



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

석사학위 청구 논문

학생 건강체력평가(PAPS)의  
심폐지구력 검사종목 간 동등화

지도교수 : 김 미 예

제주대학교 교육대학원

체육교육전공

김 경 오

2015년 8월

학생 건강체력평가(PAPS)의  
심폐지구력 검사종목 간 동등화

지도교수 김 미 예

김 경 오

이 논문을 교육학 석사학위 논문으로 제출함

2015년 7월

김경오의 교육학 석사학위논문을 인준함

심사위원장 류 지 강 (인) 

위 원 김 영 표 (인) 

위 원 김 미 예 (인) 

제주대학교 교육대학원 체육교육전공

2015년 8월

# 목 차

I. 서 론	1
1. 연구의 필요성 및 목적	1
2. 연구문제	4
3. 연구의 제한점	4
4. 용어 정의	4
II. 이론적 배경	5
1. 학생체력검사의 변천과정	5
2. 학생 건강체력평가의 이해	5
1) 학생 건강체력평가의 개념	5
2) 학생 건강체력평가의 특징	7
3) 학생 건강체력평가의 구성내용 및 세부항목	7
4) 학생 건강체력평가의 평가기준	9
3. 학생 건강체력평가 측정 종목 및 방법	9
1) 심폐지구력	9
2) 유연성	10
3) 근력 근지구력	11
4) 순발력	12
5) 체지방	12
4. 동등화의 이해	13
1) 동등화의 개념	13
2) 동등화 방법	14
3) 동등화의 조건	15
4) 동등화 오차 산출 및 타당성 검증	16
5. 선행연구	17
1) 동등화 관련 연구	17
2) 학생 건강체력평가 관련 연구	18

III. 연구방법 .....	20
1. 연구대상 .....	20
2. 측정방법 .....	20
1) 왕복 오래달리기 .....	20
2) 오래달리기 걷기 .....	21
3) 스텝검사 .....	21
3. 동등화 절차 .....	22
1) 검사의 유사성 검증 .....	23
2) 변환등급 산출 .....	23
3) 동등화 오차산출 .....	24
4. 자료분석 .....	24
IV. 연구결과 .....	25
1. 기술통계 .....	25
2. 검사종목 간 평가등급차이 .....	26
1) 빈도분석 .....	26
2) 일치도 검증 .....	27
3. 동등화 .....	28
1) 검사의 유사성 검증 .....	28
2) 백분위 점수 산출 .....	28
3) 등급 변환 .....	30
V. 논의 .....	32
VI. 결론 .....	35
참고문헌 .....	36
<Abstract> .....	41

## 표 목차

표 1. 학생체력검사제도의 변천과정 .....	6
표 2. 학생 건강체력평가(PAPS)의 세부 측정내용 .....	8
표 3. PEI 계산 공식 .....	10
표 4. 연계화 방법의 종류와 활용 .....	13
표 5. 동등화 방법의 종류 .....	14
표 6. 연구대상의 신체적 특성 .....	20
표 7. 심폐지구력 검사종목 평가기준표 .....	22
표 8. 심폐지구력 검사종목별 기술통계 .....	25
표 9. 심폐지구력 검사종목 간 평가등급의 빈도분석 .....	26
표 10. 심폐지구력 검사결과의 평가등급의 일치도 .....	27
표 11. 심폐지구력 검사종목 간 상관관계 .....	28
표 12. 변환등급 .....	31

## 그림 목차

그림 1. PAPS필수평가 항목 .....	7
그림 2. 왕복 오래달리기 .....	21
그림 3. 오래달리기 걷기 .....	21
그림 4. 스텝 검사 .....	22
그림 5. 5학년 남학생의 심폐지구력 검사종목 등급의 백분위 .....	29
그림 6. 5학년 여학생의 심폐지구력 검사종목 등급의 백분위 .....	29
그림 7. 6학년 남학생의 심폐지구력 검사종목 등급의 백분위 .....	30
그림 8. 6학년 여학생의 심폐지구력 검사종목 등급의 백분위 .....	30

<국문초록>

## 학생 건강체력평가(PAPS)의 심폐지구력 검사종목 간 동등화

김 경 오

제주대학교 교육대학원

체육교육전공

지도교수 김미예

2009년 건강관련체력 종목 위주로 구성된 학생 건강체력평가(Physical Activity Promotion System; PAPS)가 개발되었다(교육과학기술부, 2009). 기존 학생체력검사와 2009년 새롭게 개발된 학생 건강체력평가(PAPS)의 가장 큰 차이점은 각 체력 항목별로 복수의 측정 종목이 제공되어 그 중 한 종목을 선택하여 측정할 수 있다는 것이다. 이처럼 요인별 단일 종목이 아닌 복수의 종목이 제공되면 해당 학교의 환경여건에 따라 가장 적합한 검사를 선택하여 측정할 수 있다는 장점이 있는 반면, 학교마다 또는 학년별로 서로 다른 종목을 선택할 경우 검사결과를 상호 비교할 수 없다는 단점이 있다. 특히 선택 종목 간 난이도가 다를 경우 평가의 공정성에 문제가 발생할 수 있다. 예를 들어 심폐지구력은 왕복 오래달리기, 오래달리기 걷기, 스텝검사 중 하나의 종목을 선택할 수 있는데, 만약 이 종목들의 난이도가 달라 쉬운 종목을 선택한 학생과 어려운 종목을 선택한 학생 간에 체력 차이가 발생한다면 공정한 평가가 어려울 것이다. 따라서 서로 다른 검사에서 산출된 점수들을 상호 비교할 수 있고, 어느 종목을 선택하느냐에 관계없이 개인의 체력수준을 일관성 있고 공정하게 평가할 수 있는 동등화(equating)가 필요하다. 이 연구의 목적은 학생 건강체력평가(PAPS)의 심폐지구력 선택종목인 왕복 오래달리기, 오래달

리기 걷기, 스텝검사가 서로 동일한 구인, 즉 심폐지구력을 측정한다는 가정 하에 각 종목의 난이도를 확인하고 종목 간 동등화를 실시하여 보다 일관되고 공정한 평가 지침을 제공하는 것이다. 제주도 소재 6개 초등학교 5학년 남학생 131명과 여학생 112명, 6학년 남학생 121명과 여학생 104명, 총 468명을 대상으로 학생 건강체력평가(PAPS) 종목 중 심폐지구력 검사에 해당하는 왕복 오래달리기, 오래달리기 걷기, 스텝검사를 실시하였다. 검사종목 간 평가등급의 차이를 검정하기 위해 빈도 분석을 실시한 결과, 스텝검사, 왕복오래달리기, 오래달리기 걷기 순으로 평가등급이 높게 나타났고 통계적으로도 유의한 차이가 있었다. 또한 세 가지 검사 종목의 평가등급 일치도를 검증한 결과, 일치도 값이 .16~.44로 나타나서 종목 간 일치성이 매우 낮은 것으로 나타났다. 동등화를 실시하기 위해 심폐지구력 검사 종목이 서로 유사한 구인을 측정하는지 확인한 결과, 왕복 오래달리기와 오래달리기 걷기 간 상관( $r$ )은 비교적 높은 반면, 스텝검사와 나머지 두 종목 간 상관은 낮게 산출되었다. 검사가 서로 같은 구인을 측정하지 않는다면 동등화가 불가능하지만 전통적으로 스텝검사가 심폐지구력을 측정하는 검사로 인정받고 있고 학생 건강체력평가(PAPS) 심폐지구력 검사로 선정되어 있으므로 동일한 구인으로 판단하여 동등화를 실시하였다. 왕복 오래달리기 검사를 기준으로 동등화를 실시한 결과에 의하면 선형 동등화와 동백분위 동등화 결과는 비교적 유사한 것으로 나타났다. 또한 오래달리기 걷기는 전반적으로 등급이 상향 조정되었고 스텝검사는 하향 조정되었다. 이 연구에서 스텝검사가 다른 두 검사보다 난이도가 낮은 것으로 나타났는데 이러한 결과는 스텝 높이 때문인 것으로 판단된다. 즉, 학생 건강체력평가(PAPS)의 스텝 높이가 선행연구에서 제시된 연령별 스텝 높이보다 지나치게 낮게 설정되어 있었다. 스텝검사의 특성상 스텝 높이가 측정결과에 많은 영향을 미치기 때문에 학생 건강체력평가(PAPS)의 스텝 높이가 조정될 필요가 있다. 본 연구에서는 학생 건강체력평가(PAPS)의 심폐지구력 선택종목 간 동등화만을 실시하였으나 향후 연구에서는 유연성, 근력 등 다른 항목의 선택종목 간 동등화가 이루어져야 할 것이다. 또한 초등학교 뿐만 아니라 중, 고등학생을 대상으로 보다 많은 표본을 수집하여 평가등급의 동등화 연구가 이루어져야 할 것이다.

# I. 서론

## 1. 연구의 필요성 및 목적

학생체력검사 제도는 1951년 학생들의 체력을 육성하기 위한 목적으로 문교부령에 의해 처음 시행되었다. 당시 검사 종목은 신체능력을 측정하기 위한 폐활량, 배근력, 악력 검사와 운동능력을 측정하기 위한 달리기, 오래달리기, 턱걸이(남), 매달리기(여), 나르기, 던지기, 넓이뛰기 검사로 구성되었다. 1962년에는 검사종목 중 오래달리와 나르기 검사가 제외되었고, 1972년부터는 고등학교 입학시험에 체력검사 결과가 반영되기 시작하였다. 1979년에는 100m 달리기, 제자리 멀리뛰기, 윗몸일으키기, 턱걸이(남), 팔굽혀 매달리기(여), 던지기 검사로 종목이 변경되었다. 1993년에 고등학교 입학시험에 체력검사 반영 제도가 폐지되었고, 1999년에는 던지기과 턱걸이(남), 오래 매달리기(여) 종목이 제외되고 50m 달리기, 오래달리기 걷기 종목이 새롭게 추가되었다(교육과학기술부, 2009).

그동안 학생체력검사가 학생들의 체력을 육성하기 위한 목적으로 시행되었지만 단순히 체력을 측정하여 기록하는데 그치고 학생들의 체력과 비만을 관리하는 데는 한계가 있었다. 이러한 문제점을 보완하고자 2009년 건강관련체력 종목 위주로 구성된 신개념 학생 건강체력평가(Physical Activity Promotion System; PAPS)가 개발되었다(교육과학기술부, 2009). PAPS는 학생의 체력 뿐만 아니라 비만까지 평가할 수 있는 시스템으로 학생 스스로 체력 및 건강관리가 가능하도록 고안되었다. 측정항목은 심폐지구력, 순발력, 근력·근지구력, 유연성, 체지방의 5개이고, 각 항목별로 2~3개의 검사종목이 제공되어 그 중 한 종목을 선택하여 측정할 수 있다. 평가등급은 아주 높음, 높음, 보통, 낮음, 아주 낮음의 1등급에서 5등급으로 구분되어 있다(교육과학기술부, 2009).

기존 학생체력검사와 2009년에 개발된 PAPS의 가장 큰 차이점은 각 체력 항목별로 복수의 측정 종목이 제공되어 그 중 한 종목을 선택하여 측정할 수 있다는 것이다. 예를 들어 심폐지구력은 왕복 오래달리기(PACER), 오래달리기 걷기, 스텝 검사의 3종목 중 한 종목을 선택하여 측정할 수 있고, 유연성은 종합유연성, 앉아 윗몸앞으로굽히기 중 한 종목, 그리고 근력근지구력은 팔굽혀펴기, 윗몸 말아올리기, 악력 중 한 종목을 선택하여 측정할 수 있다. 또한 순발력은 50m 달리기, 제자

리 멀리뛰기 중 한 종목, 체지방은 BMI와 체지방율(%body fat) 중 하나의 종목을 선택하여 측정할 수 있다.

이처럼 각 학교의 환경여건에 따라 여러 종목 중 가장 적합한 검사를 선택하여 측정할 수 있다는 장점이 있는 반면, 동일한 요인을 측정해 주는 검사들이지만 검사 결과를 상호 비교할 수 없다는 단점이 있다. 특히 선택 종목 간 난이도가 다를 경우 평가의 공정성에 문제가 발생할 수 있다. 예를 들어 심폐지구력은 왕복 오래 달리기, 오래달리기 걷기, 스텝검사 중 하나의 종목을 선택할 수 있는데, 만약 이 종목들의 난이도가 달라 쉬운 종목을 선택한 학생과 어려운 종목을 선택한 학생 간에 체력 차이가 발생한다면 공정한 평가가 어려울 것이다. 따라서 서로 다른 검사에서 산출된 점수들을 상호 비교할 수 있고, 어느 종목을 선택하느냐에 관계없이 개인의 체력수준을 일관성 있고 공정하게 평가할 수 있는 방안이 필요하다.

동등화(equating)란 한 검사 결과를 다른 검사 결과와 비교하기 위해 두 검사를 동일 척도로 전환하는 것을 의미한다. 동등화는 동일한 특성을 측정하기 위해 고안된 다양한 유형의 검사들간의 난이도를 조정하고 검사결과를 직접 비교하기 위한 목적으로 여러 연구에서 활용되고 있다. 예를 들어 검사보안의 이유로 검사 유형이 매년 달라지는 대학입학 선발고사의 난이도를 조정하거나, 국가수준의 학업성취도를 학년별로 비교하고 싶을 때 동등화 과정을 적용하여 직접 비교할 수 있다. 우리나라의 경우 대학수학능력 시험에서 선택형 과목간의 난이도를 확인하고 공정성을 보장하기 위하여 사후 검증의 일환으로 동등화 방법이 응용되고 있다(김현철, 1999, 2000; 남현우, 2003).

그동안 체육분야에서도 동등화와 관련된 연구가 이루어졌는데, 고기환, 강상조(1997)는 남자고등학생을 대상으로 세 가지 배구 서브인 언더 서브, 사이드 서브, 프론트 서브 유형에 대한 기능검사를 실시하고 문항반응이론을 적용하여 동등화 및 시행 별 차이를 비교하였다. 강상조(1999)의 연구에서는 세 가지 상이한 윗몸일으키기 자세를 반복적으로 검사하여 성별, 연령별 검사차별기능을 검증하고 이들 검사로부터 얻은 점수들을 단일집단 설계를 이용하여 수평동등화와 동백분위 동등화한 후 그 결과를 비교하였다. 또한 남덕현(2002)은 남녀 고등학생을 대상으로 오래달리기 800m, 1000m, 1200m, 1500m 종목 간 동등화를 실시하였다.

동등화는 동등점수의 상호교환성이라는 장점을 지니고 있지만 대칭성(symmetry), 집단불변성(population invariance), 공정성(equity) 등과 같은 강한 통계적 가정을 요구하는 제한점이 있다(Kolen & Brennan, 2004). 체력검사의 경우 각 종목이 동일한 구인을 측정해주고 있지만 각 검사들의 측정 부위나 측정 단위가 다르기 때문

에 동등화 방법에 주의가 요구된다. 예를 들어 악력과 배근력은 모두 근력을 측정하는 종목이지만 서로 다른 부위의 체력 요소를 측정하기 때문에 동등화 할 수 있는 유사 구인으로 간주하는 데는 논쟁의 소지가 있다. 그러나 심폐지구력 측정종목인 왕복 오래달리기, 오래달리기 걷기, 스텝검사는 유연성이나 근력·근지구력 측정 종목에 비해 비교적 유사한 구인을 측정한다고 할 수 있다. 유사 구인의 경우 동등화 과정과 동일한 설계 및 통계적 조정 방법을 적용할 수 있기 때문에(김아영, 이규민, 반재천, 차정은, 2004) PAPS의 심폐지구력 검사들은 동등화가 가능하거나 최소한, 동등화 방법에 적용되는 설계를 이용하여 연계화(linking)를 실시하는 것이 가능하다.

PAPS의 개발과정에서 선택종목 간 난이도 차이가 발생하는 것을 고려하여 기준 설정에 있어 동등화 연구를 실시하였다고 보고하고 있지만(교육과학기술부, 2009), 단순히 “종목 간 기준이 비교적 일치하도록 수정되었다”고 밝히고 있을 뿐, 구체적으로 어떤 절차와 방법을 통해 동등화를 수행하였는지 확인하기 어렵다. 최우영(2009)은 PAPS 종목 간 체력급수 차이를 분석한 결과, 평가등급에 차이가 있는 것을 확인하였고, 특히 김일곤(2011)의 연구에서는 심폐지구력 검사인 스텝검사 평가등급과 다른 종목의 평가등급 간에 유의한 차이가 나타났다.

이 연구의 목적은 PAPS의 심폐지구력 선택종목인 왕복 오래달리기, 오래달리기 걷기, 스텝검사가 서로 유사한 구인, 즉 심폐지구력을 측정한다는 가정 하에 각 종목의 난이도를 확인하고 종목 간 동등화를 실시하여 보다 공정한 평가 지침을 제공하는 것이다.

## 2. 연구문제

- 1) 학생 건강체력평가의 심폐지구력 선택종목 간 평가등급 결과에 차이가 있는가?
- 2) 학생 건강체력 평가의 심폐지구력 선택종목들은 동일한 구인(construct)을 측정하는가?
- 3) 학생 건강체력평가의 심폐지구력 선택종목의 선형 동등화와 동백분위 동등화 결과에 차이가 있는가?
- 4) 학생 건강체력평가 심폐지구력 선택종목의 동등화 전후 평가등급에 차이가 있는가?

## 3. 연구의 제한점

이 연구에 참여하는 연구대상은 제주도에 거주하는 초등학교생으로 한정하였다.

## 4. 용어 정의

- 1) 건강관련체력  
일상생활을 영위하는데 필요한 신체능력을 의미하며 건강관련체력을 구성하는 요인은 심폐지구력, 근력·근지구력, 유연성, 신체구성이다.
- 2) 심폐지구력  
인체의 순환계와 호흡계가 지속적으로 신체활동이 가능하도록 산소를 제공하는 능력을 나타낸다.
- 3) 동등화  
동등화(equating)란 연계화(linking)의 한 형태로, 동일한 구인을 측정하는 검사가 다양한 유형으로 제작되고 실시되었을 때, 서로 다른 검사유형의 점수를 공통 척도로 변환시키는 것을 의미한다(남현우, 2001).
- 4) 연계화  
하나의 검사 결과를 다른 검사 결과와 비교할 수 있도록 만들어 주는 과정으로, 동등화와는 다르게 서로 다른 구인을 측정하는 검사결과를 직접비교가 가능하다는 차이점이 있다(Linn, 1993; Mislevy, 1992).

## II. 이론적 배경

### 1. 학생체력검사의 변천과정

학생체력검사는 1951년 문교부령 제 15호에 의해 학교신체검사규정으로 처음 시행되었다. 1968년 이후, 청소년의 체력강화와 국방체력육성을 위하여 국제 체력검사 표준화 위원회(ICSPF)에 의해 작성된 표준체력 검사를 근거로, 1972년 문교부령 제 294호에 의해 일반체력검사와는 다른 학생체력검사를 시행하게 되었다(김기학 등, 2007). 또한 1973년부터 대학 입시에 체력장 성적을 반영하게 되어 운동 능력을 진단하는 체력검사가 입시제도의 변화로 본래의 목적과는 다르게 선발고사를 대비하여 요령을 중시하게 되었다. 따라서, 1992년 상급학교 내신 체력 검사 제도는 폐지하여 일반 학생 체력검사로 전환되어, 오늘에까지 체력검사가 실시되고 있다(문화체육부, 1992).

학생체력검사는 1971년 1차 개정되어 2006년 1월까지 총 7차에 걸쳐 개정되었으며, 2006년 교육인적자원부는 다시 교육인적자원부령 제 873호에 의해 학생건강검사규칙으로 법 규정을 재정비하여, 2009년에 신개념 PAPS의 도입으로 변화를 도모하였다. <표 1>은 학생체력검사제도의 변천 과정과 변경된 내용을 요약한 것이다.

### 2. 학생 건강체력평가의 이해

#### 1) 학생 건강체력평가의 개념

학생 건강체력평가(PAPS)는 초, 중, 고등학교 학생체력검사 제도를 전면적으로 개정하여 새롭고 선진화된 체력평가로서 건강체력과 비만, 그리고 선택적 심폐지구력 정밀평가와 자기신체평가, 자세평가와 같이 학생들의 신체활동과 관련된 종합적인 평가가 이루어지고 이를 토대로 신체활동처방이 제공되는 종합 평가시스템을 의미한다(교육과학기술부, 2009). 기존의 체력검사가 측정과 등급화에 초점을 두었다면 PAPS는 학생의 건강체력평가 뿐만 아니라 관리가 가능하도록 개발하였다. 평가항목은 필수항목과 선택항목으로 나뉘며 학생 개인의 체력수준별 운동처방이 이루어진다. 체력검사와 관련된 모든 자료들은 교사가 온라인으로 입력하고 학생과 학부모는 장소에 구애받지 않고 확인할 수 있으며 건강과 체력, 신체처방에 대한 자료를 제공하여 체계적인 건강관리를 이룰 수 있도록 제작되었다(김세정, 2012).

표 1. 학생체력검사제도의 변천과정(교육과학기술부, 2009)

연도	학생체력검사 실시 연혁	학생체력검사 실시종목
1951	· 학교신체검사규정 - 문교부령 제 15호 의거 일반체력검사 시행	신체기관 능력 ·3개 항목: 폐활량, 배근력, 악력
		운동능력 ·5개 종목 - 달리기: 50m/100m/800m/1000m - 턱걸이(남)/매달리기(여) - 넓이뛰기 - 던지기 - 나르기: 50m(15kg)/100m(40kg)
1962	· 문교부령 제 112호 의거 일부 개정	신체기관 능력 ·3개 항목: 폐활량, 배근력, 악력
1972	· 문교부령 제 294호 의거 9개 항목 체력장 시행 - 일반체력검사 시행 유지 - 고입 내신 체력검사 시행(8개 종목)	운동능력 ·4개 종목 - 달리기: 50m(11세 이하)/100m(12세 이상) - 턱걸이(남), 매달리기(여) - 넓이뛰기 - 던지기
		일반체력 검사 ·9개 종목 - 턱걸이(남), 매달리기(여) - 윗몸 앞으로굽히기 - 도움닫기 멀리뛰기 - 달리기: 100m - 왕복달리기 - 던지기 - 오래달리기: 600m(초), 800m(여;중·고), 1,000m(남;중·고) - 악력
1973	· 대학입학고사령 12조 시행 - 대학입시 성적 반영	고입내신 체력검사 · 일반체력검사 종목에서 악력 제외
1974	· 검사종류 구분 시행 - 상급학교 입시내신 체력검사	· 고입 내신체력검사와 종목 동일 · 이전 연도와 동일
1979	· '학교신체검사규칙'으로 명칭 변경 - 문교부령 446호 의거 변경 - 5개 종목 축소 시행 - 절대기준 평가 → 상대기준 평가	·5개 종목 - 100m - 제자리멀리뛰기 - 턱걸이(남)/팔굽혀매달리기(여) - 윗몸일으키기(30초) - 던지기
1983	· 검사종목 확대 조정(5종목 → 6종목)	· 이전 5개 종목에 오래달리기 추가
1991	· 대학 입시제도 개선안 발표 - 학력고사 및 대입학생체력검사 폐지	·5개 종목 - 100m - 제자리멀리뛰기 - 턱걸이(남)/팔굽혀매달리기(여)
1993	· 대입 학력고사 및 학생체력검사 폐지 - 체력장 존속	- 윗몸일으키기(30초)
1996		- 던지기 - 오래달리기
1999	· 종목 변경 시행	·5개 종목 - 50m 달리기 - 제자리멀리뛰기
2006	· '학교건강검사규칙'으로 개정 시행(체력 검사→신체능력검사)	- 앉아 윗몸 앞으로 굽히기 - 윗몸일으키기(60초) - 오래매달리기(여)/팔굽혀펴기(남) - 오래달리기걸기: 초(남·여:1,000m)/중·고(남: 1,600m, 여: 1,200m)

## 2) 학생 건강체력평가의 특징

PAPS는 학생들의 학업과 일상생활, 질병예방에 중요한 건강관련체력 종목 위주로 구성되어 체력을 평가하고 그 결과를 바탕으로 신체활동을 처방하여 학생 스스로 건강과 체력증진을 향상시킬 수 있도록 제작되었다(교육과학기술부, 2009). 또한 장소에 구애받지 않고 자신의 체력 평가 결과를 확인할 수 있고, 학부모의 관심이 집중되는 비만과 자세변형 등을 평가, 관리하며, IT 장비를 이용한 측정도 가능하여 학생의 흥미를 유발하고 측정의 신뢰성을 보장하고 있다(임정화, 2012).

## 3) 학생 건강체력평가의 구성내용 및 세부항목

PAPS는 건강관련체력 4개 요인(심폐지구력, 유연성, 근력·근지구력, 체지방)과 순발력에 대한 측정종목 12개 중에서 각 요인별로 한 종목씩 선택하여 평가한다. <그림 1>은 PAPS 필수평가 항목을 나타낸 것이다. 또한 심폐지구력 정밀평가, 비만평가, 자기신체평가, 자세평가의 4가지 평가를 별도로 제공하고 있다. 비만평가는 체성분 측정 장비를 이용하여 근육량과 지방량, 체지방율을 측정하고, 심폐정밀 평가는 심폐능력 측정장비를 이용하여 필수항목 중 심폐지구력의 종목과 동일하게 측정한다. 자기 신체평가는 20개의 설문문항을 이용하며, 자세평가는 올바르지 못한 습관으로 발생하는 자세이상, 신체뒤틀림 등을 평가하여 의사에게 진단과 치료를 유도하도록 하였다(교육과학기술부, 2009). <표 2>는 PAPS의 세부 측정내용을 요약한 내용이다.

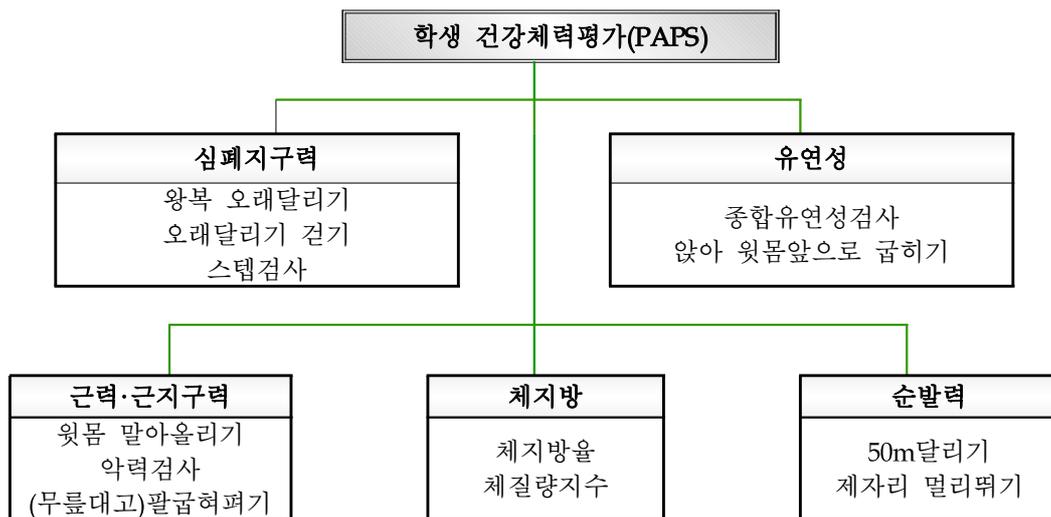


그림 1. PAPS 필수평가 항목(교육과학기술부, 2009)

표 2. 학생 건강체력평가(PAPS)의 세부 측정내용(교육과학기술부, 2009)

평가 요인 및 종목	구분	검사 대상					측정소요 예상 시간	검사방법
		초등			중	고		
		4	5	6				
심폐 지구력	왕복 오래달리기	○	○	○	○	○	약 10분	초등학생 15m 중·고등학생 20m
	오래 달리기걸기	/	○	○	○	○	약 10분	초등학생 1000m 중·고 여학생 1200m 중·고 남학생 1600m
	스텝검사	/	○	○	○	○	약 12분	초등학생 높이 20.3cm 중학생 높이 45.7cm 고등(여)높이 45.7cm 고등(남)높이 50.8cm
유연성 필수 평가	앞아랫몸 앞으로굽히기	/	○	○	○	○	약 30초	초등학생 5~6학년 중·고등학생 전체 실시
	종합 유연 성	어깨	○	○	○	○	약 2분	초, 중, 고등 전체 학생 좌/우 실시
		몸통	○	○	○	○		
		옆구리	○	○	○	○		
하체		○	○	○	○			
근력· 근지구 력	(무릎대고) 팔굽혀펴기	/	/	/	○	○	·	중·고등 남학생: 팔굽혀펴기 중·고등 여학생: 무릎대고 팔 굽혀펴기
	윗몸 말아올리기	○	○	○	○	○	약 10분	초, 중, 고등 전체 학생 실시
	악력검사	○	○	○	○	○	약 1분	초, 중, 고등 전체 학생 좌/우 실시
순발력	50m 달리기	○	○	○	○	○	약 30초	초, 중, 고등 전체 학생 실시
	제자리 멀리뛰기	○	○	○	○	○	약 30초	초, 중, 고등 전체 학생 실시
체지방	체지방률	○	○	○	○	○	약 2분	체지방자동측정장비(BIA방법)
	체질량지수	○	○	○	○	○		BMI = 체중(kg)/신장(m) <sup>2</sup>
선택 평가	왕복 오래달리기	○	○	○	○	○	약 10분	
	오래 달리기걸기	/	○	○	○	○	약 10분	심박수 측정 장비 사용
	스텝검사	/	○	○	○	○	약 12분	
	비만평가	○	○	○	○	○	약 2분	체지방 자동 측정 장비
	자기신체평가(검사지)	○	○	○	○	○	약 20분	자기신체평가 20문항, 웹에서 실시권장
	자세평가	○	○	○	○	○	약 2분	자세평가 설문지 사용

#### 4) 학생 건강체력평가의 평가기준

PAPS의 평가기준은 실측 자료와 확장된 Angoff의 방법(Hambleton & Plake, 1995)에 의한 전문가들의 심층판단을 근거로 설정하여 아주 높음, 높음, 보통, 낮음, 아주 낮음의 1등급에서 5등급으로 구분되어 있다. 전체 20점 만점 중 1등급은 16~20점, 2등급은 12~15점, 3등급은 8~11점, 4등급은 4~7점, 5등급은 3점 이하에 해당한다. 또한 8점 이상은 양호구간, 8점 미만은 우려구간으로 설정하여 우려구간의 경우 별도의 지도를 하도록 설계되었다(교육과학기술부, 2009).

### 3. 학생 건강체력평가의 측정 종목 및 방법

다음은 학생 건강체력평가(PAPS)의 각 요인별 측정 종목 및 방법(교육과학기술부, 2009)을 나타낸 것이다.

#### 1) 심폐지구력

##### (1) 왕복 오래달리기(PACER)

① 측정장비: 바닥이 평평한 공간(체육관 또는 운동장), 줄자, 앰프, 고깔

##### ② 측정방법

가. 초등학생은 15m, 중, 고등학생은 20m 떨어진 직선 거리에 고깔을 설치한다.

나. 출발 신호와 함께 달리기 시작하여 다음 신호음이 울리기 전에 맞은편 지점에 도착하고 다시 신호음에 맞추어 돌아오는 동작을 반복한다.

##### ③ 유의사항

가. 신호음 간격이 단계(level)별로 줄어든다는 것을 미리 안내해야 한다.

나. 신호음이 울리기 전에 맞은편 지점에 도착하지 못하였을 때는 다시 제자리로 돌아오며, 기록지에 “△”로 표시한다. 신호음에 맞추어 이동하지 못한 횟수가 2회이면 측정을 종료하고 기록지에 “×” 라고 표시한 후 직전 횟수를 최종기록으로 기록한다.

##### (2) 오래달리기 걷기

① 측정장비: 바닥이 평평한 공간(운동장 트랙), 초시계, 호루라기

② 측정방법: 초등학교 5, 6학년은 1000m, 중, 고등학교 여학생은 1200m, 중, 고등학교 남학생은 1600m 트랙을 달리거나 걷는다.

③ 유의사항: 결과는 분, 초 단위로 기록하며, 파울 시 1회당 5초 추가한다.

(3) 스텝 검사

① 측정장비: 바닥이 평평한 공간(체육관 또는 운동장), 스텝박스, 초시계, 앰프,

② 측정방법

가. 스텝 높이는 초등학생 5, 6학년 20.3cm, 중학생 45.7cm, 고등학생(여) 45.7cm, 고등학생(남) 50.5cm로 준비한다.

나. 초등학생 및 중학생과 여자고등학생은 분당 24회, 남자고등학생은 분당 30회의 실시간격으로 신호음에 따라 스텝박스에 올라갔다 내려오는 동작을 3분간 반복한 후 안정시 심박수를 3회 측정한다.

다. 스텝운동을 마치고 1분간 안정을 취한 후 촉진법에 의해 1회(1분~1분30초), 2회(2분~2분30초), 3회(3분~30분 30초), 총 3회에 측정된 맥박수를 <표 3>에 제시된 신체효율지수(Physical Efficiency Index: PEI) 계산 공식을 이용하여 PEI 값을 산출한다.

③ 유의사항: 스텝검사 종료 직후 3분 동안 심박수가 올라가지 않도록 최대한 편안하게 안정을 취한다.

표 3. PEI 계산 공식

대 상	공 식
초등학생(5, 6학년)	$PEI = D / (2 \times P) \times 100$ D: 스텝운동 지속시간(초) P: 3회 맥박수 합계
중학생(남, 여)	
고등학생(여)	
고등학생(남)	$PEI = D \times 100 / \{5.5 \times P\} + \{0.22 \times (300 - D)\}$

2) 유연성

(1) 앉아 윗몸앞으로 굽히기

① 측정도구: 바닥이 평평한 공간(체육관, 운동장), 앉아 윗몸앞으로 굽히기 측정기

② 측정방법

가. 앉은 자세에서 측정기에 피험자의 발바닥이 닿도록 하고, 양팔을 앞으로 쭉 뻗으면서 윗몸을 앞으로 굽힌 자세에서 손가락 끝이 모이도록 자세를 교정한다.

나. 윗몸이 굽혀진 상태에서 2초 이상 자세를 유지하여 수치를 기록한다.

③ 유의사항

가. 양팔을 동시에 뻗은 자세가 유지되어야 하며, 허리반동을 이용하거나, 무릎을

굽히지 않도록 한다.

나. 2회 측정 중 좋은 기록을 최종기록으로 한다.

## (2) 종합유연성

① 측정도구: 매트, 숫자카드

② 측정방법

가. 어깨, 옆구리, 몸통, 하체 총 4부위의 유연성을 측정한다.

나. 어깨의 경우, 양손을 X자 자세로 한손은 머리 뒤쪽으로, 다른 한손은 옆구리 옆에서 등으로 뺀어 양손을 잡는다.

다. 옆구리의 경우, 발을 어깨넓이로 벌리고 한손은 머리에, 다른 손은 무릎 뒤 오금부위에 닿게 한다.

라. 몸통의 경우, 양발 뒤꿈치에 숫자카드를 놓고 허리를 돌려 뒤꿈치 숫자를 읽는다.

마. 하체의 경우, 왼발은 곧게 뻗은 자세에서 오른발은 굽혀 세워 양손이 왼발 끝에 닿게 해야 하며, 반대로 오른발이 곧게 뻗은 자세에서 왼발은 굽혀 세워 양손이 오른발 끝에 닿게 한다.

자. 양쪽이 모두 성공 할 경우 2점, 한쪽만 성공 할 경우 1점, 양쪽 모두 실패했을 경우 0점으로 기록 한다.

③ 유의사항: 모든 자세는 3초간 정지해야 하며, 몸의 반동을 이용하지 않는다.

## 3) 근력 근지구력

(1) 팔굽혀펴기(남), 무릎대고팔굽혀펴기(여)

① 측정도구: 팔굽혀펴기대, 계수기, 매트

② 측정방법: 남자와 여자의 경우 동일하게 양발을 모으고 양팔은 어깨 넓이만큼 벌려서 팔굽혀펴기대의 봉을 잡는다. 팔을 굽혀서 내려오는 동작 중, 팔굽혀펴기대 봉과 가슴사이가 10cm이하가 되어야하고, 팔의 각도는 90°를 유지해야 한다.

③ 유의사항: 상체만 위 아래로 움직이지 않도록 주의한다.

(2) 윗몸 말아올리기

① 측정도구: 앰프, 매트

② 측정방법

가. 매트위에 양발이 완전히 닿아, 무릎을 굽히고 상체를 완전히 눕혀 양손은 허벅지 위에 올려, 3초에 1번씩 신호음에 맞추어 손이 허벅지 위를 타고 올라가 손바

닥으로 무릎을 감싸면서 상체를 맡아 올린다.

③ 유의사항: 1회 3초 간격을 지키지 못할 경우, 처음에는 허용하고 두 번째 지키지 못할 경우 측정을 종료한다.

### (3) 악력

① 측정도구: 악력계

② 측정방법: 양발을 어깨너비로 벌리고 악력계를 중지의 두 번째 마디가 걸림쇠와 직각이 되도록 잡는다. 측정은 오른쪽과 왼쪽 각각 2회 실시한다.

③ 유의사항: 팔을 굽히거나, 상체를 위아래로 움직이면서 측정하지 않도록 한다.

## 4) 순발력

(1) 50m 달리기

① 측정도구: 바닥이 평평한 공간(체육관 또는 운동장), 호루라기, 초시계, 깃발

② 측정방법: 출발지점에서 호루라기 또는 깃발 신호에 맞춰서 50m 전방 도착점까지 가장 빠른 속도로 달려야 한다.

③ 유의사항: 반드시 직선주로에서 측정하고, 단위는 0.01까지 기록한다.

(2) 제자리 멀리뛰기

① 측정도구: 구름판이 있는 모래 터, 삼, 줄자

② 측정방법: 구름판에 그려진 흰색 출발선을 밟지 않고 제자리에서 몸의 반동을 이용하여 최대한 멀리 뛴다. 총 2회를 실시하여 좋은 기록을 최종기록으로 한다.

③ 유의사항: 뛰는 순간 발이 구름판 흰색 선에 닿으면 무효로 처리한다.

## 5) 체지방

(1) 신체질량지수(BMI)

신체질량지수는 신장과 체중치를 이용하여 아래 공식을 이용하여 산출한다.

$$BMI = \text{kg/m}^2$$

(2) 체지방률(%body fat)

① 측정도구: 생체전기저항분석기

② 측정방법: 개인의 필수입력 항목을 입력하고, 양발은 발바닥의 전기분석기 위

치에 맞게 자세를 취한다. 피험자는 전기분석기의 전극을 양손으로 잡으며, 2분간 직립자세를 유지한다.

#### 4. 동등화의 이해

##### 1) 동등화의 개념

동등화(equating)는 측정 또는 검사를 시행하여 얻은 점수들을 상호 교환하고 보정하기 위한 장치로 연계화(linking)의 한 유형이라고 할 수 있다. 연계화는 동등화를 포함하여 조정(calibration), 예측(prediction), 통계적 조절(statistical moderation), 사회적 조절(social moderation) 방법이 있다. <표 4>는 Linn(1993)이 제시한 연계 방법을 정리한 것이다.

표 4. 연계화 방법의 종류와 활용(성태제, 2000)

유형	개념	활용
1. 동등화 (equation)	가장 강력한 통계적 연계방법. 동등화된 검사는 상호교환 가능	동일한 특성을 측정하는 여러 동형검사 비교하는 경우
2. 조정 (calibration)	개인 혹은 집단 수준에서 검사 결과를 비교할 때 사용	동일한 특성을 측정하지만 신뢰도와 난이도가 다른 검사 간 점수들을 비교하는 경우
3. 통계적 조절 (statistical moderation)	다른 특성을 재는 검사 간 공통된 검사를 실시하여 평가 결과를 비교	단일 요인을 측정하는 여러 검사가 존재할 때, 공통검사를 이용해 다른 검사 결과와 비교
4. 예측 (prediction)	가장 약한 통계적 연계 방법. 상관을 이용하여 한 검사 점수로 다른 검사의 점수 예측. 집단, 상황, 시간에 민감	1차 검사 후, 2차 검사 결과를 예측하고자 하는 경우
5. 사회적 조절 (social moderation)	통계적인 방법을 적용하는 것이 아닌, 평가자들의 합의를 통해 공통된 평가기준을 마련한 후 여러 검사 간 비교	검사에 관련된 평가기준의 활용도에 대한 의사결정을 하는 경우

## 2) 동등화 방법

동등화 방법은 전통적인 방법과 문항반응이론(IRT: item response theory)에 의한 방법으로 구분된다. 전통적 방법으로는 선형동등화(linear equating)와 동백분위 동등화(equi-percentile equating), 평균동등화(mean equating) 방법이 있다(Kolen & Brennan, 1995, 2004). 문항반응이론에 의한 방법은 능력점수를 동일척도로 만드는 방법과 진점수동등화(true score equating)방법, 관찰점수동등화(observed estimation) 방법으로 구분된다(Stocking & Lord, 1983). <표 5>는 동등화 방법의 종류와 그 개념을 설명한 것이다.

표 5. 동등화 방법의 종류

구분	종류	개념
전통적 방법	평균동등화	서로 다른 검사의 평균 차를 구해서 평균이 낮은 검사를 치른 피험자에게 그 차이점수를 더하는 방법. 예를 들어 A검사를 치른 피험자의 평균이 80이고, B검사를 치른 피험자의 평균이 83점이라면, A검사를 치른 피험자에게 3점을 더해주는 방법
	선형동등화	두 검사를 치른 피험자의 결과 분포가 동일한 평균과 표준편차를 갖도록 한 검사형의 측정치를 변환하는 방법. 두 검사의 표준점수가 같을 경우 동등점수로 간주
	동백분위동등화	두 검사자료의 누적빈도가 같다면, 즉 같은 백분위점수는 동등점수로 간주. 예를 들어 A검사에서 점수 X를 받은 피험자와 B검사를 치른 피험자 중 동일한 백분위에 해당하는 피험자에게 같은 X점수를 부여하는 방법
IRT 방법	능력점수의 동일척도화 <ul style="list-style-type: none"> <li>• 진점수 동등화</li> <li>• 관찰점수 동등화</li> </ul>	<p>두 검사의 문항모수추정치(item parameter estimates)를 동일 척도 위에 올려놓는 방법. 동형 또는 평형인 검사의 모수 추정치 값이 동일 척도 위에 놓이게 되면 개인의 능력(체력) 모수치는 검사유형과 관계없이 자동적으로 같은 척도로 간주하며, 추가적인 작업 없이 두 검사 형으로부터 능력(체력) 점수 상호 비교 가능</p> <p>오래달리기 오래달리기 걸기</p>

### (1) 선형동등화

선형동등화 방법은 각 검사의 표준점수가 같을 경우 동등점수로 간주한다. 검사 결과의 분포가 유사할 경우 연구 대상자의 수가 적더라도 비교적 정확한 동등화 결과를 나타낸다. 또한 평균 점수 근처에서는 선형방법이 비교적 정확한 동등 점수를 산출할 수 있다. 검사 X의 선형동등 전환 점수는 검사 X의 표준점수에서 검사 Y의 표준편차를 곱하고 평균을 더한 값과 같다(공식 2.4.1)(남현우, 2001).

$$l(x) = \sigma(y) \{ [x - \mu(x)] / \sigma(x) \} + \mu(y) \quad \text{----- (2.4.1)}$$

### (2) 동백분위 동등화

동백분위 동등화 방법은 평균을 이용한 것으로 선형변환이 갖는 특성을 모두 가지고 있고, 다양한 검사 점수의 분포를 동일하게 만들어 준다는 장점이 있다. 즉 상대적으로 동일한 위치에 있을 경우 서로 같은 점수를 받을 수 있도록 조정하는 방법이다. 예를 들어 왕복 오래달리기 검사에서 상위 15%에 해당하고 오래달리기 걷기 검사에서 상위 15%에 해당할 때 동일한 표준점수를 받을 수 있도록 조정하는 것이다. 동백분위 동등화는 두 단계에 걸쳐 이루어지는데 먼저 각 검사결과의 백분위 함수(percentile rank function)를 구한 후, 같은 백분위에 해당하는 대응 점수를 찾는다. 동등화된 X검사 점수 x의 분포 함수를 F라 하며, 검사 Y의 점수 y의 분포 함수를 G라 하고, 검사 X를 검사 Y에 동등화 하는 함수를 e(x)라 할 때, 점수 x와 점수 y는 동등점수가 된다(공식 2.4.2)(남현우, 2001).

$$F[e(x)] = G(y) \quad \text{----- (2.4.2)}$$

동백분위 동등화법의 동등화 함수는 표집의 오차로 인해 불규칙적(irregular)으로 나타나므로 이를 교정하는 완곡화(smoothing)가 필요하다. 완곡화는 사전 완곡화(pre-smoothing) 또는 사후 완곡화(postsmoothing)로 구분할 수 있다. 사전 완곡화 방법으로 자주 활용되는 것은 평균이동방법(rolling average methods)(Kolen, 1991)과 로그선형방법(log loinear method)(Holland & Thayer, 1987) 또는 강진점수방법(strong true score method)(Kolen & Brennan, 1995)이다.

### 3) 동등화의 조건

동등화를 하기 위해서는 다음과 같은 4가지 조건이 충족되어야 한다(Angoff,

1971; Lord, 1980). 첫째, 같은 능력을 조건으로 두 검사는 모두 동일한 잠재적 특성 혹은 능력을 측정해야한다. 특히 문항 반응이론 방법으로 동등화를 할 경우에는 두 검사들이 동일한 특성을 측정하는 것 뿐만 아니라, 그 특성은 일차원적일 때 동등화가 가능하다. 둘째, 공정성의 조건으로 각 피험자가 X와 Y의 검사 중 어느 검사를 치르든 관계없이 공정한 검사가 되어야한다. 즉, 두 검사의 결과 값이 신뢰도가 높을 때 동등화 조건이 성립된다(Lord, 1980). 셋째, 모집단 불변성의 조건으로 검사 X와 Y의 동등화를 위한 변환은 피험자들의 모집단 특성에 의해 영향을 받지 않아야한다. 즉, 검사 X와 Y가 어느 피험자 집단에 실시되든 상관없이 동등화 변환은 변하지 않아야 한다. 넷째, 대칭성의 조건으로 두 검사의 점수가 동등하려 면 두 검사의 결과 점수는 일대일로 대응해야 한다. 또한 동등화 변환은 역으로도 가능하다. Petersen, Kolen과 Hoover(1989)의 연구에 의하면 동일능력, 공정성, 모집단 불변성 그리고 대칭성의 4가지 조건 중 최소한 모집단이 불변성과 대칭성의 조건은 반드시 충족해야만 동등화가 가능하다(남현우, 2001).

#### 4) 동등화 오차 산출 및 타당성 검증

동등화 오차는 무선오차(random error)와 체계적 오차(systematic error)로 구분할 수 있다. 무선오차는 표본자료를 이용하기 때문에 나타나는 오차로 피험자 수를 늘이면 오차를 줄일 수 있다. 무선오차는 일반적으로 동등화의 표준오차로 확인 가능하다. 동백분위 동등화 방법의 표준오차는 Lord(1981), Jarjoura & Kolen(1985), Holland, King, & Thayer(1989), 선형동등화 방법의 표준오차는 Kolen(1984), Zeng(1991), IRT 방법의 표준오차는 Lord(1981), WIngersky & Lord(1984)가 제시한 방법에 의해 확인할 수 있다. 다양한 연구에서 Bootstrap 방법을 이용하여 동등화 오차를 산출하는데, 이 방법은 각 점수별로 동등화 오차가 얼마나 되는지 자세히 알 수 있다. Bootstrap의 절차는 다음과 같다(김아영, 이규민, 반재천, 차정은, 2004).

- (1) 표본 집단으로부터 측정된 체력검사 결과 자료와 사례수가 동일한 Bootstrap 표본을 추출한다. 특히 이 표본 추출 방식은 복원(replacement) 방식을 이용한다.
- (2) Bootstrap 표본을 이용해 평균과 표준편차 등 통계치를 산출한다.
- (3) 위 단계를 500회 반복한다.
- (4) 위 500개의 Bootstrap 방법에 의해 산출된 표준편차가 Bootstrap표준오차(estimated Bootstrap standard error)가 된다. 이 표준오차를 비교한다.

## 5. 선행연구

### 1) 동등화 관련 연구

동등화에 대한 논의는 전통적인 동등화 방법을 소개한 Angoff(1971; 1984)에 의해 시작되었다. 이후 Lord(1980), Hambleton(1983), Hambleton & Swaminathan(1985)이 문항반응 이론을 적용한 동등화 방법을 소개하였고, Petersen 등(1989)은 고전검사 이론에 의한 검사 동등화 방법과 문항반응 이론에 의한 동등화 방법을 비교 분석하여 체계적으로 동등화를 설명하였다. Holland & Rubin(1982)은 동등화 검사에 관한 이론과 분석방법에 대하여 구체적인 실례를 들어 설명하였다. 또한 Jarjoura와 Kolen(1985, 1987), Harris와 Kolen(1990)등도 동등화에 대한 실증적인 연구를 수행하였다.

체육학에서는 Zhu(1998)에 의해 동등화의 개념 및 필요성이 소개되어, 문항반응 이론에 입각한 운동기능검사의 동등화 연구(Zhu, 2001), 신체활동 분야에 사용되고 있는 다양한 설문지를 문항반응이론 적용하여 단일척도로 변환하는 동등화 연구(Zhu, 2002)가 수행되었다.

국내에는 최초 교육 분야를 시작으로 이종성(1985)에 의해 검사동등화 개념이 소개되었으며, 이후 문항반응이론을 적용한 동등화 검사가 소개되었다(이종성, 1990). 김현철(1999)은 동일한 방법으로 선택 교과 점수를 연계화한 다음 표준점수를 산출하는 선형변환 표준점수체제와 비선형변환 표준점수체제의 점수분포와 특징을 비교 분석하였다. 분석 결과, 선형변환 표준점수와 비선형 변화 표준점수는 총점의 분포 면에서 거의 비슷하다고 제시하였고, 원점수와 점수 차이나 또는 순위변동이 선형변환척도 점수에서 더 적게 일어나는 것을 규명하였다. 부재율(1999)은 검사 동등화의 고려 사항에 대한 지침을 제시하였다. 성태제(2000)는 Linn(1993)과 Mislevy(1992)가 제안한 5가지 점수 연계화 방법으로 동등화, 조정, 예측, 사회적 조절, 통계적 조절을 소개하고 여러 통계적 조절 방법을 비교하여 각 방법의 타당성을 분석하고 장단점을 밝혔다. 남현우(2003)는 2005년 수능의 탐구영역 내 과목 간 점수조정에 다양한 동등화 방법을 비교하고 합집단 방법을 이용하여 연계화 방법의 적용 가능성을 탐색하였다. 김아영, 이규민, 반재천, 차정은(2004)은 개정된 성인용 직업적성검사와 기존의 성인용 직업적성검사를 연계화하고, 개정된 성인용 직업적성검사 종목의 타당성을 검증하였다. 이규민(2005)은 초등학교 3학년 국가수준 기초학력 진단평가의 다년간 결과를 동등화하여 진단평가의 적절성을 비교하고 현

장에서 사용 가능한 구체적인 방안을 제시하였다.

체육분야의 동등화 연구를 살펴보면 고기환, 강상조(1997)가 배구의 다양한 서브 유형에 대한 기능검사를 실시하여 문항반응이론을 이용한 동등화 및 각 시행별 차이를 비교 분석하였다. 박문식(1998)은 오래달리기 거리 간 동등화를 실시하기 위해 중학교 남녀학생을 대상으로 동백분위 동등화 방법을 이용하여 오래달리기 기록 간 비교가 가능한 동등화 표를 개발하였다. 정재운(1999)는 초, 중학교 재학 중인 남학생과 여학생 총 1597명을 대상으로 6가지 종류의 순발력 검사로부터 얻은 점수를 RAGE 컴퓨터 프로그램을 통해 동등화를 실시하여 종목 간 변환점수표를 개발하였다. 또한 남덕현, 윤상화, 김도준, 박재우(1999)는 다양한 순발력 검사종목 간 동등화를 실시하여 체육현장에서 보다 간편하고 편리한 순발력 검사가 가능하도록 하였다. 남덕현(2002)은 고등학생을 대상으로 심폐지구력 검사점수를 비완곡화, 동백분위, 완곡화 동백분위, 선형동등화 방법을 거쳐 표준오차를 검증하였다. 한태호는(2011)는 척수장애인 149명을 대상으로 유연성 검사를 실시하여 일반인과 척수장애인의 등급별 체력기준을 비교 및 평가가 가능하도록 동등화를 실시하였다. 또한 권오형(2013)은 시각장애 학생과 비장애 학생의 PAPS 점수를 분석하여, 시각장애 학생들이 시각정보 습득의 불리함에 따른 체력의 차이를 규명하였다.

## 2) 학생 건강체력평가 관련 연구

이해희(2014)는 왕복 오래달리기(PACER)를 활용하여 중학교 축구선수, 주 3회 이상 운동하는 학생, PAPS 평가등급이 2등급 이상인 학생과 4~5등급인 학생을 집단별로 심폐지구력 수준을 비교 분석하여, 축구선수와 주 3회 이상 운동한 집단이 다른 집단보다 심폐지구력 수준이 높은 것을 규명하였다. 김경훈(2012)은 고등학교 남학생의 PAPS 종목별, 학년별 평가등급을 분석하여 각 등급에 알맞은 신체활동과 청소년의 전반적인 체력수준을 파악하여 교육정책의 기초자료를 제공하였다. 기선경(2013)은 초등학생 5학년과 6학년 총 358명을 대상으로 PAPS 13개 종목을 측정 후 신체적 특성과 체력특성에 따른 클럽활동의 차이를 검증하였고, 체격요인과 스포츠 잠재력 관계를 분석을 실시하여 스포츠 영재발굴을 위한 기초자료를 제공하였다. 배성웅(2014)은 3년간 PAPS 종합체력점수와 학업 성적 간 연관성을 분석하기 위해 고등학생 220명을 대상으로 3년간 종단연구를 실시하여, PAPS의 종합체력 점수와 학업성적의 상관성은 약한 정적상관관계가 있음을 밝혔다.

최인엽(2011)은 초, 중, 고 학생 총 191명을 대상으로 PAPS 건강체력과 학업성취도와의 관계를 규명하였다. 연구결과에 의하면 PAPS의 종합 체력점수가 높거나,

체력 수준이 높은 학생일수록 학업성적이 우수한 것으로 나타났다. 배소라(2013)는 PAPS의 심폐지구력, 순발력, 근력 근지구력, 유연성, 체지방의 항목에서 낮은 등급을 받을수록 신체적 자기개념이 낮아진다고 보고하였다. PAPS를 활용하여 비만 아동을 과학적이고 체계적으로 지도하기 위해 양인철(2010)은 초등학교 5, 6학년 과체중, 정상체중, 비만아동의 종목 별 차이를 분석한 결과, 체중이 높을수록 체력이 저하된다고 보고하였다. 이경수(2012)는 초등학생의 신체질량지수(BMI), 즉 비만도가 높을수록 심폐지구력, 유연성 능력이 낮아지는 것을 규명하였다.

엄용현(2014)은 PAPS를 활용하여 건강체력 상태와 비만결정요인의 차이를 분석하여 근지구력과 근력, 심폐지구력이 비만결정요인에 영향을 미친다고 보고하였다. 나법경(2014)은 PAPS의 체력요인 5가지 항목 간 성별, 학년별 차이를 분석하였고, 방과 후 스포츠클럽 활동 참여집단과 비참여집단을 비교하여 체력발달에 미치는 요인을 규명하였다. 연구결과 남학생, 여학생 모두 스포츠클럽 참여집단이 비참여집단에 비해 건강체력 요소 측정값이 우수한 것으로 나타났다. 박태섭(2001)은 학생들의 신체발육정도 상, 중, 하에 따른 평가 기준표를 개발하였다. 초등학생 400명을 대상으로 5가지 종목을 측정한 결과, 체력검사 제도의 등급분포와 체격발육을 고려한 체력검사 등급분포 간에는 차이가 있었고, 학생 체력검사제도의 등급 분포는 1등급에 편포되었으나, 체격을 고려한 등급분포는 3등급에 더 많은 분포가 나타나는 것을 확인하였다.

### Ⅲ. 연구방법

#### 1. 연구대상

이 연구의 대상은 제주도 제주시 소재 2개 초등학교, 서귀포 소재 4개 초등학교, 총 6개 초등학교의 5학년 남학생 131명, 여학생 112명, 6학년 남학생 121명, 여학생 104명으로 모두 468명이다. <표 6>은 연구대상자의 신체적 특성을 나타낸 것이다.

표 6. 연구대상의 신체적 특성

학년	성별	n	신장(cm)	체중(kg)
5학년	남자	131	144.3±5.5	40.9±8.4
	여자	112	145.8±5.8	45.5±13.3
6학년	남자	121	149.6±6.8	44.8±9.2
	여자	104	151.1±7.4	47.8±11.2

#### 2. 측정방법

PAPS 종목 중 심폐지구력 검사에 해당하는 왕복 오래달리기(PACER), 오래달리기 걷기, 스텝검사를 실시하였다. 이 연구는 단일집단설계로서 모든 피험자가 심폐지구력 검사인 오래달리기 걷기, 왕복 오래달리기, 스텝검사에 모두 참여하였다. 검사는 피로가 회복되는 3~7일 이상의 간격을 두고 하루에 한 종목씩 실시하였고, 모든 실험은 1개월 이내에 측정하여 성숙에 의한 오차가 발생하지 않도록 통제하였다. 각 종목별 측정은 2009년도 교육과학기술부에서 제시한 PAPS 검사 지침서(교육과학기술부, 2009)에 따라 실시하였다.

##### 1) 왕복 오래달리기

왕복 오래달리기(PACER)는 체육관에서 정해진 신호음에 맞추어 15m 거리를 왕복으로 달리는 동작을 반복하였다. 신호음과 함께 15m 반대편으로 신속히 이동하고, 다음 신호음을 기다렸다가 다시 돌아오도록 하였다. 신호음이 울리기 전에 맞은편 지점에 도착하지 못하였을 때는 경고를 주고 신속히 제자리로 돌아와 다음

신호를 기다렸다가 다시 출발하도록 하였다. 1회 미완수시 기록지에 “△” 로 표시하고 2회 미완수하면 측정을 중단하고 기록지에 “×” 로 표시한 후 직전횟수를 최종기록으로 기입하였다.



그림 2. 왕복 오래달리기

## 2) 오래달리기 걷기

중학생 남녀 모두 운동장 트랙 1000m 거리를 본인이 이동할 수 있는 범위내에서 최선을 다해 달리거나 걷도록 하였다. 측정치는 0.1초 단위를 버리고 분, 초 단위로 기록하였다.



그림 3. 오래달리기 걷기

## 3) 스텝검사

피험자들은 신호음에 맞춰 20.3cm의 스텝박스에 올라갔다, 내려오는 동작을 반복하여 분당 24회 속도로 3분간 측정하였다. 3분간 스텝운동을 실시한 후 휴식하는 동안 촉진법에 의해 1회(1분~1분30초), 2회(2분~2분30초), 3회(3분~30분30초)에 측정된 맥박수를 기록하였다. 본인의 맥박수가 최초 안정시 심박수와 비교하여 더 낮을 경우에는 측정값의 신뢰도를 위해 다음날 재검사를 실시하였다. 측정된 맥박수는 아래 신체효율지수(Physical Efficiency Index: PEI) 계산 공식에 의해

PEI 값을 산출하였다.

$$PEI = D / (2 \times P) \times 100$$

D : 스텝운동 지속시간(초)

P : 3회 맥박수 합계



그림 4. 스텝검사

표 7. 심폐지구력 검사종목 평가기준표(교육과학기술부, 2009)

검사종목	학년	성별	아주 낮음 (5등급)	낮음 (4등급)	보통 (3등급)	높음 (2등급)	아주 높음 (1등급)
왕복 오래달리기 (lap)	5	남	22~28	29~49	50~72	73~99	100~107
		여	18~22	23~44	45~62	63~84	85~104
	6	남	22~31	32~53	54~77	78~103	104~112
		여	20~24	25~49	50~68	69~92	93~112
오래 달리기걸기 (초)	5	남	480~640	410~479	325~409	282~324	268~281
		여	502~640	442~501	360~441	300~359	268~299
	6	남	450~587	380~449	315~379	251~314	243~250
		여	480~587	430~479	354~429	300~353	243~299
스텝검사 (PEI)	전 학년		44.0~46.9	47.0~51.9	52.0~61.9	62.0~75.9	76.0~95.0

### 3. 동등화 절차

이 연구에서는 심폐지구력의 세 가지 검사 중 왕복 오래달리기(PACER) 검사를 기준으로 나머지 두 검사를 동등화 하였다. 먼저 3가지 검사 간 유사성을 검증하고 평가등급에 차이가 있는지 분석하였다. 동등화는 선형 동등화와 동백분위 동등화에 의해 오래달리기 걸기와 스텝검사의 변환등급을 산출하였다.

### 1) 검사의 유사성 검증

심폐지구력을 측정하는 왕복 오래달리기(PACER), 오래달리기 걷기, 스텝검사가 서로 유사한 구인을 측정하는지 확인하기 위해 상관계수( $r$ )를 산출하였다.

### 2) 변환등급 산출

왕복 오래달리기 등급을 기준으로 오래달리기 걷기와 스텝검사의 변환등급을 산출하기 위해 선형 동등화와 동백분위 동등화 방법을 적용하였다(남현우, 2001).

#### (1) 선형 동등화에 의한 변환점수 산출

선형 동등화는 두 검사의 표준점수가 같으면 동등 점수로 간주한다. 이 가정을 수식으로 표현하면 아래 식(3.3.1)과 같다(남현우, 2001)

$$(Y - m_y)/s_y = (X - M_x)/s_x \quad (3.3.1)$$

위 식을 일차 함수( $Y = aX + b$ )로 정리하면 다음 식 (3.3.2)와 같다.

$$Y = (s_y/s_x)X + [M_y - (s_y/s_x)M_x] \quad (3.3.2)$$

X와 Y에 관한 함수식을 결정하는 기울기(a)는 두 검사 표준편차의 비에 해당하는  $(s_y/s_x)$  로 계산되며, 절편(b)는  $[M_y - (s_y/s_x)M_x]$ 로 계산된다(남현우, 2001).

#### (2) 동백분위 동등화에 의한 변환점수 산출

동백분위 동등화는 서로 다른 종목의 측정결과와 누가빈도를 작성하여 상호 대응되는 점을 찾는다. 상대적 누가빈도 분포도를 통해 검사점수를 상대적 누가빈도에서 상호 대응 가능한 위치에 있는 점수를 말한다. 즉, 동일한 백분위에 위치한 점수의 짝을 의미한다. P(x)와 Q(y)는 백분위로서 아래의 식(3.3.3)에 의해 계산된다(남현우, 2001).

$$P(x) = 100[F(x-1) + f(x)/2] \quad (3.3.3)$$

백분위 점수를 구했다면, 다음 식(3.3.4)에 의해 동등 백분위에 해당하는 점수를 구한다(남현우, 2001).

$$e_Y(x) = Q^{-1}[P(x)] \quad (3.3.4)$$

$$= \frac{P(x)/100 - G(y_U^* - 1)}{G(y_U^*) - G(y_U^* - 1)} + (y_U^* - .5)$$

$$= K_Y + .5, \quad P(x) = 100 \text{ 일 때}$$

### 3) 동등화 오차산출

이 연구에서는 Bootstrap 방법을 적용하여 동등화 오차를 산출하였는데, 동등집단의 표본자료를 이용하여 500개의 Bootstrap 표본을 각 검사별로 추출한 후 표준 오차를 구하였다.

## 4. 자료분석

SAS version 9.3을 이용하여 왕복 오래달리기(PACER), 오래달리기 걷기, 스텝검사 측정 자료의 기술통계를 산출하였고, 각 검사의 평가등급에 차이가 있는지 확인하기 위하여 빈도분석과 카이제곱( $\chi^2$ ) 검증을 실시하였다. 또한 종목별 평가등급 결과가 얼마나 일치하는지 확인하기 위하여 교차분석을 통해 일치도 통계(%)를 산출하였다. 심폐지구력 3가지 측정 종목 간 검사의 유사성은 상관계수( $r$ )를 통해 확인하였다. 동등화 분석프로그램인 RAGE(Zeng, Kolen, & Hanson, 2005)를 사용하여 선형 동등화와 동백분위 동등화를 실시하였고 큐빅 스플라인 방법으로 완곡화를 실시하였다.

## IV. 연구결과

### 1. 기술통계

<표 8>은 초등학교 5, 6학년 남학생과 여학생을 대상으로 학생 건강체력평가(PAPS) 심폐지구력 종목인 왕복 오래달리기(PACER), 오래달리기 걷기, 스텝검사를 실시한 결과를 나타낸 것이다.

표 8. 심폐지구력 검사종목별 기술통계

검사종목	학년	성별	n	평균	표준 편차	최대값	최소값	왜도	첨도
왕복 오래달리기 (lap)	5	남	131	70.50	25.71	118.00	19.00	0.16	-0.74
		여	112	53.05	21.36	118.00	12.00	0.61	0.15
	6	남	121	79.83	24.18	118.00	24.00	-0.06	-0.85
		여	104	59.36	25.29	118.00	14.00	0.60	-0.34
오래 달리기걷기 (초)	5	남	127	369.72	89.30	681.00	230.00	0.60	0.55
		여	112	408.44	75.38	633.00	251.00	0.63	0.90
	6	남	121	359.12	90.54	720.00	229.00	1.41	3.00
		여	104	424.68	96.02	686.00	249.00	0.50	0.25
스텝검사 (PEI)	5	남	128	76.47	15.92	126.80	46.20	0.51	0.22
		여	111	74.77	13.60	112.50	53.60	0.63	-0.20
	6	남	121	75.63	15.02	126.80	51.30	0.93	0.89
		여	104	75.22	15.45	123.30	43.90	0.81	0.77

<표 8>을 살펴보면 전반적으로 학년이 높을수록 기록이 더 좋아지는 것을 확인할 수 있다. 그러나 여학생의 경우 오래달리기 걷기의 평균은 5학년(408.44)이 6학년(424.68)보다 더 우수한 것으로 나타났다. 또한 스텝검사도 남학생의 경우 5학년(76.47)이 6학년(75.63)보다 더 높은 것으로 나타났다. <표 8>에서 왕복 오래달리기의 최대값이 남녀 모두 118로 모두 동일한데 이는 PAPS에서 118이 최대값으로 지정되어 있기 때문이다.

## 2. 검사종목 간 평가등급 차이

### 1) 빈도분석

<표 9>는 초등학교 5, 6학년 남학생과 여학생을 대상으로 왕복 오래달리기, 오래달리기 걷기, 스텝검사를 실시한 후 그 결과의 평가등급 빈도와 종목별 등급 차이를 검정한 것이다.

표 9. 심폐지구력 검사종목 간 평가등급의 빈도분석

학년	성별	종목*	평가등급(빈도, %)					전체	$\chi^2$	P
			1	2	3	4	5			
5	남	A	39 (29.7%)	34 (26.0)	43 (32.8)	12 (9.2)	3 (2.3)	131 (100)	34.80	<.001
		B	33 (26.0)	29 (22.8)	39 (30.7)	19 (15.0)	7 (5.5)	127 (100)		
		C	60 (46.5)	43 (33.3)	21 (16.3)	4 (3.1)	1 (0.8)	129 (100)		
	여	A	10 (9.0)	27 (24.3)	32 (28.8)	39 (35.1)	3 (2.7)	111 (100)		
		B	6 (5.4)	26 (23.2)	45 (40.2)	25 (22.3)	10 (8.9)	112 (100)		
		C	45 (40.2)	43 (38.4)	24 (21.4)	0 (0.0)	0 (0.0)	123 (100)		
6	남	A	52 (43.0)	38 (31.4)	23 (19.0)	8 (6.6)	0 (0.0)	121 (100)	35.20	<.001
		B	30 (24.8)	39 (32.2)	33 (27.3)	12 (9.9)	7 (5.8)	121 (100)		
		C	53 (43.8)	47 (38.8)	20 (16.5)	1 (0.8)	0 (0.0)	121 (100)		
	여	A	16 (15.4)	25 (24.0)	29 (27.9)	33 (31.7)	1 (1.0)	104 (100)		
		B	11 (10.6)	15 (14.4)	35 (33.7)	26 (26.0)	17 (16.4)	104 (100)		
		C	42 (40.4)	43 (41.4)	17 (16.4)	1 (1.0)	1 (1.0)	104 (100)		
전체	A	117 (25.1)	124 (26.6)	127 (27.2)	92 (19.7)	7 (2.0)	467 (100)	223.20	<.001	
	B	80 (17.2)	109 (23.5)	152 (32.8)	82 (17.7)	41 (8.9)	464 (100)			
	C	199 (42.8)	176 (37.9)	82 (17.7)	6 (1.3)	2 (0.4)	465 (100)			

\*A: 왕복 오래달리기(lap), B: 오래달리기 걷기(초), C: 스텝검사(PEI)

<표 9>에서 음영한 부분은 각 검사에서 가장 높은 빈도가 나타난 등급을 표시한 것이다. <표 9>를 살펴보면 스텝검사, 왕복 오래달리기, 오래달리기 걷기 순으로 평가등급이 높게 나타난 것을 알 수 있다. 즉 스텝검사 결과가 다른 검사 결과보다 등급이 더 우수하게 나타난 것이다. 예를 들어 5학년 남자의 경우 1등급에 해당하는 비율이 오래달리기 걷기는 26%, 왕복 오래달리기는 29.7%이지만 스텝검사는 46.5%이다. 즉, 동일한 피험자가 세 가지 검사를 실시한 평가결과가 모두 다르고 이 차이는 통계적으로도 유의하게 나타났다( $\chi^2=34.80, p<.001$ ). 여학생의 경우에도 5학년 오래달리기 걷기의 1등급 비율은 5.4%, 왕복 오래달리기는 9.0%인 반면, 스텝검사는 40.2%로 등급별 비율의 차이가 매우 크게 나타났고, 이 차이는 통계적으로도 유의하였다( $\chi^2=106.30, p<.001$ ). 이와 같은 종목별 등급 차이는 모든 학년과 성별에서 동일하게 나타났고 전체 자료 분석에서도 차이가 나타났다( $\chi^2=223.20, p<.001$ ).

## 2) 일치도 검증

<표 10>은 심폐지구력 세 가지 검사결과의 평가등급의 일치도를 나타낸 것이다.

표 10. 심폐지구력 검사결과의 평가등급의 일치도

학년	성별	n	종목	왕복 오래달리기	오래달리기 걷기
5	남	131	오래달리기 걷기	.44	.
			스텝검사	.35	.27
	여	112	오래달리기 걷기	.34	.
			스텝검사	.22	.28
6	남	121	오래달리기 걷기	.33	.
			스텝검사	.40	.36
	여	104	오래달리기 걷기	.36	.
			스텝검사	.27	.16
전체	468	오래달리기 걷기	.37	.	
		스텝검사	.31	.27	

일반적으로 일치도는 .85 이상일 때 높다고 판단하는데(성태제, 2002), <표 10>을 살펴보면 일치도 값이 .16~.44로 나타나서 검사종목 간 평가등급 결과의 일치성이 매우 낮은 것을 알 수 있다. 특히 6학년 여학생의 오래달리기 걷기와 스텝검사 평가등급의 일치도가 .16, 즉 16%로 가장 낮게 나타났다.

### 3. 동등화

#### 1) 검사의 유사성 검증

검사가 서로 같은 구인을 측정하지 않는다면 동등화는 불가능하다. 심폐지구력을 측정하는 세 검사인 왕복 오래달리기(PACER), 오래달리기 걷기, 스텝검사가 서로 유사한 구인을 측정하는지 확인하기 위해 상관분석을 실시하였다. <표 11>은 각 종목 간 상관계수( $r$ )를 나타낸 것이다.

표 11. 심폐지구력 검사종목 간 상관관계( $r$ )

학년	성별	n	종목	왕복 오래달리기	오래달리기 걷기
5	남	131	오래달리기 걷기	-.79	
			스텝검사	.40	-.42
	여	112	오래달리기 걷기	-.69	
			스텝검사	.42	-.32
6	남	121	오래달리기 걷기	-.69	
			스텝검사	.23	-.33
	여	104	오래달리기 걷기	-.76	
			스텝검사	.37	-.36
전체	468	오래달리기 걷기	-.76		
		스텝검사	.33	-.38	

<표 11>을 살펴보면 왕복 오래달리기와 오래달리기 걷기 간의 상관은 학년별 성별로 -.69에서 -.79의 범위를 나타내어 검사의 유사성을 확인할 수 있다. 그러나 스텝검사와 다른 두 종목 간 상관은 전반적으로 낮게 나타났다.

#### 2) 백분위 점수 산출

동백분위 동등화를 실시하기 위해서는 각 등급에 해당하는 백분위 값(percentile rank)을 구해야 한다. <그림 5>에서 <그림 8>은 학년별 성별로 각 등급에 해당하는 백분위 값을 공식 3.3.3을 이용하여 산출한 후 세 검사의 백분위 그래프를 함께 그린 것이다. 예를 들어 <그림 5>에서 백분위 43에 해당하는 왕복 오래달리기의 등급이 2등급인 반면 스텝검사의 등급은 약 1.4이고 오래달리기 걷기의 등급은 약 2.3인 것을 알 수 있다. 즉, 왕복 오래달리기 평가등급 2등급에 대해 나머지 두 종목을 동등화 했을 때 스텝검사는 1.4, 오래달리기 걷기는 2.3등급으로 변환된다.

각 그림을 살펴보면 전반적으로 스텝검사의 그래프가 가장 왼쪽에 위치해 있고 오래달리기 걷기 그래프가 가장 오른쪽에 위치한 것을 알 수 있다. 즉 스텝검사 결과의 평가등급이 가장 우수하고 오래달리기 걷기 평가등급이 가장 낮은 것이다. 특히 <그림 6>의 5학년 여학생의 경우 스텝검사 결과에서 4, 5등급에 해당하는 학생이 나타나지 않을 것을 알 수 있다.

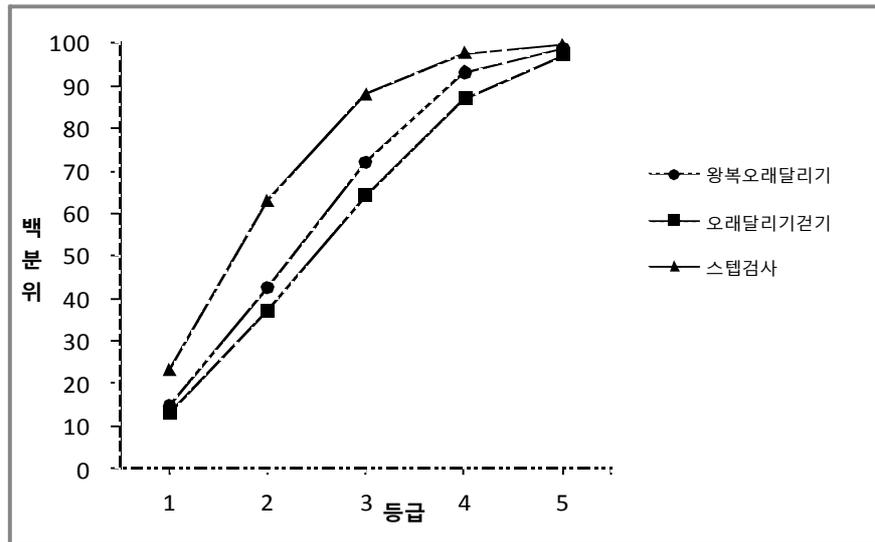


그림 5. 5학년 남학생의 심폐지구력 검사종목 등급의 백분위

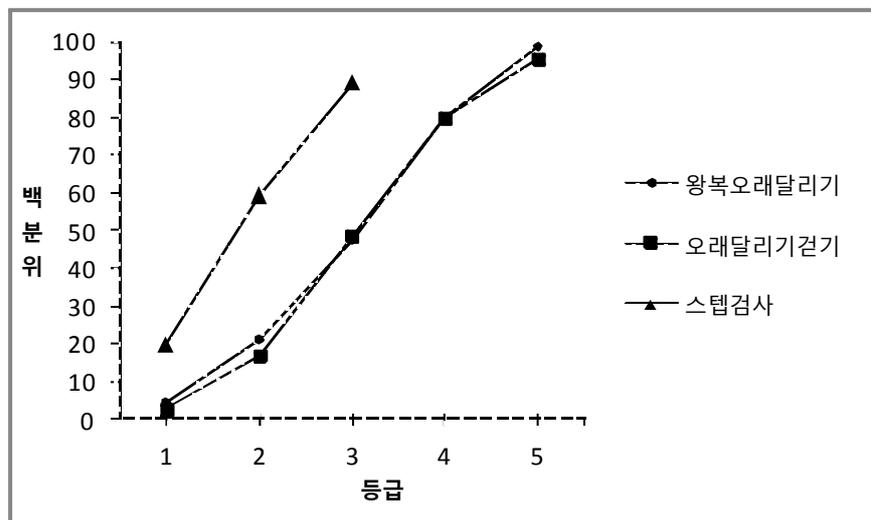


그림 6. 5학년 여학생의 심폐지구력 검사종목 등급의 백분위

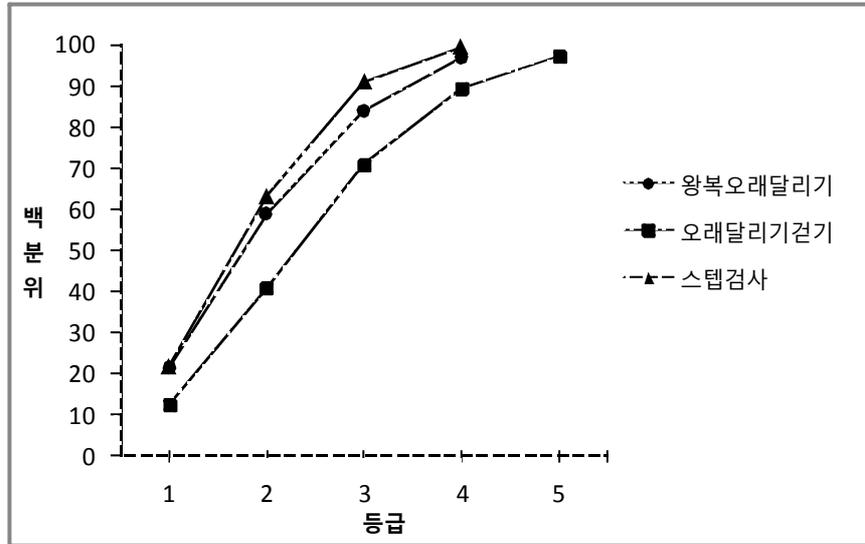


그림 7. 6학년 남학생의 심폐지구력 검사종목 등급의 백분위

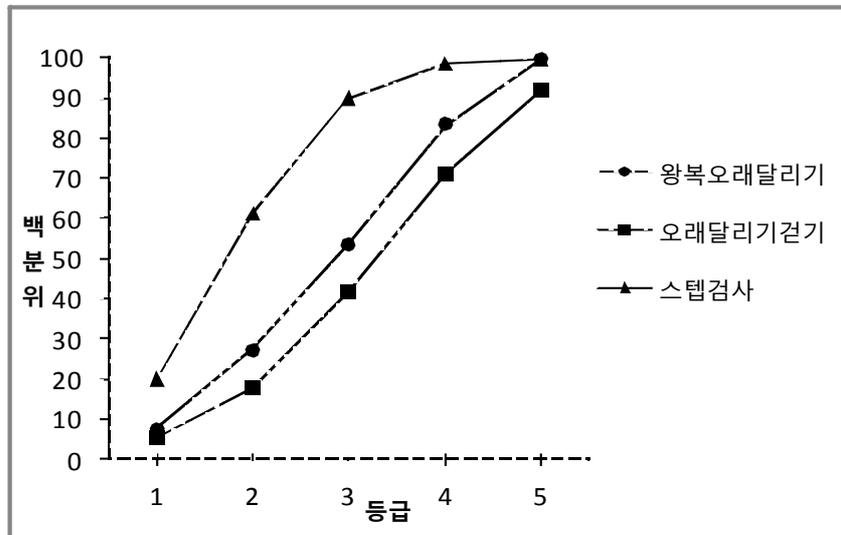


그림 8. 6학년 여학생의 심폐지구력 검사종목 등급의 백분위

### 3) 등급 변환

<표 12>는 공식 3.3.3에 의해 산출된 백분위 자료를 기초로 RAGE(Zeng, Kolen, & Hanson, 2005) 프로그램을 이용하여 사후 완곡화를 거쳐 왕복 오래달리기 등급에 동등화 된 오래달리기 걷기와 스텝검사의 최종 변환등급을 나타낸 것이다. <표 12>를 살펴보면 전반적으로 스텝검사 등급이 동등화하기 전 보다 낮게 변환 된 것을 확인할 수 있다.

표 12. 변환등급

학년	성별	왕복 오래 달리기 등급	변환등급			
			오래달리기 걷기		스텝검사	
			선형	동백분위	선형	동백분위
5	남	1	1	1	1	1
		2	2	2	1	1
		3	3	3	2	2
		4	5	5	2	2
		5	5	5	3	3
		SE	1.23	1.23	0.54	0.54
	여	1	2	2	.	.
		2	2	2	.	.
		3	3	3	1	1
		4	4	4	1	1
		5	5	5	2	3
	SE	0.88	0.95	0.16	0.32	
6	남	1	2	2	1	1
		2	3	3	2	2
		3	5	4	2	2
		4	5	5	3	3
		5	5	5	.	.
		SE	1.20	0.93	0.60	0.60
	여	1	2	2	.	.
		2	3	3	1	1
		3	4	4	1	1
		4	5	5	2	2
		5	5	5	2	4
	SE	1.06	1.06	0.46	0.53	

<표 12>를 살펴보면 5학년 여학생의 스텝검사 5등급, 6학년 남학생의 오래달리기 3등급, 6학년 여학생의 스텝검사 5등급에 대한 결과를 제외하고 선형동등화 결과와 동백분위 동등화 결과가 일치하는 것으로 나타났다. 이 연구에서는 왕복 오래달리기 검사를 기준으로 나머지 두 검사를 동등화 하였기 때문에 왕복 오래달리기와 오래달리기 걷기, 왕복 오래달리기와 스텝검사 간의 등급 호환만 가능하다. 예를 들어 <표 12>에서 5학년 남학생이 오래달리기 걷기 검사에서 5등급을 받았을 경우 왕복 오래달리기 등급으로 변환할 경우 4등급에 해당하게 된다. 마찬가지로 스텝검사서 1등급을 받았다면 이 결과를 왕복 오래달리기 등급으로 변환할 경우 2등급을 받게 된다.

## V. 논의

본 연구에서는 학생 건강체력평가(Physical Activity Promotion System; PAPS)의 심폐지구력 선택종목인 왕복 오래달리기(PACER), 오래달리기 걷기, 스텝검사에서 산출된 등급들을 상호 호환적으로 비교가 가능하고, 어느 종목을 선택하느냐에 관계없이 개인의 체력수준을 일관성 있고 보다 공정하게 평가할 수 있도록 검사등급의 동등화를 실시하였다.

먼저 초등학교 5학년과 6학년 학생들을 대상으로 단일집단 설계에 의해 PAPS의 심폐지구력 선택종목의 평가결과에 차이가 있는지 분석하였다. 분석결과에 의하면 동일한 피험자임에도 불구하고 검사 종류에 따라 결과 등급에 차이가 있는 것으로 나타났다. 특히 스텝검사의 경우 다른 두 검사에 비해 평가등급이 더 높은 것으로 나타났다. 즉, 왕복 오래달리기와 오래달리기 걷기는 피험자가 2, 3, 4등급에 걸쳐 비교적 고르게 분포한 반면 스텝검사는 대부분의 피험자가 1등급과 2등급에 분포하고 있다. 이러한 결과는 선택종목의 검사 난이도에 차이가 있음을 의미한다. 현재와 같이 종목 선택이 가능한 경우 어떤 검사를 하느냐에 따라 서로 다른 평가등급을 받기 때문에 평가의 공정성에 문제가 발생할 수 있다. 김일곤(2011)은 PAPS의 선택종목 간 평가결과에 차이가 있는지 분석하였는데, 본 연구결과와 마찬가지로 스텝검사와 나머지 두 검사인 왕복 오래달리기와 오래달리기 걷기 간 평가등급에 유의한 차이가 있는 것으로 나타났다. 그러나 김일곤(2011)의 연구는 대상이 중학생이었고 스텝검사 등급이 더 낮게 나타나서, 스텝검사 결과가 더 높게 나타난 본 연구결과와는 상반되었다. 본 연구에서는 평가등급 간 차이 뿐만 아니라 세 검사의 평가등급 일치율도 확인하였는데, 가장 높은 일치도 값이 .44에 불과하여 일치도 기준인 .85을 만족시키지 못하는 것으로 나타났다. 즉, 심폐지구력 선택종목 간 평가결과에 유의미한 차이가 있기 때문에 동등화가 필요한 것으로 판단하였다.

검사가 서로 유사한 구인을 측정하지 않는다면 동등화가 불가능하다. 동등화에 앞서 PAPS 심폐지구력 종목이 유사한 구인을 측정하는지 상관계수( $r$ )를 산출하여 검증하였다. 본 연구에서는 왕복 오래달리기와 오래달리기 걷기 간 상관은 비교적 높은 편이었지만, 왕복 오래달리기와 스텝검사, 오래달리기 걷기와 스텝검사는 상관이 낮은 것으로 나타났다. 최우영(2009)의 연구에서도 이와 유사한 결과가 나타났는데, 왕복 오래달리기와 오래달리기 걷기의 상관( $r$ )은 .59로 산출되었으며, 교육과

학기술부(2009) 연구에서는 심폐지구력의 세 검사종목 간 상관( $r$ )이 .3이하로 나타났다. 스텝검사와 다른 두 검사와의 상관성이 낮게 나타났지만 전통적으로 스텝검사가 심폐지구력을 측정하는 검사로 인정받고 있고 현재 PAPS에서 스텝검사가 심폐지구력 선택종목에 포함되어 있으므로 이론적으로 동일한 구인으로 판단하여 동등화를 실시하였다.

이 연구에서 동등화는 선형 동등화와 동백분위 동등화 방법을 사용하였고 왕복 오래달리기 검사를 기준으로 나머지 두 검사인 오래달리기 걷기와 스텝검사를 동등화 하였다. 왕복 오래달리기를 기준 검사로 선정한 이유는 현재 학교 현장에서 심폐지구력 검사종목으로 가장 많이 사용되고 있기 때문이다. 분석 결과에 의하면 선형 동등화와 동백분위 동등화에 의한 변환등급에는 큰 차이가 없는 것으로 나타났다. 또한 오래달리기 걷기 검사는 전반적으로 높은 등급으로 변환되었고 스텝검사는 낮은 등급으로 변환되었다. 즉 검사의 난이도를 고려하여 검사 간 등급이 조정되었다.

이 연구에서 스텝검사의 난이도가 다른 두 검사에 비해 지나치게 낮은 것을 확인할 수 있었다. 스텝검사의 난이도가 낮은 것은 종목 자체의 특성 때문이기도 하지만 스텝 높이가 결과에 영향을 미치기 때문이다. 이명철(1999)의 연구에 의하면 스텝 높이를 5cm를 올렸을 경우 심박수가 평균 10회 이상 증가하는 것으로 나타났다. PAPS 스텝검사에서 초등학생의 스텝 높이는 20.3cm이고 중학생은 45.7cm로 지정되어 있다. 초등학생과 중학생 간 스텝 높이의 차이가 20cm 이상인 것은 성장 발달을 고려하더라도 과도한 차이라고 사료된다. Montoye(1969)는 적정 스텝 높이에 있어 10~12세는 12inch(30cm), 12~14세는 18inch(45cm)로 제시하였고, 퀸스 칼리지 스텝검사에서 제시한 적정 스텝높이는 16.5inch(41.8cm)이며, 일본 문부성에서는 초등학생의 적정 스텝높이를 35cm로 제시하였으므로 현재 PAPS 스텝 높이를 수정할 필요가 있다. 또한 박태섭(2001)에 의하면 현재 체력검사가 학생들의 체격 특성을 고려하지 않았다는 문제점이 발견되므로 단순히 연령별로 스텝 높이를 지정하는 것이 아니라 피험자 개인의 신장을 고려하여 스텝의 높이를 조정하는 것이 필요하다.

본 연구는 PAPS의 심폐지구력 검사종목 간 난이도를 확인하여 어느 종목을 선택하느냐에 관계없이 개인의 체력수준을 일관성 있고 공정하게 평가할 수 있도록 왕복 오래달리기, 오래달리기 걷기, 스텝검사 간 동등화를 실시하여 서로 다른 검사에서 나온 점수들을 상호 비교가 가능하도록 등급을 변환하였다. 향후 연구에서는 유연성, 근력 등 다른 종목에서도 동등화가 이루어져야 할 것이다. 또한 초등학

생 뿐만 아니라 중, 고등학생을 대상으로 보다 많은 표본을 수집하여 평가등급의 동등화가 이루어져야 할 것이다. 현재 PAPS 평가등급의 차이가 규명된 시점에서 종목별, 성별, 학년에 대한 여러 종목의 타당성 검증을 통한 종목의 선택적 적용이 가능한 연구가 지속적으로 이루어져야 할 것이다. 또한 이러한 문제를 해결하기 위해서 교육청 단위 또는 개별학교 단위로 종목을 일치시켜 측정하는 현실적인 방안을 검토해야 할 것이다.

## VI. 결론

본 연구는 초등학교 5학년, 6학년을 대상으로 학생 건강체력평가(PAPS)의 심폐지구력 선택종목인 왕복 오래달리기(PACER), 오래달리기 걷기, 스텝검사 간 동등화를 실시하여 공정한 평가를 지침을 제공하는 것이 목적이다.

이 연구를 수행한 결과 다음과 같은 결론을 도출하였다.

1) 심폐지구력의 왕복 오래달리기, 오래달리기 걷기, 스텝검사 간 평가등급의 차이를 검증하기 위해 빈도분석을 실시한 결과, 모든 집단이 등급 간 유의한 차이가 있었다( $p < .001$ ).

2) 심폐지구력 선택종목이 서로 유사한 구인을 측정하고 있는지 상관분석을 통해 검증한 결과, 왕복 오래달리기와 오래달리기 걷기 간 상관은 비교적 높았지만 왕복 오래달리기와 스텝검사, 오래달리기 걷기와 스텝검사는 상관이 낮았다.

3) 심폐지구력 선택종목의 선형 동등화, 동백분위 동등화는 5학년 여학생의 스텝검사 5등급, 6학년 남학생의 오래달리기 3등급, 6학년 여학생의 스텝검사 5등급 결과를 제외하고 선형 동등화 결과와 동백분위 동등화 결과가 일치하였다.

4) 학생 건강체력평가의 심폐지구력 선택종목을 동등화 한 결과, 오래달리기 걷기 검사는 전반적으로 등급이 상향 조정되었고, 스텝검사는 하향 조정되었다. 즉 검사의 난이도를 고려하여 검사 간 등급이 조정되었다.

## 참고문헌

- 고기환, 강상조(1997). 배구 서브 기능검사 점수의 동등화를 위한 IRT 적용. **한국체육학회지**, 36(3), 283-295.
- 강상조(1999). 윗몸일으키기 검사(sit-up)의 동등화와 검사차별기능 검증. **한국체육학회지**, 38(1), 683-697.
- 교육과학기술부(2008). **학생건강체력평가제 확대시행 계획**. 학생건강안전과.
- 교육과학기술부(2008). **PAPS 시범학교·학교스포츠클럽 운영 보고회**. 교육과학기술부.
- 교육과학기술부(2009). **학교건강 체력증진을 위한 PAPS 효율적 운영 방안**.
- 교육과학기술부(2009). **학생건강체력 평가 매뉴얼**. 교육과학기술부.
- 김경훈(2011). **고등학교 남학생의 건강체력평가시스템(PAPS) 종목별, 학년별 평가 등급 분석**. 석사학위 논문. 목원대학교 교육대학원.
- 김기학, 김기봉, 최민동, 허정, 이동수, 박정화, 조국래, 김현경, 정도상(2007). **체육 측정평가**. 서울: 형설출판사.
- 김세정(2013). **초등학교 학생건강체력평가제(PAPS) 운영 실태 및 개선 방안**. 석사학위 논문. 한국교원대학교 교육대학원.
- 김아영, 이규민, 반재천, 차정은(2004). 성인용 직업적성검사 연계화. **교육평가연구**, 17(2), 107-129.
- 김일곤(2011). 신체활동증진시스템(PAPS)의 심폐지구력 유형별 평가기준의 차이 분석. **한국체육과학회지**, 20(3), 1563-1572
- 김현철(1999). 1999학년도 대입수능시험 선택과목 원점수와 선형 또는 비선형 표준점수의 비교 및 표준화 효과의 실증분석. **교육학연구회**, 37, 219-243.
- 김현철(1999). 척도화 방법들의 무작위 오차크기의 실증비교. **교육평가연구**, 12(2), 148-168.
- 김현철(2000). 순환동등화에 의한 2000학년도 대입수능시험 선택과목 점수 동등화 결과의 비교. **교육평가연구**, 13(2), 153-168.
- 기선경(2013). **학생 건강체력 평가시스템을 활용한 스포츠 텔런트 평가 방안 모색**. 미간행 박사학위 논문. 한국체육대학교 대학원.
- 권오형(2013). **중등 시각장애 학생과 비장애 학생의 학생건강체력평가 점수 비교**

- 분석. 미간행 석사학위 논문. 용인대학교 교육대학원.
- 나법경(2014). **학생건강체력평가시스템(PAPS)을 통한 중학생의 체력분석**. 석사학위 논문. 명지대학교 대학원.
- 남덕현(2002). 심폐지구력 검사 동등화. **한국사회체육학회지**, 17, 183-195.
- 남덕현(2002). 오래달리기 검사의 동등화 표준오차 검증. **인문사회논총**, 8, 73-99
- 남덕현, 윤상화, 김도준, 박재우(1999). 동등백분위법을 이용한 순발력 검사 방법간의 동등화 방안. **용인대학교 체육과학연구논총**, 9(1), 113-128.
- 남현우(2003). 수능의 탐구영역 내 과목 간 점수조정에서 합 집단 연계화방법의 적용가능성. **교육평가연구**, 16(1), 163-182.
- 남현우(2001). **검사동등화 방법**. 서울: 교육과학사.
- 문화체육부(1992). **학생체력검사 제도 개선 연구**. 체육청소년부
- 박문식(1998). **오래달리기 기록의 거리 간 동등화 방안**. 석사학위 논문. 수원대학교 교육대학원.
- 박태섭(2001). 초등학생의 체격발육을 고려한 체력평가 기준표 개발. **한국발육발달학회지**, 9(1), 21-34.
- 부재울(1999). 검사동등화 시의 고려사항. **교육평가연구**, 12(1), 43-64.
- 배성웅(2014). **학생건강체력평가시스템(PAPS)의 종합체력점수와 학업성적의 관계에 관한 종단적 연구**. 석사학위 논문. 서강대학교 교육대학원.
- 배소라(2013). **초등학생의 학생건강체력 평가(PAPS)에 따른 신체적 자기개념**. 석사학위 논문. 서울교육대학교 교육대학원.
- 성태제(2002). **타당도와 신뢰도**. 서울: 학지사
- 성태제(2000). 다른 교과와 점수를 연결하는 통계적 조절 방법의 타당성. **교육평가연구**, 13(1), 63-99.
- 양인철(2010). **초등학교 고학년의 비만이 학생건강체력평가 측정 결과에 미치는 영향**. 석사학위 논문. 한국체육대학교 교육대학원.
- 엄용현(2014). **학생 건강체력평가(PAPS)를 활용한 중학생의 건강체력과 비만 결정요인 분석**. 석사학위 논문. 한국체육대학교 대학원.
- 이종성(1985). **행동과학 연구를 위한 측정이론의 기초**. 서울: 중앙적성출판사.
- 이종성(1990). **문항반응이론의 기초**. 서울: 대광문화사.
- 이경수(2012). **초등학생의 신체질량지수(BMI)가 건강 체력평가에 미치는 영향**. 석사학위 논문. 건국대학교 교육대학원.
- 이규민(2005). **초등학교 3학년 국가수준 기초학력 진단평가 동등화 방안**. 교육평가

- 연구, 18(1), 125-151.
- 이명철. (1999). Step test 시 고관절 굴곡각에 의한 승강대 높이 적용의 타당성 연구. *한국체육과학회지*, 8(2), 461-474.
- 이해의(2014). *중학교 축구 선수의 심폐지구력 수준평가: PAPS의 활용*. 석사학위 논문. 한국체육대학교 교육대학원.
- 임정화(2001). *학생건강체력평가 운영에 대한 중학교 체육교사의 인식 연구*. 석사학위 논문. 이화여자대학교 교육대학원.
- 정재윤(1999). *순발력 검사의 동등화*. 박사학위 논문. 한국체육대학교.
- 최인엽(2011). *PAPS를 활용한 초, 중, 고 학생의 건강 체력과 학업 성취도*. 석사학위 논문. 한국체육대학교 교육대학원.
- 최우영(2009). *학생건강체력평가의 종목선정에 따른 체력급수의 차이*. 석사학위 논문. 인천대학교 교육대학원.
- 한태호(2011). *척수장애인의 유연성 평가를 위한 동등화*. 석사학위 논문. 한국체육대학교 일반대학원.
- Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike(Ed.), *Educational measurement (2nd edition)*. Washington D. C.: American Council on Educational
- Angoff, W. H. (1984). *Scales, norms, and equivalent scores*. Educational testing service.
- Hambleton, R. K.(Ed.)(1983). *Applications of IRT*. Vancouver: Educational Research Institute of British Columbia.
- Hambleton, R. K., & Swaminathan, H. (1985). Item response theory: *Principles and applications (Vol. 7)*. Springer Science & Business Media
- Hambleton, R. K., & Plake, B. S. (1995). Using an extended Angoff procedure to set standards on complex performance assessments. *Applied Measurement in Education*, 8(1), 41-55.
- Harris, D. J., & Kolen, M. J. (1990). A comparison of two equipercentile equating methods for common item equating. *Educational and Psychological Measurement*, 50(1), 61-71.
- Holland, P. W. & Rubin, D. B.(Ed)(1982). *Test equating*. NY: Academic Press
- Holland, P. W., & Thayer, D. T. (1987). Notes on the use of log-linear models for fitting discrete probability distributions (Technical Report 87-79).

- Princeton, NJ: ETS.
- Holland, P. W., King, B. F., & Thayer, D. T. (1989). *The standard error of equating for the Kernel Method of equating score distributions* (TR No. 89-83). Princeton : ETS.
- Jarjoura, D. & Kolen, M. J. (1985). Standard errors of equipercentile equating for the common item nonequivalent populations design. *Journal of Educational Statistics, 10*(2), 143-160.
- Kolen, M. J. & Jarjoura, D. (1987). Analytical smoothing for equipercentile equating under the common item nonequivalent populations design. *Psychometrika, 52*(1), 43-59
- Kolen, M. J. (1991). Smoothing methods for estimating test score distributions. *Journal of Educational Measurement, 28*, 257-282.
- Kolen, M. J., & Brennan, R. L. (1995). *Test Equating: Methods and Practices*. New York: Springer.
- Kolen, M. J., & Brennan, R. L. (2004). *Test Equating, Scaling, and Linking: Methods and Practices*(2nd ed.). New York: Springer.
- Kolen, M. J. (1984a). Effectiveness of analytic smoothing in equipercentile equating. *Journal of Educational Statics, 9*(1), 25-44.
- Mislevy, R. J. (1992). *Linking educational assessment: Concepts, issues, methods and prospects*. Princeton, NJ: Educational Testing Service, Policy Information Center.
- Montoye, H. J., Willis III, P. W., Cunningham, D. A., & Keller, J. B. (1969). Heart rate response to a modified Harvard step test: males and females, age 10 - 69. *Research Quarterly. American Association for Health, Physical Education and Recreation, 40*(1), 153-162.
- Petersen, N. S., Kolen, M. J., & Hoover, H. D. (1989). *Scaling, Norming, and equating*. In R. L. Linn(Ed.), *Educational measurement* (3rd ed.). NY: Macmillan.
- Linn, R. L. (1993). Linking results of distinct assessment. *Applied Measurement in Education, 6*, 83-102.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.

- Lord, F. M. (1981). *Standard error of and equating by IRT*. Princeton: ETS.
- Lord, F. M., & Wingersky, M. S. (1984). Comparison of IRT true-score and equipercentile observed-score equatings. *Applied Psychological Measurement*, *8*, 452-461.
- Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Journal of Applied Psychological Measurement*, *7*, 201-210.
- Wingersky, M. S. & Lord, F. M. (1984). An investigation of methods for reducing sampling error in certain IRT procedures. *Applied Psychological Measurement*, *8*(3), 347-364.
- Zeng, L. (1991). *Standard errors of linear equating for the single-group equating*(ACT Research Reaport series). American College Testing.
- Zeng, L., Kolen, M. J., & Hanson, B. A. (2005). *RAGE*. [Computer Program]. Iowa City, IA: University of Iowa.
- Zhu W. (2002). *Physical activity assessments for health-related research*. Champaign, IL: Human Kinetics.
- Zhu W. (2001). An empirical investigation of Rasch equating of motor function task. *Adapted Physical Activity Quarterly*, *18*, 72-89.
- Zhu, W. (1998). Test equating: what, why, how. *Research Quarterly for Exercise and Sport*, *69*, 11-23.

## Abstract

### Equating Among Cardiovascular Endurance Test Item of Physical Activity Promotion System(PAPS)

Kim, Gyeong-o

Physical Education Major  
Graduate School of Education  
Jeju National University  
(Adviser: Professor Kim, Miye)

Initially the students' physical fitness test system was decreed in 1951 by the Education Ministry to promote students' physical fitness. In 2009, Physical Activity Promotion System(PAPS) which was composed mainly of health-related physical fitness test was developed(Ministry of Education, Science and Technology, 2009). Physical Activity Promotion System(PAPS) is evaluation measure that can be utilized as scale of both physical fitness and obesity items. The significant difference of 2 scales is that the latter has not only multiple items of physical strength but also can choose one item arbitrarily when comparing the latest revised PAPS(2009) with the former. That is, The latter has advantage that the most suitable item can be chosen and measured relative to environment and condition but disadvantage that cannot be compared with the result directly even though the same measure. Particularly when the level of difficulty among items are different, the problem of fairness can rise. Accordingly the equating test when the scores from different tests can be compared on the basis of the inter-compatible ground and the individual's level of physical strength, can be evaluated consistently and fairly no matter which item was chosen. Therefore the purpose of the study was to provide fairer evaluation guideline by confirming the levels of difficulty of the items and conducting equation among the items on the assumption of measuring PAPS'

cardiovascular endurance optional items, such as distance run, walk test, PACER and step test. To achieve the aim, the subjects participated in the study composed of total 486 elementary students from each 2 schools of Jeju-Si and 4 schools of SeogWipo-Si(male=131, female= 112 of 5<sup>th</sup> grade and male=121, female=104 of 6<sup>th</sup> grade). The difference between the rating of the optional items was verified through the frequency analysis and the conformity degree. The analyzed results showed rather higher correlation among PACER, distance run and walk test, but rather lower correlation among PACER, step, distance run and step test. It was found that there were no big differences in converted grades by linear equating and percentile equating, and the difficulty level of the test of distance run and walk of equated converted grade was relatively high and it was converted to equal position as other tests, i.e. in general, converted to be high. And, the difficulty level of the step test was relatively very low, so in general it was converted to be low. The new rank among distance run and walk test and step test which was equated from PACER showed rather more stable result than the result of previous rank. In this research, it was confirmed that the difficulty level of step tests was lower than that of the other two tests. By the nature of step test, the height of step can have big effects on measurement results, so it is necessary to adjust the height of step in PAPS. In this research, only equating between cardiovascular test of PAPS were conducted, but in follow-up researches, equating between the other tests such as flexibility, muscular strength, etc. should be conducted. Also, equating of evaluated grades should be attained through collecting more samples not only from elementary school students but also from junior high and high school students.