

가정과 수행평가의 타당성과 신뢰성 연구

- 경남 농촌지역의 소인수 학급을 대상으로 -

A Study on the Validity and Reliability of Performance Assessment in Home Economics

육중고등학교
교사 윤성희
경상대학교 가정교육과
조교수 유태명

Ok-Jong High School

Teacher : Sung-Hee Yoon

Dept. of Home Economics Education, Gyeong Sang National University

Assistant Professor : Taemyung Yoo

◀ 목 차 ▶

- | | |
|------------|-------------|
| I. 서론 | IV. 결과 및 논의 |
| II. 이론적 배경 | V. 결론 및 제언 |
| III. 연구방법 | 참고문헌 |

<Abstract>

The purpose of the study was to verify validity and reliability of performance assessment in home economics by applying performance assessment tools developed by Korean Curriculum Assessment Institute and by the researcher to a class of 17 high school female students in a rural area. For the purpose of this study, the research question was formulated and addressed: "How are the validity and reliability of performance assessment in home economics?"

The results of the study were as follows.

1. The results of testing validity
 - a. Thirteen high school home economics teachers, whose average teaching years is ten years verified high validity in the contents of performance assessment.
 - b. In the validity of the items that is verified being based on the scores of high school entrance examination of 2000 school year, the essay type showed its characteristics as performance

assessment.

c. The generalizability among assessment methods was not high.

2. The results of testing reliability

Both Cronbach's α , inter-item reliability of the assessment tools used in the research, and reliability among the raters were high. But reliability among the tasks was relatively low.

This study proposed that the diverse performance methods, the increasing number of assessment, the explicit assessment criterion, and scoring criterion are needed to satisfying validity and reliability of performance assessment.

주제어(Key Words): 수행평가(performance assessment), 타당성(validity), 신뢰성(reliability), 가정과교육(home economics education)

I. 서론

1. 연구의 필요성 및 목적

세계화 정보화 시대를 맞이하여 우리 사회는 획일성과 경직성을 극복하고 다양화 된 시대를 주도해 나갈 수 있는 사고력과 창의력, 문제 해결능력을 중요시하게 되었다. 그에 따라 학교 교육은 학습자의 다양한 개성을 존중하고 개인의 무한한 잠재력을 찾아 길러주며 인성 및 창의력을 신장시키는 전인교육의 방향으로 변화가 요구되고 있다.

그러나 전통적으로 학교 교육 현장에서 시행되어 온 평가 방식은 주로 선택형 지필 검사를 이용하여 학생을 선발·분류·배치하기 위한 양적인 측정에만 치중해 왔으며 이는 채점의 객관성과 공정성을 확보할 수 있다는 이유로 정당화 되어왔다. 이러한 평가 방식은 단순한 지식이나 정보의 습득 여부에 대한 평가를 하기에는 좋은 방법이나 학생들의 창의성이나 문제 해결력 등 고등사고기능을 평가하기는 어려운 방법이다(백순근, 1999a).

과학기술의 발전과 급변하는 사회 구조 속에서 교육은 실생활에 적응해 나갈 수 있는 유능한 생활인을 기를 것을 요청 받고 있다. 따라서 가정교과에서는 가정생활을 중심으로 생활에 대한 기초 기능과 지식, 정보를 습득하게 함으로써 기본적인 인간생활의 적응력과 미래 사회의 변화에 대처할 수 있는 능력을 기르도록 해야 한다.

가정교과의 교수학습은 실험 실습, 토의, 조사, 관

찰 활동을 중심으로 학생의 체험적 경험을 중시하고 평가도 역시 실제적인 수행 능력을 중심으로 이루어질 필요가 있다. 습득한 지식과 기능을 실제에 적용하고 실천하는 것을 강조하는 가정교과는 그 특성이 교수학습 방법과 평가에 반영되어야 하고 배우는 과정과 결과를 모두 중시하여야 한다. 그러나 윤인경(1999)은 현재 새롭고 다양한 교수·학습 방법이 개발되어 적용되고는 있지만 그에 따르는 적절한 평가가 소홀하였음을 지적하였다.

전통적인 평가 방식의 문제점을 인식하여 「교육비전 2002: 새 학교 문화 창조」(교육부, 1998)에서는 학생들의 창의력과 문제해결능력을 길러 주고 개성과 다양성을 조장하는 평가, 석차 매기기를 위한 평가가 아니라 학습활동을 돕기 위한 평가 체제로 변화시키기 위해 수행평가의 당위성을 제시하였고, 교육현장에서 1999학년도부터 수행평가가 전면적으로 시행되고 있다. 그러나 개발 보급된 평가 도구가 거의 없고 일선교사가 개발하였더라도 타당성과 신뢰성 확보에는 많은 어려움을 겪고 있다. 더욱이 수행평가의 결과가 대학 입시 내신자료로 반영되는 현실에서는 평가 도구의 타당성과 신뢰성 검증은 절실하며 교육부 및 시·도 교육청에서도 수행평가의 신뢰성 확보를 강조하고 있다.

본 연구의 예비조사로서 경남 지방의 93개 고등학교의 가정과 교사를 대상으로 1999년 11월 수행평가의 실태를 설문 조사한 결과 아래와 같은 문제점을 발견하게 되었다.

첫째, 1999년 수행평가를 처음 실시할 즈음에는

교사들의 수행평가에 대한 연수 부족으로 수행평가에 대한 이해가 부족하였다. 응답자의 63.6% 교사가 학교자체 또는 개별적인 수행평가 연수를 하였고 도교육청 단위의 연수는 시·군별 1, 2명의 연수자에 불과하여 수행평가에 대한 포괄적인 이해가 부족하였다.

둘째, 수행평가 도구 개발과 평가기준의 설정에 많은 어려움을 겪고 있었다.

셋째, 수행평가 영역은 가정교과 전 영역에 고루 분포되지 못하였고 특히 학기말이나 학년말에 편성된 단원은 거의 수행평가가 실시되지 못하였다.

넷째, 다양한 방법의 수행평가가 실시되지 않고 있었다.

이와 같은 상황은 단지 경상남도에 국한되지 않음은 잘 알려지고 있다. 시의 적절하게 한국 교육과정 평가원은 1999년에 고등학교 10개 교과에 대한 「국가 교육과정에 근거한 성취 기준 및 평가 도구」를 개발하였으며 수행평가 도구와 평가기준 및 채점기준이 개발되어 많은 도움이 되리라 본다. 왕석순(1999)의 「고등학교 가정-국가 교육과정에 근거한 성취기준 및 평가 도구」에서는 현재의 교실상황이나 지역여건 등을 충분히 고려하여 개발하지 못하였으므로 '현장적용을 통한 실험연구가 뒤따라야 할 것'을 제언하였다.

본 연구에서는 농촌지역의 남녀 공학 소규모 학교, 소인수 학습의 여건에서 「국가 교육과정에 근거하여 개발한 성취기준 및 평가도구」 중 가족의 구조와 기능 변화, 부모 됨의 의미와 역할, 임신과 출산에 관한 서술형과 논술형, 선택형의 평가 도구와, 연구자가 개발한 단답형의 평가 도구를 적용하여 평가 문항의 난이도와 변별도, 평가 도구의 타당성과 신뢰도를 검증하여 현장적용을 위한 기초자료로 제공하고자 한다.

2. 연구 문제

위와 같은 연구목적은 이루기 위해 다음과 같은 연구문제와 하위 연구문제를 설정하였다.

연구문제 : 가정교과 수행평가 방법의 타당성과 신

뢰성은 어떠한가?

하위 연구문제:

- (1) 서술형, 논술형, 선택형, 단답형 문항의 질은 어떠한가?
 - 1) 각 평가 도구의 난이도는 어느 정도인가?
 - 2) 각 평가 도구의 변별도는 어느 정도인가?
- (2) 서술형, 논술형, 선택형, 단답형의 평가도구는 타당한가?
 - 1) 각 평가 방법의 내용 타당도는 어떠한가?
 - 2) 각 평가 방법의 준거 관련 타당도는 어떠한가?
 - 3) 각 평가 방법의 구인 타당도는 어떠한가?
- (3) 서술형, 논술형, 선택형, 단답형 평가도구는 신뢰로운가?
 - 1) 문항 내적 일치도는 어느 정도인가?
 - 2) 평정자간 신뢰도는 어느 정도인가?
 - 3) 과제간 신뢰도는 어느 정도인가?

II. 이론적 배경

1. 수행 평가의 개념과 특징

교육부(1998b, 1998c)는 '수행평가 (performance assessment)란 평가자가 학습자들의 학습과제 수행 과정 및 결과를 직접 관찰하고 그 관찰 결과를 전문적으로 판단하는 평가 방식을 의미한다.'고 정의하였다. 윤인경(1999)은 '수행평가는 근본적으로 학생이 무엇을 알고 있는가만 측정하기보다는 학생이 알고 있는 지식을 바탕으로 무엇을 할 수 있는지를 재려는 데 목적을 두고 지식(knowledge), 기능(skills), 가치(values)를 모두 평가의 대상으로 포함하는 평가 방식'이라고 정의하였다. 최호성(1999)은 '학생이 과제를 수행해 가는 모습을 관찰하고 학생의 작품을 살펴보고 그 질을 판단하는 일'이라고 하였으며 수행평가의 구성은 수행과제와 수행기준으로 구성되는데 수행과제는 학생 자신이 알고 있고 실행 할 수 있음을 증거하기 위해서 관여하는 활동이라고 하였다. 백순근(1999a)은 '학생 스스로 자신

의 지식이나 기능을 나타낼 수 있도록 산출물을 만들거나 행동으로 나타내거나 답을 작성하도록 요구하는 평가 방식'이라고 하였다. 이상을 종합해 볼 때 수행평가는 지식이나 기능, 가치를 종합하여 학생이 실제로 활동하는 전 과정과 활동의 결과를 평가하는 것이라고 할 수 있다. 그러므로 수행평가는 교수·학습과 평가의 통합이라고 할 수 있으며 학습자의 총체적인 발달에 초점을 두고 평가하는 평가방법이다.

수행평가와 기존의 평가 방식의 특징을 비교해 보면 <표 1>과 같다.

수행평가에서는 학생이 답안을 작성, 구성하거나 행동으로 나타내는 것을 통해 지식이나 기능을 직접적으로 측정·평가하고자 한다. 전통적 평가와 대비되는 수행평가의 일반적인 특징을 한국교육과정평가원(1999)은 다음과 같이 제시하였다.

첫째, 수행평가는 학생이 문제와 정답을 선택하는 것이 아니라, 자기 스스로 답을 작성, 서술, 구성하거나 행동으로 나타내도록 하는 평가 방식이다.

둘째, 수행평가는 추구하고자 하는 교육목표의 달성 여부를 가능한 한 실제 상황 하에서 파악하고자 하는 평가 방식이다.

<표 1> 수행평가와 전통적 평가방식의 특징 비교

구 분	전통적 평가체제(선택형 지필 평가)	새로운 평가 체제(수행평가)
시대적 상황	산업화시대 소 품종 대량생산	정보화시대 다 품종 소량 생산
학습관	추상적 객관적 상황 중시 학습자의 기억·재생산 중시	구체적 주관적 상황 중시 학습자의 이해 성장 중시
평가체제	규준 지향 평가(상대 평가) 양적 평가	준거 지향 평가(절대평가) 질적 평가
평가 목적	선발·분류·배치 한 줄로 세우기	지도·조언·개선 여러 줄 세우기
평가 내용	내용적 지식 학습 결과 중시 학문적 지능의 구성요소	절차적 지식 학습과정 중시 실천적 지능의 구성요소
평가 방법	선택형 평가 위주 표준화 검사 중시 대규모 평가 중시 일회적·부분적인 평가 객관성·일관성·공정성 강조	관찰자 판단 위주 개별교사에 의한 평가 중시 소규모 평가 중시 지속적·종합적인 평가 전문성·타당도·적합성 강조
평가 시기	학습활동이 종료되는 시점 교수·학습과 평가활동 분리	학습활동의 모든 과정 교수·학습과 평가 활동 통합
교사의 역할	지식의 전달자	학습의 안내자·촉진자
학생의 역할	수동적인 학습자 지식의 재생산자	능동적인 학습자 지식의 창조자
교과서의 역할	교수·학습·평가의 핵심내용	교수·학습·평가의 보조자료
교수·학습 활동	교사 중심 인지적 영역 중심 암기 위주 기본 학습능력 강조	학생 중심 지·정·체 모두 강조 탐구 위주 창의성 등 고등사고 기능 강조

(자료 출처: 교육부, 1998b)

셋째, 수행평가는 교수·학습의 결과뿐만 아니라 교수·학습의 과정도 함께 중시하는 평가 방식이다.

넷째, 수행평가는 단편적인 영역에 대해 일회적으로 평가하기보다는, 학생 개개인의 변화·발달과정을 종합적으로 평가하기 위해 전체적이면서도 지속적으로 이루어지는 것을 강조하는 방식이다.

다섯째, 수행평가는 개개인을 단위로 해서 평가하기도 하지만, 집단에 대한 평가도 중시하는 평가방식이다.

여섯째, 수행평가는 학습과정을 진단하고 개별학습을 촉진하려는 노력을 중시하는 평가 방식이다.

일곱째, 수행평가는 학생의 인지적인 영역(창의성이나 문제 해결력 등 고등사고기능을 포함)뿐만 아니라, 학생 개개인의 행동발달 상황이나 흥미·태도 등 정의적인 영역, 그리고 운동기능 영역에 대한 종합적이고 전인적인 평가를 중시하는 평가이다.

마지막으로 수행평가는 기억, 이해와 같은 단순 사고능력보다는 창의, 비판, 종합과 같은 고등 사고능력의 측정을 중히 여기는 평가 방식이다.

2. 수행평가 방법

수행평가는 교사가 학습자의 학습과제, 수행과정이나 결과를 직접 관찰하고 관찰 결과를 전문적으로

로 판단하는 평가 방식으로 평가의 대상, 평가하고자 하는 요소에 대한 측정 방법, 측정결과에 대한 평가 방식 등 수행평가 방식을 구성하는 핵심요소들이 다른 종류의 평가 방식을 구성하는 요소들과 어떻게 다른지에 따라 정의되는데 그 형태는 다양

<표 2> 수행평가의 본질을 구현하는 정도

수행평가 본질의 구현정도	평가 방법*
↑ 매우 높음 ↓ 매우 낮음	실제 상황에서의 평가
	실기시험, 실험·실습법, 관찰법
	면접법, 구두시험, 토론법
	자기평가 및 동료평가 보고서법
	포트폴리오
	연구보고서법
	논술형
	서술형
	단답형
	완성형(괄호형)
	선다형
	연결형(줄 긋기형)
	진위형(○×형)

(자료 출처: 백순근, 1999c)

*특정의 평가 방법이 어떤 상황에서 어떠한 방식으로 시행되느냐에 따라 위치가 달라질 수 있음.

<표 3> 수행평가에 포함되는 평가 방법

평가 대상	구성적 반응요구	특정 산출물 요구	특정활동 요구	과정을 밝힘
수행평가방법	· 완성형 문항* · 단답형 문항* · 도표나 그림에 제목 붙이기 · 과제물 붙이기 · 시각적 자료 만들기(개념도, 흐름도, 그래프, 표, 도안 등)	· 수필 · 연구보고서 · 과제일지 · 실험 보고서 · 이야기/극본 · 시 · 포트폴리오 · 미술작품 전시 · 과학 프로젝트 · 비디오구성/ 오디오 구성	· 구두발표 · 무용/동작 발표 · 과학 실험 시연 · 체육 경기 · 연극 · 토론 · 음악 발표	· 구두 질문 · 관찰 · 면담 · 회의 · 과정에 대한 기술 · 생각하는 과정을 말로 표현 · 학습일지

(자료 출처: 한국교육과정 평가원, 1999)

*완성형과 단답형 문항이 단순 암기력만 평가하기 위한 것이라면 수행평가에 해당되지 않는다.

하며 교과와 내용적 특성에 따라 다양한 방법이 적용되어야 한다. 다양한 평가 방법들은 수행평가의 본질을 구현하기 위한 하나의 도구들이며 수행평가가 전제하고 있는 중요한 특성들이 평가 상황에서 제대로 구현될 때 비로소 수행평가가 시행되고 있다고 할 수 있다(한국교육과정 평가원, 1999a).

수행평가의 본질을 구현하는 정도에 따른 평가 방법을 분류하면 <표 2>와 같다.

동일한 평가 방법이라고 하더라도 교과목의 특성이나 평가 하고자 하는 능력의 특성에 따라 수행평가의 본질은 달라진다는 것을 주지해야 한다. <표 2>에 제시된 수행평가 방법에 대한 구분은 상호 배타적이기보다는 상호 보완적이며 창의성이나 문제해결력 등 고등사고 기능의 신장을 중시하고, 교수·학습의 과정을 개선하는 것을 중시하며 개별학생에게 지도·조언하고 충고하기 위한 목적으로 사용되기만 한다면 어떤 평가 방법도 수행평가 방법에 포함 될 수 있다. 특히 수행평가에서는 교수·학습활동과 평가활동이 상호 통합적으로 진행되는 것을 강조하기 때문에 다양한 교수·학습방법 들이 곧 수행평가를 위한 좋은 평가 방법이 될 수 있다(백순근, 1999c).

수행평가에 포함되는 평가 방법을 구성적 반응, 특정 산출물, 학생의 특정활동이나 활동의 과정 등의 평가 대상별로 분류해 보면 <표 3>과 같다.

3. 평가도구의 질을 판단하는 준거

1) 문항 분석

평가도구의 질을 판단하는 데에는 일반적으로 도구의 신뢰도와 타당도에 준한다. 그러나 도구의 신뢰도와 타당도를 논하기 전에 도구를 구성하는 문항 하나 하나가 타당하고 신뢰로운 것인가에 대한 분석도 매우 중요하다. 문항에 대한 통계적 분석 중 보편적으로 사용하는 난이도와 문항 변별도를 중심으로 살펴보기로 한다.

(1) 난이도

난이도는 문항 곤란도 지수로 나타낸다. 난이도를 결정하는 방법에는 전문가의 판단에 의하는 방법,

한 문항을 해결하는 속도에 의하는 방법, 한 문항에 의한 정답 수 또는 정답비율에 의한 방법 등이 있다. 앞의 두 방법은 적성검사, 지능검사, 성취검사 등의 표준화 검사를 제작할 때 우선 고려해야 할 첫 단계이다. 한 문항에 대한 정답 수 또는 정답비율에 의하는 방법은 한 검사를 실험적으로 실시한 후에 그 결과를 바탕으로 하나 하나의 문항을 객관적, 통계적으로 분석하는 것으로서 난이도를 결정하는데 널리 사용되고 있다. 문항 곤란도 지수는 한 문항에 정답 한 학생의 백분율로 표시하는데 그 수식은 다음과 같다.

$$P = \frac{R}{N} \times 100$$

P : 문항 곤란도 지수(난이도)

N : 전체 사례 수

R : 정답자 수

문항 곤란도 지수(P)가 클수록 곧 정답자수가 많을수록 쉬운 문항이다. 선택형 문항에서는 우연히 정답을 맞출 가능성을 제거하고 문항 곤란도를 결정한다(백용덕, 1995).

(2) 문항 변별도

문항 변별도는 한 문항의 타당도를 말한다. 문항 변별도는 문항 하나 하나가 얼마나 피험자의 상하능력을 잘 변별해 내느냐를 뜻하는 것으로 검증 결과로 얻어진 수치를 변별도 지수(discriminating index)라 한다. 변별도 지수는 -1~+1사이에 분포하며 +0.2이상의 변별도 지수가 되는 문항은 무리 없이 사용할 수 있다. 변별도 지수는 곤란도가 50%정도일 때 가장 클 수 있다(김옥환 외, 1983).

2) 타당도

타당도란 하나의 평가 도구 혹은 검사가 '측정하려고 하는 것을 얼마만큼 충실하게 측정하여 주고 있는지'를 의미하는 것이다. 타당도를 결정하는 방법에는 여러 가지 종류가 있겠으나 여기서는 내용 타당도, 준거타당도, 그리고 구인타당도를 살펴보기로 한다(백용덕, 1995, 남명호, 1995, 이연숙, 1997, 박도순, 2001).

(1) 내용타당도(content validity)

내용타당도는 적절성과 깊은 관련이 있다. 평가도구는 일반적으로 평가하려고 하는 내용영역에 대한 질문의 표집일 수밖에 없다. 그러므로 내용타당도는 평가도구의 내용이 문항 모집단을 잘 대표하고 있는가에 의해 결정된다. 내용타당도는 그 검증방법을 검증자의 논리적 판단에 두고 있으므로 해당 분야의 전문가에 의해 검증된다. 이럴 경우 전문가의 주관에 개입될 가능성이 있고, 이를 최소한으로 줄이기 위하여 일반적으로 두 사람 이상의 전문가로 하여금 내용타당도를 판단하게 한다. 본 연구에서는 교직경력 10년 이상의 고등학교 가정과 교사 13명으로 하여금 내용타당도를 검증하였다. 내용타당도의 객관성을 높이기 위한 통계적 방법으로 문항-목표 합치도 지수 산출, 카파계수와 내용타당도 지수 산출, 기대 난이도 수준과 관찰 난이도 수준의 비교 방법을 사용할 수 있다.

(2) 준거관련 타당도(criterion related validity)

준거관련 타당도는 도구가 목적하는 준거가 될 만한 행동 특성을 정확하게 예언할 수 있는 정도를 판단하는 예언타당도와 이미 타당성이 인정된 기준 도구의 측정치와 공통성 정도를 판단하는 공인타당도로 분류할 수 있다. 준거관련 타당도는 검사 X와 준거 Y와의 상관계수로 표시된다. 상관계수가 높으면 예언의 정확성이 크고, 준거와 공통된 요인이 크다는 것을 의미한다. 그러나 본 연구에서는 준거인 전통적인 지필평가와 수행평가와의 공통요소보다는 차별성에 중점을 두어 해석할 수 있다. 즉, 수행평가가 측정하는 부분과 전통적인 평가 방법인 선발고사가 측정하는 부분은 다르므로 두 검사간의 상관성이 높지 않은 것이 수행평가의 타당성을 간접적으로 반증하게 된다. 이는 Marshall(1991, 남명호, 1995 재인용)의 이질 특성·이질 방법이 낮은 상관을 보일 경우 변별적 타당도의 증거가 된다는 이론에 기초한 것이다.

(3) 구인타당도(construct validity)

구인은 개념과 비슷한 뜻이며 인간 행동의 어느 측면을 설명하기 위해 만들어 낸 가설적 모형으로, 동기, 욕구, 흥미 등이 예가 된다. 그러므로 어떤 검

사가 특수한 개념을 측정한다고 하면 그에 관한 구인타당도를 밝혀야 한다. 구인타당도는 검사에서 나온 점수의 의미를 심리학적 개념으로 분석하는 것이다(황정규, 1972, 백용덕, 1995 재인용). 구인이라는 용어는 Cronbach와 Meel이 “심리검사의 구인타당도”란 논문에서 처음 사용되었다. 그들은 구인은 경험을 여러 범주로 조직하는 수단이며, 검사가 조작적으로 정의되지 않은 어떤 속성이나 질의 측정치로서 해석되기 위해서는 반드시 구인타당도가 입증되어야함을 주장하였다(남명호, 1995, 36).

3) 신뢰도(reliability)

신뢰도는 한 측정도구가 측정하고자 하는 내용을 정밀하게 측정하고 있는 정도를 말한다. 즉, 측정의 오차가 얼마나 적으냐는 문제이며 항상성, 일관성, 신빙성을 의미한다(박도순, 2001). 일반적으로 신뢰도는 측정결과의 일관성을 판단하는 재검사 신뢰도, 동형검사 신뢰도, 반분 신뢰도 등으로 측정한다. 수행평가의 경우 일반적인 표준화된 성취검사와 달리 측정 오차의 원인의 대부분이 평정자 판단인 경우가 많다(남명호, 1995). 그러므로 수행평가의 신뢰도 계수는 이 오차를 고려하여야 하며, 표준화 검사에서 얻은 신뢰도 계수보다 적은 것이 일반적이다. AERA, APA, and NCME(1985)의 「Standards for educational and psychological testing」은 평가도구의 신뢰도로서 단일한 신뢰도 계수를 사용하기보다는 가능한 한 다양한 방법에 의한 신뢰도를 구하도록 제안하고 있다. 본 연구에서는 수행평가의 특성을 고려하여 문항 내적 일치도, 평정자간 신뢰도, 과제간 신뢰도를 산출하였다.

III. 연구 방법 및 절차

1. 연구대상과 절차

서부 경남 농촌(면)지역 소규모 고등학교 1학년 여학생 17명을 대상으로 수행평가를 실시한 후 평가도구의 타당도와 신뢰도를 다음의 절차에 의해

검증하였다.

우선 현장 적용에 대한 충분한 준비기간 없이 적용되고 있는 가정과 수행평가 실시 현황을 1999년 9월부터 11월까지 경남지역 93개 고등학교를 대상으로 조사하였다. 이 예비조사를 통하여 수행평가도구 개발의 어려움과 검증된 평가도구에 대한 요구를 파악하였다. 이를 위하여 2000년 1월~2월에 수행평가 도구의 개발 및 수정, 2000년 3월~6월에 수행평가 도구 적용 및 평가, 2000년 6월~7월에 적용한 수행평가 도구의 검증을 하였다.

2. 측정도구

1) 수행평가 도구

(1) 서술형 수행평가 도구

서술형 수행평가 도구로 변화하는 현대가족, 부모의 양육태도, 임신과 출산의 3과제에서 한국교육과정평가원에서 개발한 각 1문항씩의 수행평가 도구를 이용하였다. 각 과제에 사용된 평가 문항의 예는 다음과 같다.

- 현대 사회의 다양한 가족의 구조 중 5가지를 제시하고 그 특징을 서술하시오.
- 부모의 양육 태도 중 3가지를 제시하고 자녀에게 미치는 영향을 서술하시오.
- 가족 계획의 의미와 목적을 설명하고 수태 조절 방법을 3가지 이상 서술하시오.

평정은 고등학교 교사 3명의 평정결과의 평균을 이용하였고 평가 기준은 아래와 같다.

과제	평가기준	상	중	하
변화하는 현대가족		8-6점	5-2점	1-0점
부모의 양육태도		8-6점	5-2점	1-0점
임신과 출산		6-5점	4-2점	1-0점

배점 및 통계 처리 시 환산 점수는 아래와 같다.

과제	배점 및 환산점수	배점	통계 처리시 환산점수
변화하는 현대가족		8점	5점
부모의 양육태도		8점	5점
임신과 출산		6점	5점

(2) 논술형 수행평가 도구

서술형 평가도구와 마찬가지로 논술형 수행평가 도구로 변화하는 현대가족, 부모의 양육태도, 임신과 출산의 3과제에서 한국교육과정평가원에서 개발한 각 1문항씩의 수행평가 도구를 이용하였다. 각 과제에 사용된 평가 문항의 예는 다음과 같다.

- 가정의 기능 중 사회화된 기능에 대하여 설명해 보고 현대 사회에서 강화되고 있는 가정의 기능과 관련하여 가정이 더욱 중요시되는 이유에 대해 논하시오.
 - 다음의 예시문을 읽고 자녀에 미치는 영향과 올바른 자녀 양육태도와 관련하여 자신의 생각을 논술하여 보시오(예시문 생략).
 - 요즘 10대 미혼모의 임신과 출산의 문제점에 대하여 생각해 보고 자신의 의견을 논술하시오.
- 평정은 고등학교 교사 3명의 평정결과의 평균을 이용하였고 평가 기준은 아래와 같다.

과제	평가기준	상	중	하
변화하는 현대가족		18-14	13-8	7-6
부모의 양육태도		18-14	13-8	7-6
임신과 출산		18-14	13-8	7-6

배점 및 통계 처리 시 환산 점수는 아래와 같다.

과제	배점 및 환산점수	배점	통계 처리시 환산점수
변화하는 현대가족		18점	5점
부모의 양육태도		18점	5점
임신과 출산		18점	5점

(3) 단답형 수행평가 도구

단답형 수행평가 도구는 연구자가 개발한 평가도구를 사용하였으며 각 과제에 사용된 평가 문항의 예는 다음과 같다.

- 사회가 복잡해지고 고도의 경쟁의식과 개인주의가 팽배하면서 과거와는 달리 더욱 중요한 기능으로 부각된 가정의 기능은 무엇인가?
- 전통사회의 자녀교육에 있어 아버지와 어머니의 역할이 구분되었으나 현대사회의 자녀교육에서 부모 모두의 역할이 중요시되고 있다. 가장 우선

적인 이유는 무엇인가?

- 임신 중 태아의 정상적 발달을 확인하고 자궁외 임신, 유산, 조산, 임신중독증 등의 이상증세를 예방하도록 하기 위해 꼭 필요한 것은 무엇인가?
 문항 수, 배점, 통계처리 시 환산 점수는 다음과 같다.

과제	점수	문항수	배점	통계처리시 환산 점수
변화하는 현대가족		6문항	6점	5점
부모의 양육태도		6문항	6점	5점
임신과 출산		8문항	8점	5점

(4) 선택형 수행평가 도구

선택형 수행평가 도구로 한국교육과정평가원에서 개발한 평가도구를 사용하였다. 각 과제에 적용되는 선택형 평가도구는 한가지 과제에 대하여 2 문항 1조로 구성하고 두 문항의 난이도의 차이를 두어 사용하였다. 한 과제에 사용된 평가 문항의 예는 다음과 같다.

- 다음은 현대 사회의 가족 구조의 특성을 설명한 것이다. 바르게 설명한 것은?(난이도 상/중)
 - ① 핵가족은 가족 간의 권리와 의무를 강조한다.
 - ② 핵가족은 가족의 문제에 있어 친척 등 외부의 영향을 적게 받는다.
 - ③ 확대가족은 모자가 가족관계의 중심이다.
 - ④ 재혼가족은 자녀양육을 혼자 담당해야 하므로 경제부담이 크다.
 - ⑤ 확대가족은 가족 스스로 의사결정을 내어 해결한다.
- 아래 설명은 현대 사회의 가족형태에 대한 것이다. 바른 것은?(난이도 중/하)
 - ① 핵가족은 부모와 미혼 자녀로 구성된다.
 - ② 편부모가족은 편부모의 재혼으로 형성된 가족형태이다.
 - ③ 현재 우리 나라의 가족형태는 확대가족이 70% 정도로 점차 증가하고 있다.
 - ④ 핵가족은 현대 사회의 대표적인 가족형태로 3세대 이상의 가족형태이다.

- ⑤ 요즈음은 사별로 인한 편부모 가족이 증가하고 있다.

2) 준거 평가도구

준거 평가도구로는 2000학년도 경상남도 교육청 주관 고입 선발고사 문항 중 가정과목 8문항을 사용하였고, 배점 8점을 5점으로 환산하여 통계 처리하였다.

3. 통계분석 방법

SAS 통계패키지 이용하여 문항의 난이도와 변별도, 평가도구의 타당도와 신뢰도를 검증하였다. 난이도와 변별도를 알아보기 위해 문항의 변별도 순수지수를 계산하였다. 타당도 중 내용 타당도는 고등학교 교사 13명에 의뢰하여 검증하였고, 준거관련 타당도는 Pearson의 적률상관 계수로, 구인 타당도는 ANOVA 분석으로 판단하였다. 신뢰도 중 문항 내적 일치도는 Cronbach α 로, 평정자간 신뢰도와 과제간 신뢰도는 ANOVA 분석으로 검증하였다.

IV. 연구 결과 및 논의

1. 검사 문항 분석

검사 문항의 난이도와 변별도 검증을 위하여 변별도 순수지수를 산출한 결과는 <표 4>와 같다.

상/하 집단으로 나누어 정답 율을 더한 결과를 난이도로 채택하였다. 난이도는 0 ~ 2사이 값이고, 1 정도면 난이도가 보통이고 2에 가까울수록 쉽고 0에 가까울수록 어렵다. 위의 표에서 난이도는 선택형이 쉽고 단답형, 서술형, 논술형 순으로 어렵게 나타났다. 단답형 문항 중 부모의 양육태도의 문항 외는 대체로 쉬운 편이었다. 변별도는 -1 ~ +1의 범위로 전체적으로 낮게 나타났다.

평가 방법별로 살펴보면 서술형과 단답형은 다른 평가 방법에 비해 상대적으로 변별도가 낮은 편이며 논술형이 가장 변별도가 낮은 것으로 나타났다.

〈표 4〉 문항 반응 이론에 의한 문항분석

		난이도		변별도	
		범 위	평 균	범 위	평 균
선택형	변화하는 현대가족	1.24~1.63	1.38	.38~.63	.51
	부모의 양육태도	1.43~1.88	1.66	-.32~.13	-.10
	임신과 출산	1.26~2.00	1.63	.00~.51	.25
서술형	변화하는 현대가족	.11~1.88	1.15	.05~.44	.22
	부모의 양육태도	0~2	1.16	0~0.5	.16
	임신과 출산	.46~1.89	1.14	-.29~.39	.15
논술형	변화하는 현대가족	0~1.04	.36	-.01~.43	.17
	부모의 양육태도	0~1.18	.41	-.14~.32	.06
	임신과 출산	0~1.39	.34	-.18~.32	.01
단답형	변화하는 현대가족	1.15~2.0	1.61	0~4	.24
	부모의 양육태도	.11~1.88	.81	.11~.33	.19
	임신과 출산	.46~1.89	1.35	-.11~.53	.13
중간고사		0~2	1.38	-.22~.51	.19

〈표 5〉 문항 변별도 범위와 언어적 표현

문항변별도 범위	언어적 표현
0.6 이상	매우 좋은 변별력
0.4~0.6	이주 좋은 변별력
0.2 이상	좋은 변별력
0.2 이하	낮은 변별력

이런 현상은 논술형을 처음 대하는 학생이 대부분 이었고 연구대상자의 학력이 낮은 이유도 있다고 사료된다. 중간고사의 변별도가 낮은 것으로 봐서 과제의 내용과도 관련이 있는 것으로 판단된다. 과제별로 살펴보면 변화하는 현대가족은 다른 과제에 비해 변별도가 상대적으로 높은 편이고 임신과 출산이 상대적으로 낮게 나타났다.

2. 타당성 검증

평가방법 및 평가 도구의 타당성을 검증하기 위해 내용타당도, 준거 관련 타당도, 구인 타당도를 검증하였다.

1) 내용 타당도(content validity)

내용타당도 검증을 위해 평균 교직경력 10년 이상

되는 고등학교 가정과 교사 13명에게 의뢰하여 서술형, 논술형, 선택형, 단답형 문항을 검토하게 하여 내용 타당성을 검증받았다. 그 결과는 다음과 같다.

- 1) 수행평가 도구에 사용한 성취기준 및 평가기준이 고등학교 1학년 수준의 성취 목표로 타당한지의 질문에 12명(93%)이 적합하다고 대답하였다.
- 2) 사용한 평가도구가 학생들의 학습상황에 도움을 줄 수 있는지의 여부에 대하여는 전 응답자가 도움이 될 것으로 응답하였다.
- 3) 예시된 평가 도구가 학교 현장의 교사가 바로 사용할 수 있는지의 실제성 여부에 대하여는 12명(93%)이 적합하다고 대답하였으나 1명은 부분적으로 수준과 시대를 감안함이 좋겠다고 대답하였다.
- 4) 예시된 평가도구의 이원목적 분류 표에 대하여는 12명(93%)이 적절함을 응답했으나 1명은 평가 도구의 내용적 측면이 약간 부족하므로 더 세분화되어야 할 것으로 응답하였다.
- 5) 예시된 평가도구가 가족생활 단원의 평가에 적절한 방법이며 학생들의 수행 행동의 평가로 적절한 방법인지 여부에 대한 질문에서 11명(84.6%)은 적절하다고 하였으나 1명은 수행평가의 경우 서술형이 적당하다고 대답하였고 1명은

이해부분이므로 지필 평가가 더 적당하다고 대답하였다.

- 6) 예시된 평가 도구가 내용특성을 포괄하기에 적절한 평가 방법인지의 여부에 대하여는 11명(84.6%)은 적절하다고 응답하였으나 2명은 단답형의 경우 암기 위주로 될 수 있어 이해력이 부족할 수 있다고 대답하였다.
- 7) 단답형과 수행평가와의 재고자 하는 내용의 동일성에 대하여는 전 응답자가 동일하다고 판단하였다. 또 선택형과 수행평가와의 재고자 하는 내용의 동일성에 대한 여부는 전 응답자가 동일하다고 판단하였고, 중간고사와 수행평가와의 재고자 하는 내용의 동일성 여부에 대하여 전 응답자가 동일하다고 판단하여 단답형, 선택형, 중간고사가 수행평가와 재고자 하는 내용으로 동일하다고 판단하였다.

2) 준거관련 타당도(criterion-related validity)

준거관련 타당도는 준거검사와의 상관계수로 나타낼 수 있다. 일반적으로 준거검사와 상관이 높으면 타당도가 높고 상관이 낮으면 타당도가 낮다고 할 수 있다. 그러나 본 연구에서는 전통적인 평가와 수행평가와의 차별성에 중점을 두어 해석하였다. 이론적 배경에 설명된 바와 같이 Marshall(1991, 남명호, 1995 재인용)의 이론에서 전통적인 지필평가와 수행평가의 경우와 같이 이질 특성·이질 방법이 낮은 상관이 높은 변별적 타당도의 증거가 됨을 근거로 해석한 것이다.

변화하는 현대가족의 과제에서 선발고사를 준거로 했을 때 선택형과 서술형은 선발고사와 유의한

상관이 존재하는 것으로 나타났다. 이는 변화하는 현대가족 과제는 중학교 교과과정에서의 선수학습이 영향을 미친 것으로 해석된다. 단답형은 선발고사와 상관이 낮았고 특히 논술형은 상관이 아주 낮게 나타났다. 논술형은 선발고사와는 다른 측면을 측정하는 방법이므로 상관이 낮게 나타난 것으로 보여진다.

부모의 양육태도 과제에서 선발고사를 준거로 했을 때 모든 형태에서 상관이 낮고 통계적으로 유의하지 않게 나타났다. 이런 현상은 중학교 가정과 교과과정에서 부모의 양육태도와 관련된 선수학습이 없었던 것과도 관련이 있을 것으로 사료된다. 선택형 문항만으로 구성된 선발고사와 다른 종류의 수행평가의 상관을 볼 때 예상한대로 상관이 낮은 것으로 나타났고 이는 수행평가의 특성에 있어서 그 타당성을 입증하는 것이 된다.

임신과 출산과제에서 선발고사를 준거로 했을 때 선택형은 통계적으로 5%수준에서 유의한 상관이 존재하는 것으로 보여진다. 그러나 서술형과 논술형, 단답형의 세 가지 방법에서 준거와 관련된 상관이 낮게 나타나므로 수행평가의 특성상 그 타당성

<표 7> 평가방법간의 상관계수(부모의 양육 태도)

	선택형	서술형	논술형	단답형
선택형	1.00			
서술형	.17	1.00		
논술형	-.13	.10	1.00	
단답형	.05	.24	.26	1.00
선발고사	.08	.21	.26	.43

사례수 : 17

<표 6> 평가방법간의 상관계수 (변화하는 현대가족)

	선택형	서술형	논술형	단답형
선택형	1.00			
서술형	.22	1.00		
논술형	.73***	.05	1.00	
단답형	.64**	.03	.40	1.00
선발고사	.54*	.48*	.03	.44

사례수 : 17, *: p <.05, **: p <.01, ***: p <.001

<표 8> 평가방법간의 상관계수(임신과 출산)

	선택형	서술형	논술형	단답형
선택형	1.00			
서술형	.04	1.00		
논술형	-.28	.24	1.00	
단답형	-.18	.62**	.21	1.00
선발고사	.48*	.20	-.02	.07

사례수 : 17, *:p<.05, **:p <.01

이 입증되었다. 이러한 현상은 선발고사의 내용 중 임신과 출산에 관한 내용이 없었고 중학교 가정과 교과과정에서 임신과 출산에 관련된 선수학습이 없었던 것으로 사료된다.

3) 구인 타당도(construct validity)

구인타당도를 확립하는 하나의 방법으로서 평가 방법들이 서로 동일한 성취영역을 측정하는 방법, 곧 수렴적 타당도를 검증하는 방법이 있다. 수렴적 타당도를 검증하려면 검사 점수의 변량성분들을 추정하고 추정된 변량요소 추정치를 가지고 상대적 해석을 위한 타당도 계수를 계산해야 한다. 이 계수는 수렴적 타당도의 근삿값이 되며 통상적인 상관적 방법을 따른다. 다시 말해 수렴적 타당도의 근삿값으

로 일반화 가능도 계수를 사용할 수 있다. <표 9>는 선택형, 서술형, 논술형, 단답형간의 호환가능성을 알아보기 위해 일반화 가능도 연구 중 각 변량성분 추정치와 이 추정치를 가지고 일반화 가능도 계수(ρ^2)를 계산한 결과를 요약한 것이다. 일반화 가능도 계수는 '관찰점수 변량에 대한 전집점수 변량' 비로 정의된다. 이를 식으로 나타내면 다음과 같다.

$$\rho^2 = \frac{\sigma^2}{\sigma_p^2 + \sigma_{p.e}^2}$$

여기서

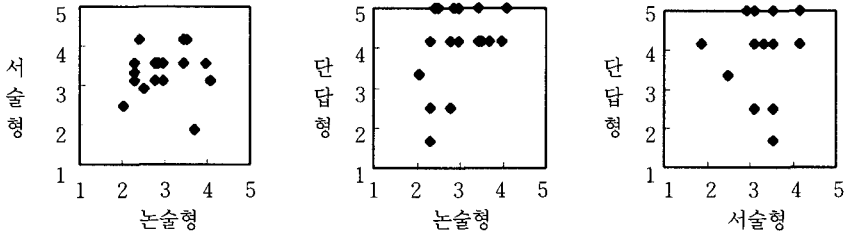
σ_p^2 : 전집점수 변량

$\sigma_p^2 + \sigma_{p.e}^2$: 관찰점수 변량

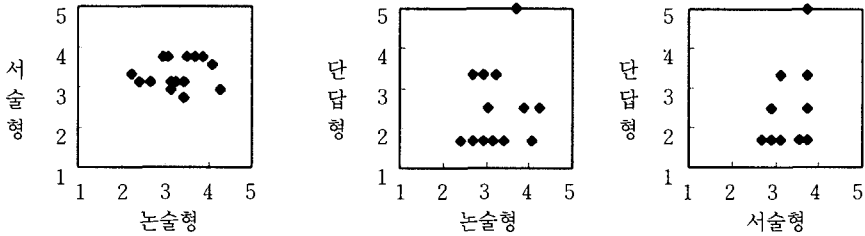
<표 9>에서와 같이 모든 평가 방법 간의 일반화

<표 9> 피험자 × 방법 일반화 가능도 연구

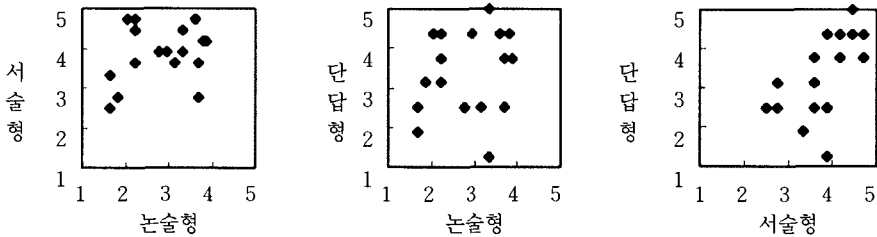
측정방법/변량성분추정치		수행평가		
		변화하는 현대가족	부모의 양육태도	임신과 출산
논술형과 선택형	피험자	.75	.00	.00
	방법	.04	.31	.74
	피험자×방법, 오차	.78	1.42	1.35
	일반화 가능도(ρ^2)	.49	.00	.00
논술형과 서술형	피험자	.02	.02	.13
	방법	.05	.00	.48
	피험자×방법, 오차	.34	.20	.42
	일반화 가능도(ρ^2)	.05	.09	.23
논술형과 단답형	피험자	.26	.18	.18
	방법	.58	.63	.11
	피험자×방법, 오차	.45	.72	.70
	일반화 가능도(ρ^2)	.36	.20	.20
서술형과 선택형	피험자	.21	.10	.04
	방법	.00	.24	.00
	피험자×방법, 오차	1.30	1.11	.95
	일반화 가능도(ρ^2)	.13	.08	.04
서술형과 단답형	피험자	.05	.11	.48
	방법	.22	.77	.08
	피험자×방법, 오차	.63	.69	.34
	일반화 가능도(ρ^2)	.07	.13	.58
선택형과 단답형	피험자	1.06	.10	.00
	방법	.18	2.01	.17
	피험자×방법, 오차	.79	1.78	1.56
	일반화 가능도(ρ^2)	.57	.05	.00



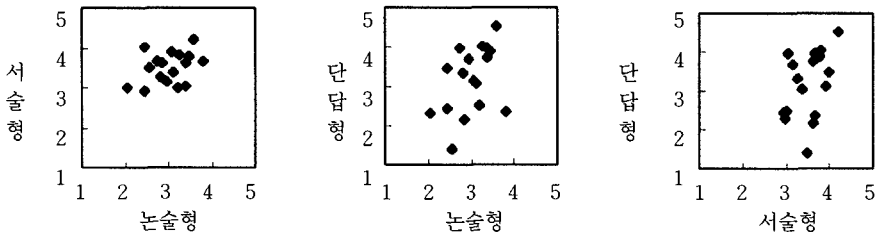
<그림 1> 수행평가 방법 간의 상관도(변화하는 현대가족)



<그림 2> 수행평가 방법 간의 상관도(부모의 양육태도)



<그림 3> 수행평가 방법 간의 상관도(임신과 출산)



<그림 4> 수행평가 방법 간의 상관도(과제 합계)

가능도는 대체로 낮은 것을 알 수 있다. 변화하는 현대가족 과제에서 논술형과 선택형의 호환성은

0.49로서 보통정도로 나타났고 논술형과 단답형에서 호환성은 0.36으로 낮은 편이다. 변화하는 현대가족

에서 논술형과 서술형의 호환성은 아주 낮다. 즉, 두 측정 방법 간에는 변량성분은 아주 낮으며 피험자와 방법 간의 상호작용 효과도 낮아 호환성이 없는 편이다($\rho^2 = 0.05$). 부모의 양육태도 과제에서는 전반적으로 평가 방법 간의 호환성이 낮게 나타났다. 임신과 출산의 과제에서도 서술형과 단답형의 호환성은 0.58로 보통정도인 것을 제외하고는 모든 평가 방법 간의 호환성이 아주 낮게 나타났다.

평가 방법 간의 호환성의 정도는 두 평가 방법 간의 상관도를 보면 분명하게 나타난다.

<그림 1>, <그림 2>, <그림 3>, <그림 4>에서 보는 바와 같이 논술형과 선택형, 논술형과 서술형, 논술형과 단답형의 상관이 대체적으로 높지 않게 나타났다. 그러므로 수행평가를 할 때는 동일한 한가지 방법만으로 피험자를 측정한다는 것은 학생들에 대한 종합적이고 전인적인 평가를 할 수 없고 단편적이고 부분적인 평가가 될 수 있어 학생을 바르게 평가할 수 없다. 따라서 수행평가에서는 다양한 방법을 적용하여야 하며 평가 과제수도 늘려야만 학생들의 전반에 관한 평가를 할 수 있음을 알 수 있다.

3. 신뢰성 검증

평가 및 검사도구의 신뢰도를 알아보기 위하여 문항 내적 일치도와 평정자간 신뢰도, 과제간 신뢰도를 구하였다.

1) 문항 내적 일치도

각 문항의 문항 내적 일치도인 Cronbach's alpha

<표 10> 문항 내적 신뢰도

	Cronbach's alpha		
	변화하는 현대가족	부모의 양육태도	임신과 출산
선택형	.62	.55	F2=1이므로 계산불가능
서술형	.70	.89	.85
논술형	.83	.69	.96
단답형	.53	.66	.64
선발고사	.51	.51	

를 구하였고 그 결과는 <표 10>과 같다.

일반적으로 수행평가도구의 신뢰도는 낮은 것으로 알려졌으나, 본 연구에서는 세 가지 과제에서 문항 내적 일치도는 선택형, 서술형, 논술형, 단답형 모두 Cronbach's alpha 0.5이상으로 신뢰도가 높은 편으로 나타났다.

2) 평정자간 신뢰도

평가 및 평가도구의 신뢰도를 측정하는 두 번째 방법으로 서술형이나 논술형과 같이 세 평정자가 평정한 경우는 문항 내적 일치도 대신 일반화 가능도를 이용한 평정자간 신뢰도를 구했다. 평정자간 신뢰도는 한 평정자의 평정점수를 어느 정도 일반화 할 수 있는지 (신뢰할 수 있는지)를 보여주는 지수이다.

본 연구에서는 서술형과 논술형에 대해 세 명의 평정자 효과를 알아보기 위해 두 국면 G 연구(P×R모형) 설계를 시행하였다. G연구에 의한 변화하는 현대가족 서술형의 평정자간 신뢰도를 구하는 절차를 예시하면 다음과 같다.

첫째, 변량분석을 한다(표 11 참조).

둘째, 변량분석을 하여 얻은 값을 가지고 각각 피험자 효과, 평정자 효과, 피험자×평정자 상호작용 효과의 변량분석을 계산한다(표 12 참조).

<표 11> 피험자×평정자 설계 변량분석표 (변화하는 현대가족 서술형의 경우)

효과	자유도	SS	MS
피험자	16	16.52	1.03
평정자	2	27.02	13.51
피험자×평정자	32	10.21	.31
전체	50	53.75	

<표 12> 변량성분 요소표

효과	자유도	변량성분 요소
피험자	16	.24
평정자	2	.78
피험자×평정자	32	.31

<표 13> 피험자×평정자 일반화 가능성도 연구

측정방법/변량성분추정치		수행평가		
		변화하는 현대가족	부모의 양육태도	임신과 출산
서술형	피험자	.24	.11	.41
	평정자	.78	.01	.03
	피험자×평정자, 오차	.31	.06	.22
	일반화 가능성도(ρ^2)	.44	.64	.65
논술형	피험자	.30	.21	.62
	평정자	.54	.22	.02
	피험자×평정자, 오차	.25	.31	.06
	일반화 가능성도(ρ^2)	.55	.40	.91

<표 14> 피험자×과제 일반화 가능성도 연구

변량성분 추정치	선택형	서술형	논술형	단답형
피험자	.00	.06	.10	.48
과제	.05	.06	.02	1.00
피험자×과제, 오차	2.28	.25	.34	.72
일반화가능도(ρ^2)	.00	.19	.23	.40

이와 같은 방법으로 나머지 평가 방법의 변량 성분과 신뢰도를 계산하여 정리한 것이 <표 13>이다.

평정자간 신뢰도는 0.40~0.91로 나타났으며 과제 중에서 임신과 출산과제의 신뢰도가 높게 나타났다. 임신과 출산 과제에서 변량성분을 살펴보면 평정자의 변량성분은 거의 0에 가까운데 이는 평정자간의 평정의 차이를 거의 무시해도 좋다는 것을 의미한다. 이는 특별한 훈련을 받지 않은 교사들의 평정을 신뢰할 수 있다는 것을 나타낸다.

3) 과제간 신뢰도

과제간 신뢰도는 한 과제의 점수를 가지고 어느 정도 일반화할 수 있는지를 보여주는 지수라고 할 수 있다. 과제간 신뢰도를 측정하기 위해 피험자×일반화 가능성도를 계산한 결과는 <표 14>와 같다.

<표 14>에 따르면 각각의 평가도구의 신뢰도는 높으나(<표 10> 참조) 과제간 신뢰도는 대부분의 평가 방법에서 낮게 나타났다. 특히 선택형의 과제간 일반화 가능성도 계수는 0의 수치를 나타냄으로써 과제의 표집에 따른 변산성(sampling variability)이 상

당히 크다는 것을 알 수 있다. 선택형으로 측정하려면 동일한 내용에 대한 문항 수를 늘려 피험자의 능력을 측정하여야 한다. 이러한 사정은 정도의 차이는 있지만 다른 평가방법에서도 마찬가지로 나타났다. 이는 과제 수를 더 늘여야 할 필요가 있음을 말해 준다.

V. 결론 및 제언

교육현장에 1999학년도부터 교육부 정책으로 도입되어 실시하고 있는 수행평가는 그 우수성이 높게 주장되고 있다. 그러나 일선 교사들은 수행평가 도구의 부족함을 느끼고 개발된 평가도구도 그 신뢰성과 타당성에 대한 검증된 것이 없기에 시행하는데 많은 어려움을 겪고 있다.

본 연구는 농촌지역의 고등학교 한 학급 여학생 17명을 대상으로 변화하는 현대 가족, 부모의 양육태도, 임신과 출산 영역에 대한 서술형, 논술형, 선택형, 단답형 수행평가 도구의 질을 입증하기 위해

평가 방법 및 평가 도구의 타당성과 신뢰성을 검증하였다. 연구 결과는 다음과 같다.

1. 타당도 검증 결과

첫째, 고등학교 교사 13명에게 의뢰하여 검토한 결과 내용 타당도가 검증되었다.

둘째, 경상남도 교육청 주관 2000학년도 고입 선발고사의 가정교과 점수를 준거로 하여 검증한 결과 선택형, 서술형, 단답형은 선발고사에 비교적 상관성이 있는 것으로 나타났다. 논술형은 매우 낮은 상관관계를 나타내 변별적 타당도의 원리에 의해 해석하여 보면 논술형이 수행평가로서의 특성을 가장 잘 나타내고 있음을 알 수 있었다.

셋째, 평가 방법 간의 일반화 가능성을 알아보기 위해 구인타당도 검증을 한 결과 평가 방법 간에 호환성이 높지 않게 나타나 평가 방법 간 일반화 가능성이 낮게 나타났다.

2. 신뢰성 검증결과

수행평가의 신뢰성은 일반적으로 낮은 것으로 인식되고 있으나 본 연구에 사용된 평가 도구의 Cronbach's α 는 0.5~0.9로 비교적 높게 나타났다. 평정자간 신뢰도는 0.40~0.91로 과제간에 차이를 보였는데 임신과 출산 과제는 높은 편이었고 변화하는 현대 가족 과제는 높은 편은 아니었다. 과제간 신뢰도는 대체로 낮았다.

이상의 연구 결과를 바탕으로 다음과 같은 결론을 얻었다.

첫째, 평가 방법 간의 일반화 가능성과 평가 방법 간의 상관계수가 낮으므로 다양한 수행평가 방법의 실시가 요구된다.

둘째, 과제간 일반화 가능성이 낮게 나타났으므로 과제 수를 늘려 실시하여야만 수행평가의 신뢰도를 높일 수 있다.

위와 같은 결론은 본 연구의 예비조사에서도 나타난 바와 같이 교육현장에서 학기별로 평균 2번 정도 수행평가를 실시하고 그 방법도 서술형이나

보고서형 등으로 국한되어 있는 점을 감안한다면 수행평가 실시 운영에 문제점이 있다고 말할 수 있다. 그러므로 중간고사와 학기말 고사의 경우와 같이 한 학기 2번의 평가를 위한 수행평가로 생각하기보다는 수업 과정 중에 수시로 간단한 수행평가 방법이라도 실시하는 것이 다양한 수행평가 방법과 과제 수의 증가에 대한 요구를 동시에 충족시킬 수 있는 길일 것이다.

끝으로 본 연구는 다음과 같은 제언을 하고자 한다.

첫째, 수행평가 도구의 개발은 모든 교사들이 필요로 하고 있다. 그러나 평가 도구를 개발하는 데 많은 어려움을 느끼므로 수행평가의 정착을 위해서는 타당화 검증을 거친 평가 도구의 공유화를 위한 노력을 기울여야 한다.

둘째, 도시와 농촌지역 학생을 대상으로 한 비교 연구, 다인수와 소인수 학급을 대상으로 한 수행평가의 타당화 연구가 요구된다.

셋째, 일반적으로 일선에서 수행평가는 교육 영역 별로 실시하는 경우가 많은데 특정 영역에 편중하지 말고 여러 영역에 골고루 수행평가를 실시할 필요가 있다.

본 연구는 '수행평가는 그 장점을 최대한 살려 점수를 내기 위한 수행평가가 아니라 교수·학습활동의 개선과 학습자들의 고등사고 기능 파악, 그리고 개별학습 신장을 위해 적용되어야 하며 결과만 중시할 것이 아니라 과정과 결과 모두를 중시하여 적용되어야 한다'는 방향 제시를 위한 기초 자료 제공에 그 의의가 있다.

■ 참고문헌

- 교육부(1997). 실과(기술·가정)교육과정. 교육부 고시 제1997-15호(별책10).
- 교육부(1998a). 수행평가 그 현장적용 실제. 교육월보, 204, 44-79.
- 교육부(1998b). 수행평가의 이해. 교육부 교사연수자료.
- 교육부(1998c). 수행평가의 이해. 교육홍보자료.

- 교육부(1998d). 새로운 대학입학제도와 교육비전 2002: 새학교 문화 참조.
- 국립교육 평가원(1999). 수행평가의 이론과 실제. 대한 교과서 주식회사.
- 김기춘, 소경희(1999). 수행평가 정착을 위한 교육과정 및 수업운영방안. 한국교육과정 평가원.
- 남명호(1995). 수행평가의 타당성 연구(과학 실기평가, 실험보고서 평가, 컴퓨터시뮬레이션 평가의 비교). 고려대학교 박사학위 논문.
- 박도순(2001). 교육연구방법론. 문음사.
- 백순근(1997). 개인차 존중의 평가 방법. 교육월보, 186, 40-43.
- 백순근(1998). 교육평가체제의 재 구조화(수행평가를 중심으로). 교육월보, 204, 44-48.
- 백순근(1999a). 중학교 각 교과별 수행평가의 이론과 실제. 도서출판 원미사.
- 백순근(1999b). 수행평가 정착을 위한 교육평가 실천 방안. 한국교육과정 평가원.
- 백순근(1999c). 수행평가의 이론적 기초 및 실천 방안. 한국교육과정 평가원.
- 백용덕(1995). 교육통계. 형설출판사.
- 부재열(1999). 서술형 및 논술형 수행평가 (경남 교원 연수교재). 경남 교원연수원.
- 석문주(1999). 학습을 위한 수행평가. 교육과학사.
- 성태제(1999). 교육평가의 새로운 동향 (경남 교원연수교재). 경남 교원연수원.
- 안상대(1998). 수행평가의 적용과 실시상의 문제. 강원교육.
- 윤인경 외(1999). 중학교 가정교과의 수행평가의 이론과 실제. 교학사.
- 왕석순 외(1999). 고등학교 가정-국가 교육과정에 근거한 성취기준 및 평가도구. 한국교육과정 평가원.
- 이상봉(1998). 기술과 수행평가의 실제. 교육월보, 204, 68-73.
- 이연숙(1997). 가정과 교수법. 신광출판사.
- 이종성(1996). 교육연구의 설계와 자료분석. 교학연구사.
- 이춘식 외(1999). 중학교 기술 가정과 수행평가 시행 방안 및 자료개발 연구. 한국교육과정평가원.
- 이춘식 외(2000). 제 7차 교육과정에 따른 성취기준과 평가기준 개발 연구-중학교 기술 가정-. 한국교육과정평가원.
- 조필교 외(1999). 중학교 가정과 수행능력 평가방법 및 도구 개발. 연구보고서 RR97-IV-10. 한국 교원 대학교 부설 교과 교육공동 연구소.
- 최봉규(1998). 수행평가에서의 평정자내 및 평정자간 신뢰도 분석. 경남대학교 석사학위 논문.
- 한국교육과정평가원(1998). 국가교육과정에 근거한 평가기준 및 도구개발 연구. 고등학교 영어. 연구보고 RRE98-3-11.
- 한국교육과정평가원(1999). 수행평가 정책시행 실태 분석과 개선대책 연구. 연구보고 CRE 99-2.
- 허경철(1999). 수행평가의 올바른 이해. 한국교육과정 평가원.
- AERA, APA & NCME (1985). *Standards for educational and psychological testing*. American Psychological Association, Inc.
- Allen, M. J. & Yen, W .M. (1979). *Introduction to measurement theory*. CA:Brooks/Cole Publishing Company.
- Baron, M. A. & Boschee, F. B. (1995). *Authentic assessment: The key to unlocking student success*. PA: Technomic.
- Berlak, H. (1992). Toward the development of a new science of a new science of educational testing and assessment. In H. Berlak et al. *Toward a new science of educational testing & assessment*. NY : State University of New York Press.
- Bulger, S. E. & Burger, D. L. (1994). Determining the validity of performance-based assessment. *Educational Measurement: Issues and Practice*. 13(1).
- Glaser, R., Lesgold, A., & Lajoie, S. (1987). Toward a cognitive theory for the measurement of achievement. In R. R. Ronning, J. A. Glover, J. C. Conoley, & J. C. Witt (Eds.) *The influence of cognitive psychology on testing*. Hillsdale, NJ:

- Erlbaum, 1-8.
- Gronlund, N. E. & Linn, R. L. (1990). *Measurement and evaluation in teaching* (Sixth edition). NY: Macmillan Publishing Company.
- Marshall, J. E. (1991). *Construct validity of multiple choice and performance-based assessments of Basic Science Process Skills: A multitrait-multimethod analysis*. A paper presented at the annual meeting of the National Association for Research in Science Testing. ERIC ED 331 719.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18, 5-11.
- Wiggins, G. (1993). Assessment : Authenticity, context, and validity. *Phi Delta Kappan*, Nov., 200-214.