	어렵다	2	10	20
합계		300		16,510

(Kosecoff & Fink, 1982, 73-78)

다. 최근의 기준 설정 방법

1) 검사결과로 얻은 점수를 중심으로 합격 여부를 판단하는 기준인 합격선을 결정하는 문제는 자격을 부여하는 최저능력수준을 나타내는 기준 점수로서 역할 충실히 수행하는 도구, 타당도, 난이도 만족

2) 성취수행 기준 설정 네 가지

- Group referencing: 소속집단 구성원의 평균적 성취업적을 기준으로 사용
- Task referencing: 이미 설정된 기준이나 수월성 기준활용 법, 준거 참조 평가관에 입각하여 등급 정하는 일 해당
- Expectation referencing: 각 대상별, 여건 및 능력 수준 고려하여 개별적 기준 설정하고 기대수준이하 성취자, 기대수준 이상 성취자로 분류활용
- Self referencing: 평가 대상의 사전 성취수준을 기준선, 출발선으로 삼고 특정 기간 이후의 성취와 비교하여 진전도나 향상도 중심의 평가기준 설정
- 가장 교육적이며 인간적인 기준설정 방법: 특정기간의 개인별 변화초점두어 개별적 평가기준 비추어 그 변화정도 판정하는 평가논리

다. 최근의 기준 설정 방법

1) 평가기준 설정 방법: 수행평가 중심, 구체적인 성취 결과에 초점 둔
기준을 설정하는 방법이 주종 이룸

2) 구체적인 기준 설정 방법

- Body of Work(BoW) 기법
- 북마크(Bookmark) 기법
- 군집분석을 활용한 기준설정

다. 최근의 기준 설정 방법: **Body of Work(BoW) 기법**

1) 둘 이상의 집단으로 분류하기 위해 객관적 방법 활용

- 과목을 이수하기 위한 학점은 누가 취득해야 하는가?
- 누가 고등학교를 졸업할 수 있는가?
- 구성반응(constructed-response)이 요구되는 문항으로 이루어진 학력검사(자극에 따른 사고결과를 기술하는 지필검사)가 주된 대상

2) 성취수준과 일반적 설명(표11-4, 508쪽)

- BoW의 기준설정 위해 목표 조준, 범위 설정, 훈련 등 세 가지 과제에 대한 준비 필요

improvement; 더 노력이 요구됨), 낙제(failing; 실망스런 수준)와 같은 성취 수준을 설정했다.

표 11-4 성취수준과 일반적 설명

탁월함	이 수준의 학생들은 심화된 주제에 대해 광범위하고 깊이 있는 이해를 하며, 복잡한 문제에 대해서도 적합한 해결책을 내놓을 수 있다.
우수함	이 수준의 학생들은 도전적인 과제에 대해 견고한 이해를 하며, 다양한 문제를 해결할 수 있다.
개선 요망	이 수준의 학생들은 주제에 관한 부분적인 이해를 하며, 몇몇 단순한 문제를 해결할 수 있다.
낙제	이 수준의 학생들은 주제에 대해서 최소수준의 이해에 그치며, 단순한 문제조차도 해결하지 못한다.

이처럼 각 성취수준에 대한 일반적인 기술과 함께, 내용 전문가는 각 기준내용에 대해 무엇을 한다. MCAS에서는 8학년 수학평가에서 '우수한(proficient)' 수준의 학생을 다음

다. 최근의 기준 설정 방법: 북마크(Bookmark) 기법

1) 시험이나 검사의 경계점수를 설정하기 위한 방법

▪ 북마크 기법의 목적

- 선택형 반응과 구성반응의 문항방식 통합
- 판단상 인지적 부담을 줄임으로써 판단적 과제 단순화하는 것
- 경계점수를 설정하는 판단과정과 측정모델을 연계
- 시험 내용과 성취 수준에 대한 설명을 연계

2) 북마크 절차 3단계

- 1단계: 최초 큰 단위의 집단의 검토, 다음 작은 단위의 집단에서 순서화된 문항책자 검토
- 2단계: 논의는 계속 소그룹 차원에서 진행되고 집단의 리더는 제1단계에서 도출된 경계점수에 대한 피드백 시작함
- 3단계: 결과 자료를 큰 집단에게 발표함으로써 시작됨, 각 성취수준에 해당하는 학생의 기대비율이 언급됨

다. 최근의 기준 설정 방법: 군집분석을 활용한 기준설정

1) 이분법적으로 점수를 부여한 문항과 다중 군으로 점수화시킨 문항, 혹은 둘 모두를 포함하는 시험에 대한 기준 설정이나 평가하는 추가방법 고찰

- 통계적인 분석에 의해 기준 설정
- 군집분석방법 활용
- 기준설정에 관한 기존 연구로부터 도출된 보충자료의 의미 지님

2) 군집분석 수행위한 두 요소: 대상과 변수

- 대상: 군집되는 것, 변수: 군집 형성에 사용되는 준거
- 시험응시자의 군집화 변수 선정 위한 3가지 선택사항
 - 시험을 구성하는 모든 개별 문항을 활용하는 것
 - 문항수준 요인 분석으로부터 얻은 직교요인점수 활용
 - 시험의 주요내용 부분을 구성하는 문항들로부터 추출된 하위점수 활용

다. 최근의 기준 설정 방법: 군집분석을 활용한 기준설정

1) 군집분석 2가지

- 위계적(N-군집) 군집분석: 모든 대상을 최초 독립된 유일한 군집들로 생각함, 최종 하나의 군집단위로 모일 때까지 한 번에 한 단계씩 군집을 병합하는 과정에 의해 분석함

2) 군집분석의 유용성

- 주관적이 아니라 경험적으로 패널 후보의 개요를 추출하는 경우
- 패널 없이 기준을 추출하는 경우
- 다면평가에서 기준을 설정하는 경우
- 다양한 문항형태가 담긴 시험의 기준을 설정하는 경우

다. 최근의 기준 설정 방법

1) 평가기준 설정 방법: 수행평가 중심, 구체적인 성취 결과에 초점 둔
기준을 설정하는 방법이 주종 이룸

2) 구체적인 기준 설정 방법

- Body of Work(BoW) 기법
- 북마크(Bookmark) 기법
- 군집분석을 활용한 기준설정

라. 기준 설정 과정에 대한 질 관리: 개요

- 1) 평가 목적을 포함하여 적절한 평가기준 설정방법을 선정하고 충실히 시행하여 평가목표를 만족스럽게 달성하는 일: 평가자나 교원의 중요한 책무임
- 2) 평가목적에 입각한 판단과정 통해 수행기준 설정방법과 설정과정의 자율적 질 관리 노력 필요

라. 기준 설정 과정에 대한 질 관리 20가지

- 패널을 대표하는 그룹과 각 그룹의 구성 비율에 관하여 고려하였는가?
- 패널은 교육평가에서 성취기준을 설정하는데 적절한 판단이 가능한 충분한 인원수의 적합한 인사들로 구성되었는가?
- 두 팀의 패널이 성취기준의 일반화 가능성을 점검하도록 구성되어 활용되었는가?
하나의 패널그룹 안에 하위 패널들이 독립집단 내 성취기준의 일관성을 점검하도록 구성되었는가?
- 연구를 적절히 수행하도록 충분한 자원이 할당되었는가?
- 기준 설정 방법을 수행 기준 수립 연구에 사용할 수 있도록 미리 충분하게 현장검증하였으며, 그 검증 결과에 따라 그 방법을 수정하였는가?
- 기준 설정 방법이 특정 교육평가 활동에 적절한 것이며 그에 관하여 구체적으로 기술하였는가?
- 패널에게 초기 기준설정 회의를 통하여 교육평가의 목적과 검사 점수의 활용에 대하여 설명하였는가?

라. 기준 설정 과정에 대한 질 관리 20가지

- 패널의 자격에 관한 자료나 관련 인구통계학적 자료를 수집하였는가?
- 패널이 교육평가 행사를 관리하거나 적어도 평가 행사의 일부분에라도 참여하였는가?
- 패널들은 수행기준을 설정하는 방법에 관하여 적절히 훈련 받았는가?
- 수행 범주는 기준설정 과정에서 패널이 효과적으로 사용할 만큼 명확히 기술되었는가?
- 등급의 차이에 관하여 토의하고 조정하기 위한 반복적인 과정이 있었다면, 패널들에게 제공된 피드백은 명확하고 이해가능하며, 유용하였는가?
- 모든 과정은 효율적으로 수행되었는가? 등급 매기는 형식은 사용하기 쉬웠는가?
- 패널이 성취에 관한 근거자료를 기초하여 그리고 그 자료의 활용방법에 기초하여 등급을 정하였는가?
- 정보를 활용방법과 결과에 대하여 심사숙고할 수 있도록 패널에게 결과 자료를 제공하였는가?
- 최종 성취기준을 명확하게 기술하고 적합한 것이 될 수 있도록 접근방식을 취하였는가?

라. 기준 설정 과정에 대한 질 관리 20가지

- 기준설정 과정에 대한 평가가 패널 멤버들에 의하여 이루어졌는가?
- 성취기준의 타당도를 지지하기 위하여 근거자료를 적절하게 편집하였는가?
- 모든 기준설정 과정을 기록하여 문서화하였는가?
- 성취기준을 관련자들에게 의사소통하기 위하여 효과적인 단계를 취하였는가?

감사합니다.

고맙습니다.