

## 스포츠심리학에서 상황판단검사의 필요성과 채점방식에 따른 타당성 비교

신명진\*(서울대학교)

공통방법편의(CMB: Common Method biases)는 측정도구의 문제가 아니라 측정 시 유발되는 다양한 오류들을 의미한다. 이를 극복하기 위한 질차적 방안 중 하나는 측정방법의 분리이지만 최근 10년 동안 스포츠심리학회지에 투고된 논문들을 분석한 결과 측정방법의 분리가 이루어진 논문은 총 197편 중 24편이었고, 대부분의 측정방법은 리커트 유형이었다. 따라서 본 연구에서는 리커트 유형의 질문지 이외의 다른 방법으로 심리변수를 측정할 수 있는 상황판단검사(SJT: Situational Judgement Test) 방법을 소개하고 개발과정과 기존 연구결과들을 정리하였다. 이와 더불어 스포츠 영역에서 개발된 상황판단검사 도구를 바탕으로 채점 방식에 따른 타당도를 비교한 결과 기존의 거리점수 만을 고려한 방식보다 순서점수를 함께 고려한 방식에서 우수한 타당도를 보였다.

주요어: 공통방법편의, 상황판단검사, 채점방식

### 서론

질문지 개발은 내용타당도, 구성타당도, 준거타당도, 수렴타당도, 문항내적 합치도, 검사/재검사 신뢰도 등 여러 타당도와 신뢰도 검증을 통해서 이루어진다. 이런 질문지를 바탕으로 운동경쟁상황과 신체활동에 따른 심리변수를 측정하여 그 관계성을 살펴보는 것은 스포츠심리학의 주요 연구 방법 중 하나이다. 그러나 최근 들어 질문지 측정 방법의 오류인 공통방법편의(CMB: Common Method Biases)의 존재로 인하여 측정 질문지의 타당성과 신뢰성의 문제가 아니라 측정 시 유발되는 다양한 편향들이 연구결과에 부정적 영향을 미치는 것으로 나타났다(신명진과 이근철, 2014; Cote & Buckley, 1987; Sharma et al., 2009).

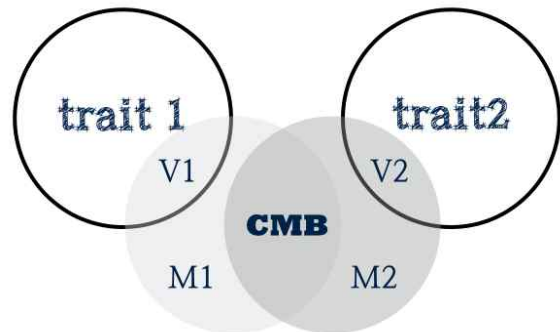


그림 1. 공통방법편의(CMB) 발생원인

CMB는 측정도구의 문제가 아니라 측정방법 시 유발되는 오류들을 의미한다(Podsakoff, et al., 2003). Podsakoff와 Organ(1986)은 CMB의 발생을 <그림 1>과 같이 현상학적으로 설명하고 있다.

<그림 1>에서 trait 1을 측정할 측정치(M1)와 trait 2를 측정하는 측정치(M2)간의 공통된 부분(V1, V2)이 존재하고 V1과 V2는 측정도구의 타당도이다. trait 1과 trait 2 간에는 상관성이 존재하지 않지만, trait 1을 측정할 측정치(M1)와 trait 2를 측정할 측정치

논문 투고일: 2015. 01. 14.

논문 수정일: 2015. 03. 12.

게재 확정일: 2015. 03. 16.

\* 저자 연락처: tennis0405@gmail.com

(M2)에 의해서 공통된 부분이 존재하여 실제 현상과 다른 결과가 도출되는데 이를 CMB라고 한다.

따라서 측정도구의 문제가 아니라 참여자의 기분 및 감정 상태(기쁨, 즐거움, 불쾌함 등), 측정시기(승리 후, 패배 후) 등과 같은 심리적, 환경적 제약 등이 측정 시 유발될 수 있는 여러 오류들이 CMB이다. Cote와 Buckley(1987)의 70편의 다특질다측정방법(MTMM) 연구(심리-사회학, 마케팅, 경영, 교육학)에 따르면, 측정분산 중 26.3%가 CMB가 존재하였고, 이중 마케팅 분야가 15.8%로 가장 낮고 교육학 분야는 30.5%로 가장 높았다. 그리고 측정변수 중 업무 수행력(job performance)은 22.5%, 태도(attitude)에는 40.7%의 CMB가 존재하였고, 공통방법편의가 존재할 때 연구 모형의 설명량은 35%인 반면 공통방법편의를 고려한 경우는 11%로 낮아졌다. Podsakoff et al.(2003)의 연구에서는 두 변수 간의 실제 상관성이 1이어야 하지만 CMB로 인하여 .52~.54였고, 변수 간 상관성이 실제 '0'일 경우에는 .07~.23의 상관성이 존재하였다. 체육학에서 이루어진 신명진과 이근철(2014)의 연구에서도 사전연구와 유사한 6.6%~36%의 CMB가 관찰되었다. 따라서 CMB로 인하여 연구 결과의 타당도와 신뢰도에 부정적 영향을 미치고 있음을 여러 연구들을 통해서 관찰되었고, 이를 해결하는 노력이 필요할 것이다.

Podsakoff et al.(2003)은 CMB의 발생원인을 응답자에 의한 편향, 설문문항에 의한 편향, 문항맥락에 의한 편향, 측정맥락에 의한 편향으로 구분하였고, 이를 해결할 수 있는 방안도 함께 제시하였다. 먼저 통계적 해결방안으로는 'Harman의 단일요인 테스트', '부분상관 통제법', '잠재측정요인 효과조절법', '단일 비잠재측정 효과조절법', '다방법요인 방법분산 조절법', 'CU 모형(Correlated Uniqueness model)' 등이 있고, 절차적 방안에는 '측정원의 다양성', '익명성 보장', '문항의 모호성 제거', '종속변수와 독립변수 간의 서로 다른 척도점 사용', '측정방법의 분리' 등이 있다. 통계적 해결방안에 관한 연구들은 여러 연구들이 진행되었고(Kline et al., 2000; Sharma et al., 2009; Williams et al., 1996), 절차적 방안에서 '측정방법의 분리'를 제외

한 기타 방법들은 연구자들의 노력으로 해결할 수 있다. 그러나 측정방법의 분리'는 질문지 측정방법 이외에 타당도와 신뢰도가 확보된 다른 측정방법이 존재하지 않는 스포츠심리학 연구의 현 상황에서는 CMB를 극복하기 위한 절차적 방안으로 적절하지 못하다.

이런 현실적 어려움은 2005년부터 현재까지 스포츠심리학회지 게재된 논문들을 분석한 결과를 통해서 알 수 있다. 지난 10년간 스포츠심리학회지에 게재된 총 515편의 논문 중, 운동 학습, 제어 및 발달을 제외한 건강운동 및 응용 스포츠심리학 분야의 논문은 376편이었다. 이 중에서 변수들 간의 관계를 살펴본 양적 연구는 196편이었으며 방법분리를 적용한 논문은 24편에 불과하였다. 방법분리가 이루어지지 않은 172편은 리커트 질문지로 여러 개의 변수를 측정한 반면, 방법분리가 이루어진 24편의 논문들은 독립변수와 종속변수 간의 측정방법이 리커트 질문지와 대화성적, 운동참여 빈도, 뇌파, 체력, 체격, 비만도 등의 측정방법을 적용하여 측정방법이 서로 달랐다. 그러나 스포츠심리학은 조작적 정의에 의해서 다양한 심리변수를 측정하여 연구가 진행되어야하므로 24편에서 이루어진 리커트 질문지 방법을 제외한 다른 측정방법은 스포츠 수행, 신체적 변화, 운동참여 행동 영역 등과 같이 매우 제한적 영역에서만 사용가능한 한계점이 존재하였다.

측정도구의 문제가 아니라 측정방법에 의해서 연구결과에 편향을 유발할 수 있는 CMB의 존재는 연구의 타당도와 신뢰도에 위협이 된다. 이를 해결할 수 있는 방안 중 하나는 측정방법의 분리이지만, 스포츠심리학에서는 동일한 방법(리커트 척도)으로 변수들을 측정하고 있으며, 제한적 영역과 논문에서만 측정방법분리가 이루어지고 있다.

본 연구에서는 리커트 유형이외의 다른 측정방법을 소개하고자 하며, 그것이 바로 본 연구의 핵심주제 중 하나인 상황판단검사(SJT: Situational Judgement Test) 기법이다. SJT는 최근 조직심리학에서 관심을 받는 검사 도구로써(Weekley & Plyhart, 2006), 업무영역에서 경험할 수 있는 상황에 대한 적응력을 살펴보는 측정도구이다. SJT는 조직심리학뿐만 아니라 교사

(백순근과 임현수, 2006), 군인(Hedlund et al., 2003) 등과 같이 특수영역에서부터 일반인의 정서지능 측정(Sharma et al., 2013)까지 측정 분야와 대상이 다양하다. 하지만 스포츠영역에서 개발된 SJT검사 도구가 전무하며 이에 대한 기본적 개념과 개발과정을 소개한 연구 또한 부족한 실정이다. 운동선수들이 경험하는 환경과 상황은 일반인들이 경험할 수 없는 특수성이 존재하므로 이를 바탕으로 개인의 다양한 심리적 변화를 측정할 수 있는 SJT는 리커트 유형의 검사지를 대처할 수 있어 CMB 해결에도 도움이 될 것이다.

따라서 본 연구에서는 리커트 척도의 질문지 측정방법과 분리되고 다양한 심리변수도 측정할 수 있는 SJT 개발과정과 기존 연구결과들을 정리하였다. 이와 더불어 스포츠 영역에서 개발된 상황판단검사를 통해서 채점방식에 따른 타당도를 비교하여 기존 채점방식을 보완할 수 있는 방안도 함께 제시하였다.

## 상황판단검사 소개

상황판단검사는 2차 세계대전에서 처음 사용되었고(Guilford & Lacey, 1947), 특정 상황을 묘사하는 지시문(stems)을 읽고 가장 적절한 답변문항(responses)을 선택하는 방법이다(Motowidlo et al., 1997; Thornton & Cleveland, 1990). 상황판단검사는 지시문 제시형태에 따라 지필형, 인터뷰형(Latham et al., 1980), 비디오형(Dalessio, 1994; Jones & DeCotiis, 1986)으로 구분될 수 있다. 본 연구에서는 지필형 상황판단검사를 중심으로 설명하였는데 그 이유는 인터뷰형과 비디오형 상황판단검사에 관한 연구가 부족하고, 상황판단검사에 대한 메타분석을 실시한 연구들(McDaniel et al., 2007; McDaniel & Nguyen, 2001)도 지필형 상황판단검사를 위주로 정리되었기 때문이다.

상황판단검사의 개발과정은 3단계에 걸쳐 이루어진다(강민우 등, 2005; 정현선과 박동진, 2009; Anderson & Wilson, 1997; Motowidlo et al., 1997). 1단계는 중요한 사건(critical incidents)을 현직 전문가들로부터 수집하여 지시문을 구성하는 단계이다. McDaniel과 Nguyen(2001)에 따르면 지시문은 다음과 같은 특성을 고려하여 제작되어야 한다고 하였다. 첫째, 사실성

(fidelity)을 고려해야 한다. 예컨대, 제시된 지시문으로 특정 상황에 대한 묘사를 응답자에게 전달할 때, 비디오로 전달하는 방법이 글로 전달하는 방법 보다 사실성이 높으므로 응답자의 문항에 대한 이해도 또한 향상되기 때문이다. 둘째, 지시문의 길이를 고려해야 한다. 셋째, 측정자들이 지시문을 쉽게 이해(comprehensibility)할 수 있어야 한다. 넷째, 지시문이 묘사하는 상황은 단순하면서도 명확한 상황들로 이루어져야 하므로 지시문의 복잡성(complexity)을 고려해야 한다. 2단계는 1단계에서 개발된 중요사건을 개선하는데 요구되거나 나타날 수 있는 다양한 행동반응과 행동대안을 수집하여 답변문항을 구성한다. 마지막 3단계는 채점용 답(scoring key)을 개발하는 것으로 채점방식은 크게 지시문 상황(시나리오 또는 갈등상황)에 대해서 주제전문가(SME: Subject Matter Expert)가 평정한 평균평정 점수 또는 응답집단의 평균점수를 사용한다.

상황판단검사는 직업 수행력, 인지능력, 성격 검사와 상관성이 높은 특성이 존재하고(Chan & Schmitt, 2002; Clevenger et al., 2001), 거짓응답(faking resistant)의 가능성도 배제할 수 없다(Nguyen et al., 2005). 거짓응답은 응답자들이 높은 점수를 받기 위해서 진심을 속이는 것으로 성격 5요인 중 정서적 안정성과 양심성이 높은 사람이 거짓 응답할 가능성이 낮은 것으로 나타났다(McFarland & Ryan, 2000). Nguyen et al., (2005)의 연구에 따르면 동일한 상황판단검사를 학술적 목적으로만 사용되므로 최대한 진실하게 답변을 요구한 집단(진실상황 집단)과 본인이 기업에 지원하고 본 상황판단검사가 합격여부에 영향을 준다고 설정한 집단(거짓상황 집단)으로 구분하여 연구를 실시하였다. 그 결과 거짓상황 집단에서 거짓응답의 효과가 나타났다.

상황판단검사가 거짓응답 유발의 가능성 문제를 지니고 있지만, 여러 연구들을 통해서 준거타당도와 구성타당도가 증명되었다(McDaniel & Whetzel, 2005; Sharma et al, 2013). 상황판단검사를 통한 실천지능 측정 도구들(CLQ: College Life Questionnaire, CSQ: Common Sense Questionnaire, ESJI: Everyday situational Judgment Inventory)에서 미국과 스페인 참여자들 모두 유사하였으므로 인종에 따른 측정도구의 모형 안정성도 검증되었다(Cianciolo et

al., 2006). 국내 연구에서는 백순근과 임현수(2006)가 중학교 교사들의 실천지능 측정도구로써 상황판단검사를 통해서 개발하였고, 구성타당도와 신뢰도가 우수한 것으로 나타났다.

### 지시문 유형과 타당성

McDaniel과 Nguyen(2001)는 상황판단검사의 지시문을 지식형(knowledge)과 행동경향(behavioral tendency)으로 구분하였다. 지식형 지시문은 각 지시문에 최선(best)이라고 생각되는 반응과 최악(worst)이라고 생각되는 반응을 모두 고르는 방식이고, 행동경향 지시문은 각 상황에 나타난 행동들에 대해서 각 대안 중 가장 할 것 같은 대안(most)과 하지 않을 것(least likely) 같은 대안을 모두 고르는 방식이다(강민우 등, 2005; McDaniel & Nguyen, 2001). McDaniel et al.(2007)은 지시문 유형에 따라 인지능력과 성격(Big 5)간의 관련성이 존재할 뿐만 아니라 준거타당도에도 영향을 미친다고 하였다. 먼저 인지능력과 지식형 지시문 간의 관련성은 .35인 반면, 행동경향 지시문은 .17이었는데, 그 이유는 지식형 지시문은 최대능력 검사(maximal performance)이고, 행동경향 지시문은 전형적 수행검사(typical performance)이므로 지식형 지시문은 인지적 요인과 관련성이 존재한다고 하였다. 최대능력검사는 학력검사나 지능검사와 같이 인간의 지적능력영역을 측정하고자 하는 검사이고, 전형적 수행검사는 시간제한이 엄격히 규정되지 않는 비속도검사를 뜻한다. 행동경향 지시문과 big 5간의 관련성이 지식형 지시문보다 높게 나타나 행동경향 지시문은 성격 등과 같이 비인지적 요인에 영향을 받으므로 사회적 바람직성(자기기만, 인상관리)에 노출될 수 있다고 하였다. 업무수행력(job performance)에 대한 준거타당도는 .26으로 McDaniel et al.(2007) 연구에서의 .34보다는 낮게 나왔으며 선택지 유형 간의 차이점이 존재하지 않았다. 그러나 메타분석을 실시한 논문 중 상황판단검사 두 유형을 모두 포함한 연구들을 고려한 후 분석한 결과, 지식형 지시문과 업무수행력 간의 관련성은 .26, 행동경향 지시문은 .12의 상관성이 나타났다. 따라서 지식

형과 행동경향 지시문을 중복으로 제공한 연구를 포함한 경우에는 지식형 지시문이 행동경향 지시문보다 준거타당도가 높았고, 그렇지 않은 경우에는 통계적 차이가 존재하지 않았으므로 어떤 지시문이 상황판단 검사에 적합하다고 판단하기는 힘들 것이다.

상황판단검사의 지식형 지시문이 행동경향 지시문보다 인지능력과 관련성이 높고, 행동경향 지시문은 성격과 상관성이 높았는데 강민우 등(2005)은 그 이유에 대해서 다음과 같이 해석하였다. 그들은 지식형 지시문의 경우 주어진 상황을 해결하기 위해서 어떠한 행동전략이 효율적인가를 인지적으로 판단하여 제시하는 대안 반응을 선택하기 때문이고, 행동경향 지시문의 경우 참여자는 인지적 판단보다 과거의 행동에 비추어 자신의 행동이 어떤 대안으로 행동할지를 예상하고 그러한 방향으로 대안을 결정하여 습관이나 과거의 행동, 성격과 같은 부분이 앞으로의 행동에 영향을 미치는 것으로 보았다.

### 채점방식과 타당성

상황판단검사의 채점기준 선정 방법은 3가지로 구분할 수 있다(정현선과 박동건, 2009; McDaniel & Nguyen, 2001). 첫째, 주제관련 전문가 또는 고성과자(excellent employee)가 제시된 상황에 대한 응답들에 대해서 효과성을 평정하도록 한 후, 이에 대한 합의 과정을 통해서 결정된 점수를 각 행동반응에 대한 채점용 답으로 설정한다. 둘째, 설문을 통해 개인반응 대안에 대한 효과성을 평정 받은 후, 평균점수를 효과성 점수로 설정하거나 효과적 또는 효과적이지 않은 반응들만을 각각 조합하여 최적/최악의 반응만으로 채점용 답을 확정하는 방법이다. 셋째, 경험적 방식으로 전기자료(biodata)에 대한 채점용 답 결정방식을 활용하는 것이다. 정현선과 박동건(2009)은 응답자들이 각 지시문에 대한 행동반응 문항에 부여한 점수와 준거(상사평점 점수)간의 상관분석을 통해서 각 행동반응에 대한 순위를 결정하였다. 따라서 관계성이 정적으로 가장 크게 나온 값이 가장 적절한 대처 방안이고 부적으로 큰 값이 최악의 대처 방안이 된다.

이와 같이 상황판단검사의 채점기준 선정은 전문가

집단의 판단과 경험이 중심인 경우, 표본의 중심경향 값 중심 그리고 특정 준거에 근거하는 방법을 적용하는데 상황판단검사의 대부분은 전문가 집단의 평점을 중심으로 이루어지고 있다(백순근과 임현수, 2006; Hedlund et al, 2003). 전문가 집단을 중심으로 이루어지는 채점방식은 시나리오 채점과 반응대안 채점으로 구분할 수 있다(강민우 등, 2005).

시나리오 채점방식은 전문가와 응답자가 같은 시나리오(지시문)와 행동반응을 제시하고 가장 할 것 같은 행동, 가장 하지 않을 것 같은 행동 2가지만을 선택하도록 한 후, 주제전문가에 의해서 선택된 혹은 합의된 가장 할 것 같은 행동을 응답자가 동일하게 선택하였다면 1점, 전문가가 가장 하지 않을 것 같은 행동 대안을 응답자가 선택하면 -1점, 다른 행동 대안을 선택하면 0점이 된다(Motowidlo et al., 1990). 따라서 전문가가 선택한 최고와 최악의 선택문항이 동일하면 2점, 서로 상반된 문항을 선택하면 -2점으로 총 점수의 분포는 -2 ~ 2점이다. 반응대안 채점은 각 행동반응 별로 주제전문가로부터 얻은 최고, 최악 행동반응에 대한 평균평점과 응답자가 최고, 최악이라고 선정한 응답반응 간의 차이값을 응답자의 점수로 부여하는 방식이다. 이 방식은 첫 번째 시나리오 방식보다 점수범위가 더 큰 점수범위 차이를 가지므로 응답자의 차이를 최대로 할 수 있는 장점이 있다(Knapp et al., 2001). 최근 상황판단검사에서 사용되고 있는 채점방식은 응답자가 각 행동반응에 주어진 평정척도에 응답하고 주제전문가가 평정한 각 행동반응 점수 간 차이의 절대값(Waugh, 2002) 또는 제공근을 점수로 사용하는 경우이다(Wagner, 1987).

이와 같이 상황판단검사는 다양한 채점방식이 존재하며 여러 연구들에서 채점방식에 따른 타당도와 신뢰도 비교연구가 진행되었지만(강민우 등, 2005; Knapp et al., 2001; Waugh, 2002; Weekly & Jones, 1999), 강민우 등(2005)은 반응대안 채점 방법에 따른 타당성을 비교한 결과 지식형은 반응대안 채점, 행동경향은 시나리오 채점 방식이 적합했으나 반응대안의 두 방법(절대값, 제공근)에 따른 차이는 보고하지 못하였다.

기존에 제시된 반응대안 방식의 한계점은 <그림 2>의

상황을 정확히 구분하지 못한다. 예를 들어 <그림 2>와 같이 전문가 집단의 지시문에 대한 선택지 평균 점수가 5점, 4점, 3점, 2점이었고, 학생A의 선택지 점수는 7점, 6점, 4점, 3점 학생 B는 3점, 2점, 2점, 3점이다. 이때 반응대안 2가지 방법(절대값, 제공근)에 따른 학생 A의 점수는 각각 6점= $|7-5|+|6-4|+|4-3|+|3-2|$ , 10점= $\{(7-5)\times 2\}+\{(6-4)\times 2\}+\{(4-3)\times 2\}+\{(3-2)\times 2\}$ 이고, 학생 B도 6점= $|3-5|+|2-4|+|2-3|+|3-2|$ , 10점= $\{(3-5)\times 2\}+\{(2-4)\times 2\}+\{(2-3)\times 2\}+\{(3-2)\times 2\}$ 으로 동일하다. 하지만, 전문가 집단의 선택지 순서를 보면 1번(5점)→2번(4점)→3번(3점)→4번(2점)의 순서로서 학생A(1번→2번→3번→4번)는 전문가 집단과 동일한 선택지 점수 순서를 보인 반면, 학생B(1번, 4번→2번, 3번)는 다른 순서를 보이고 있으므로 학생A가 학생B보다 좋은 평가를 받아야 하지만, 절대값과 제공근의 채점방식에서는 이를 구분하기가 힘들다. 따라서 본 연구에서는 기존 반응대안의 두 방법(절대값, 제공근)과 순서점수를 고려한 경우의 타당도를 비교하였다.

	학생A	전문가	학생B
선택지 1번	7점	5점	3점
선택지 2번	6점	4점	2점
선택지 3번	4점	3점	2점
선택지 4번	3점	2점	3점

그림 2. 채점방식 비교

## 연구방법\*

### 연구대상

청소년 운동선수 468명을 대상으로 채점방식에 따른

\* 상황판단검사의 채점방식에 따른 타당도를 비교하기 위해서 신명진(2014)의 연구에서 사용한 상황판단검사, 자아탄력성, 운동선수 인성, 정서지능, 지도자평점 점수를 재분석하였음.

타당도를 비교하였고, 연구 참여자의 일반적 특성은 <표 1>과 같다. 468명 참여자 모두에게 3가지 개념을 모두 측정할 경우 문항 수가 많아 측정의 피로도가 증가하여 정확한 설문조사에 어려움이 존재할 것으로 판단하였기 때문에 468명 중 자아탄력성 측정은 163명, 운동선수 인성 169명, 정서지능 136명을 대상으로 설문조사를 각각 실시하여 변별타당도를 검증하였다. 청소년 운동선수 468명 중 91명을 선정하여 수렴타당도를 검증하였고, 문항내적합치도는 468명 자료를 모두 적용하였다.

표 3. 참여자 일반적 특성

변인	구분	명	%
성별	남	349	74.6
	여	119	25.4
종목 유형	단체	231	49.4
	개인	237	50.6
주전 유무	주전	331	70.7
	비주전	137	29.3
학년	1학년	163	34.8
	2학년	161	34.4
	3학년	144	30.8
합계		468	100

## 측정도구

### 상황판단검사

상황판단검사는 신명진(2014)에 의해서 개발된 “청소년 운동선수 생활실천지능 상황판단검사”를 사용하였다. 청소년 운동선수 생활실천지능 상황판단검사는 단일 요인 13문항으로 구성되어 있으나 본 연구에서는 타당도 검증 초기 문항인 15문항을 바탕으로 타당도 차이를 살펴보았다. 그 이유는 신명진(2014) 연구의 채점기준이 절대값-순서점수를 적용하여 13문항으로 최종 선정하였으므로 이에 따른 편향을 최소화하기 위해서 초기 15문항을 모두 적용하였다.

### 자아탄력성

자아탄력성은 이준석(2013)이 개발한 운동선수 자아

탄력성 검사지를 사용하였다. 이는 5요인(능력, 동기, 사회적지지, 중재전략, 기대), 14문항 7점 척도로 구성되어 있다. 자아탄력성 질문지의 구성타당도를 살펴보기 위해서 163명을 대상으로 최대우도법으로 확인적 요인 분석을 실시하였다. 그 결과  $\chi^2$ 가 199.339, 자유도 67로 영가설을 기각하였고( $p=.000$ ), CFI=.794, TLI=.721, RMSEA=.110로 적합도 또한 만족스럽지 못하여 요인부하량이 .30이하인 7번, 13번 문항을 삭제하고 2차 확인적 요인분석을 실시하였다.  $\chi^2$ 가 99.763, 자유도 44로 영가설을 기각하였지만( $p=.000$ ), CFI=.887, TLI=.830, RMSEA=.088로 적합도는 만족스러웠으며  $\alpha$ 는 능력=.606, 동기=.743, 사회적지지=.585, 중재전략=.569, 기대=.682로 양호하였다.

### 운동선수 인성

운동선수 인성을 측정하기 위해서 권성호, 전현수, 이근철(2014)이 개발한 한국형 운동선수 인성 질문지를 사용하였다. 이 질문지는 경기상황, 일상생활로 구분되어 있으며 경기상황은 5요인(상대선수 부정적 심리 유발, 상대선수 존중, 동료선수 배려, 뒷사람과의 관계, 심판관정), 18문항, 일상생활은 4요인(예의 및 배려, 의리, 책임감, 사회적 규범), 15문항으로 구성되어 있다. 169명의 자료를 바탕으로 경기상황에 대한 확인적 요인 분석을 실시한 결과  $\chi^2$ 가 283.518, 자유도 125로 영가설을 기각하였고( $p=.000$ ), CFI=.801, TLI=.756, RMSEA=.087로 적합도 또한 만족스럽지 못하였다. 요인부하량이 .30이하를 보인 10번, 13번, 16번 삭제하고 2차 확인적 요인분석을 실시하였다.  $\chi^2$ 가 158.111, 자유도 80로 영가설을 기각하였지만( $p=.000$ ), CFI=.883, TLI=.846, RMSEA=.076로 적합도 지수는 양호하였다.  $\alpha$ 는 상대선수 부정적 심리 유발=.848, 상대선수 존중=.663, 동료선수 배려=.177, 뒷사람과의 관계=.797, 심판관정=.224로 동료선수 배려와 심판관정요인의  $\alpha$ 가 낮아 변별타당도 분석에서는 제외하였다. 일상생활 4요인에 대해서 확인적 요인분석을 실시한 결과  $\chi^2$ 가 131.159, 자유도 84로 영가설을 기각하였지만

( $p=.000$ ), CFI=.936, TLI=.920, RMSEA=.058로 적합도는 만족스러웠다. 요인 별  $\alpha$ 는 예의 및 배려=.798, 의리=.126, 책임감=.783, 사회적 규범=.719로 의리 요인의 문항내적합치도가 낮아 변별타당도 분석에서는 제외하였다.

### 정서지능

정서지능은 박치홍(2010)이 개발한 정서지능 검사지를 사용하고 이는 5요인(정서 인식 및 표현 능력, 감정 이입능력, 사고촉진능력, 정서활용능력, 정서조절능력), 27문항으로 구성되어 있다. 정서지능의 구성타당도를 살펴보기 위해서 136명의 자료를 바탕으로 확인적 요인 분석을 실시한 결과  $\chi^2$ 가 511.661, 자유도 314로 영가설을 기각하였고( $p=.000$ ), CFI=.791, TLI=.766, RMSEA=.068로 CFI, TLI 적합도가 양호하지 못하였다. 따라서 요인부하량이 .30이하인 9번, 17번 문항을 제거하고 2차 확인적 요인분석을 실시하였다.  $\chi^2$ 가 412.827, 자유도 265로 영가설을 기각하였지만( $p=.000$ ), CFI=.835, TLI=.813, RMSEA=.064로 적합도는 양호하였다.  $\alpha$ 는 정서 인식 및 표현 능력=.686, 감정이입능력=.641, 사고촉진능력=.795, 정서활용능력=.487, 정서조절능력=.517로 양호하였다.

### 자료분석

본 연구에서는 주제전문가 집단의 순서와 응답자 간의 순서의 일치도를 측정하는 순서점수를 공식 1)과 같이 산출하였다. 공식 1)에 의해서 <그림 2>에 나타난 학생 A의 순서점수는  $(5 \times 7/14) + (4 \times 6/14) + (3 \times 4/14) + (2 \times 3/14) = 5.5$ 점, 학생 B는  $(5 \times 3/14) + (4 \times 2/14) + (3 \times 2/14) + (2 \times 3/14) = 2.5$ 점이고, 순서점수가 높을수록 전문가 집단의 순서와 적은 차이를 보이고 있음을 의미한다. 절대값과 제곱근은 주제전문가 집단의 평점점수와 응답자 간 거리차이(거리점수)에서 순서점수를 뺀 새로운 채점방식을 본 연구에서는 추가하였다. 따라서 본 연구에서는 최근 상황판단검사에서 사용되는 반응대안 채점방식 중 ①절대값, ②제곱근 ③절대값-순서점수, ④

제곱근-순서점수 4가지 방식에 대한 변별타당도 비교를 실시하였다.

수렴타당도 검증을 위해서 서로 다른 팀 4명의 지도자에게 자신의 선수들이 대학팀 또는 실업팀으로 갔을 때 그 팀의 생활과 적응을 잘하는 정도를 2개 문항(“이 선수는 대학 또는 실업팀에 가서도 훈련을 성실히 할 것이다.”, “이 선수는 대학 또는 실업팀에 가서도 생활을 잘할 것이다.”)을 통해서 6점 기준으로 평정하도록 하였다. 그 이유는 운동선수 생활실천지능은 경기 외적인 상황에서 환경에 적응하는 능력이므로 선수 곁에서 오랫동안 지켜본 지도자가 선수 개인의 생활실천지능을 객관적으로 평가할 수 있기 때문에 지도자와 합숙생활을 하는 단체종목 중 설문응답을 허락한 지도자는 4명으로 총 91명의 선수를 대상으로 설문조사를 실시하였다. 따라서 실천지능은 개인 성공을 예측하는 변수이므로 미래상황의 훈련환경 및 일반생활에 대한 적응능력을 평가한 지도자 평점 점수와 운동선수 생활실천지능 상황판단검사 간의 상관성은 수렴타당도로 요인에 충족된다고 볼 수 있다.

$$\text{공식 1) OS} = \left( \frac{m_i}{\sum_{i=1}^n m_i} \cdot p_i + \frac{m_{i+1}}{\sum_{i=1}^n m_i} \cdot p_{i+1} \cdots \frac{m_{i+n}}{\sum_{i=1}^n m_i} \cdot p_{i+n} \right)$$

( $m$  = 전문가 집단 선택지 평균점수,  $n$  = 선택지 문항 번호,  $p$  = 응답자 선택지 점수)

## 결 과

<표 2>와 같이 4가지 채점방식에 따른 구성타당도를 살펴보면(AMOS 18.0, 최대우도법 적용), 4가지 방식 모두 통계적으로 구성타당도가 적합한 것으로 나타났다. 절대값 → 절대값-순서점수 → 제곱근 → 제곱근-순서점수 순으로  $\chi^2$ 값이 낮았으므로 절대값 채점방식이 468명 표본에서는 가장 적합하였다.

변별타당도 검증에서는 <표 3>과 같이 자아탄력성의 능력 요인에서 절대 값에서만 통계적으로 유의하였고, 운동선수 인성에서는 4가지 방식 모두 상대선수 존중, 예의 및 배려, 책임감 요인에서 정적 관련성이 존재하였다

표 2. 4가지 채점 방식에 따른 구성타당도

채점 방식	$\chi^2$	df	CFI	TLI	RMSEA
모형 1: 절대값	136.029***	90	.962	.956	.033
모형 2: 제공근	157.437***	90	.946	.937	.040
모형 3: 절대값-순서점수	144.366***	90	.938	.928	.036
모형 4: 제공근-순서점수	158.518***	90	.939	.929	.040

\*\* $p < .01$ , \*\*\* $p < .001$ 

표 3. 4가지 채점방식과 자아탄력성 간 판별타당도

	능력	중재전략	동기	기대	사회적지지
절대값	.168*	.055	.031	.081	.136
제공근	.128	.044	-.013	.047	.115
절대값-순서점수	.067	-.005	-.034	.040	.106
제공근-순서점수	.100	.028	-.031	.035	.107

\* $p < .05$ , \*\* $p < .01$ 

〈표 4〉, 〈표 5〉에서는 정서지능 5요인 중 사고촉진 능력에서 절대값과 제공근 방식에서 통계적 유의성이 나타났으므로 절대값-순서점수, 제공근-순서점수 방식에서 변별타당도가 높았다.

총 91명의 선수를 대상으로 수렴타당도를 분석한 결과 절대값  $- .305(p = .003)$ , 제공근  $- .316(p = .002)$ , 절대값-순서점수  $- .341(p = .001)$ , 제공근-순서점수  $- .332(p = .001)$ 로 나타났으므로 절대값-순서점수 채점방식의 수렴타당도가 가장 높았다.

## 논 의

CMB는 질문지 측정도구의 문제가 아니라 측정 시 심리적, 환경적 제약에 의해서 유발되는 오류이다. Podsakoff et al.(2003)은 CMB를 완벽하게 제거하는 일반화된 방법은 존재하지 않지만, 연구자들이 연구 진행 전 CMB를 해결할 수 있는 절차적/통계적 방법을 고려하는 자세가 중요하다고 하였다. Vealey (1986)는 스포츠자신감 질문지를 개발하면서 개발된 특성스포츠자신감 척도는 사회적 바람직성과 .23, 상태스포츠자신감 척도는 .21의 상관성이 존재한다고 언급하면서 이를 제거하기 위한 노력이 필요하다고 하였다. 사회적 바

람직성은 현재 사회의 기준이나 기준에 따라 자신의 모습을 긍정적으로 보이려는 경향(Zerbe & Paulhus, 1987)으로 CMB를 유발하는 대표적인 편향 중 하나이다. 최근 국내 체육학 영역에서 이루어진 신명진과 이근철(2014)의 연구에 따르면 사회적 바람직성이 유발하는 CMB의 규모는 6.6%~39%인 것으로 나타났으므로 이를 해결할 수 있는 방안을 간구해야 할 것이다.

CMB를 해결하는 절차적 방안 중 하나는 측정방법의 분리이다. 하지만 스포츠심리학에서 최근 10년간 이루어진 논문들을 종합해보면, 변수들 간의 관계를 살펴본 양적 연구 196편 중 방법분리를 적용한 논문은 24편으로 대화성적, 체력, 체격, BMI, 운동참여 횟수 및 빈도 등과 같이 매우 제한된 영역에서 사용하는 측정방법이었다. 심리적 변수를 리커트형 검사지를 이외의 방법으로 측정하여 방법분리가 이루어진 연구는 전무하였다. 이런 결과는 스포츠심리학에서 심리적 변수를 측정하는 개발된 검사지 대부분이 리커트 형이므로 연구자들이 측정방법분리를 통해 CMB를 고려할 수 없는 연구 환경이 고착화되고 있음을 의미한다. 따라서 본 연구에서 소개한 상환판단검사는 이런 한계를 극복할 수 있는 대안 중 하나일 것이다.

본 연구에서는 절대값, 제공근, 절대값-순서점수, 제공근-순서점수, 4가지 채점방식에 따른 타당도를 비교



표 4. 4가지 채점방식과 운동선수 인성 간 판별타당도

	부정적 심리유발	상대선수 존중	윗사람과의 관계	예의 및 배려	책임감	사회적 규범
절대값	-.120	.222**	-.045	.276**	.289**	-.045
제공근	-.133	.198**	-.012	.237**	.251**	-.064
절대값-순서점수	-.114	.166*	-.051	.172*	.181*	-.040
제공근-순서점수	-.132	.182*	-.014	.208**	.220**	-.064

\*p&lt;.05, \*\*p&lt;.01

표 5. 4가지 채점방식과 정서지능 간 판별타당도

	정서인식 및 표현능력	감정이입능력	사고촉진능력	정서활용능력	정서조절능력
절대값	.157	.131	.247**	.076	-.117
제공근	.147	.119	.231**	.058	-.157
절대값-순서점수	.105	.041	.146	-.024	-.150
제공근-순서점수	.133	.094	.203	.030	-.167

\*p&lt;.05, \*\*p&lt;.01

하였다. 구성타당도에서는 4가지 방법 모두 통계적으로 우수하였지만,  $\chi^2$ 수치에서는 절대값 채점방식이 가장 낮았다. 이와 달리 변별타당도에서 절대값 채점방식에서만 자아탄력성의 노력요인에서 관련성이 나타나 4가지 방법 중 변별타당도가 낮았다. 마지막으로 수렴타당도에서는 절대값-순서점수가 -.341로 가장 높게 나타났다. 수렴타당도가 (-)인 이유는 거리점수(절대값, 제공근)가 높을수록 실천지능 점수가 낮은 반면, 순서점수는 이와 상반된 의미를 지니고 있어 거리점수-순서점수는 점수가 높을수록 실천지능이 낮은 부적 관계를 지니고 있기 때문이다. 상황판단검사를 메타분석한 McDaniel et al.(2007)의 연구에서 준거타당도 .26이었으므로 본 연구의 .341은 기존 연구결과보다 높았다. 4가지 채점방식에 따른 3가지 타당도를 비교했을 때, 절대값-순서점수를 고려한 채점방식이 가장 적합하였고, 기존 연구들에서 제시된 거리점수(절대값, 제공근)만을 고려한 채점방식보다 본 연구에서 제기한 순서점수를 함께 고려한 방식이 타당도에 긍정적 영향을 미치는 것으로 해석할 수 있다.

일반적으로 다중회귀분석과 구조방정식에서 통계적 모형을 추정할 때는 제공근 방식인 최대우도법 또는 최소자승법을 사용한다. 본 연구에서는 제공근-순서점수보다 절대값-순서점수가 타당성이 높았으므로 일반 통

계의 추정방식과는 다른 결과가 도출되었다. 그 원인에 대해서는 명확하지 않지만, 상황판단검사의 채점방식에 절대값을 사용했을 경우에는 거리점수가 절대값 점수보다 그 수치가 급격히 증가하여 순서점수를 고려하는 비율이 부족하기 때문인 것으로 판단된다. 예컨대, <그림 1>에서 학생 A, B의 거리점수인 절대값은 6점, 제공근은 10점으로 동일한 반면, 순서점수는 학생 A=5.5점, 학생 B=2.5점이다. 따라서 절대값-순서점수는 학생 A=0.5점, 학생 B=3.5점인 반면, 제공근-순서점수는 학생 A와 B가 각각 4.5점, 7.5점으로 급격하게 증가한다. 이와 같이 제공근방식이 절대값 방식보다 점수가 증가하므로 순서점수를 고려하는 비율이 낮아 절대값-순서점수의 타당도가 더 우수하게 나타난 원인 중 하나일 것이다.

아직 스포츠심리학에서 상황판단검사 개발에 관한 연구들이 부족하고 하지만, 본 연구 내용과 결과를 바탕으로 향후 관련 연구를 위한 연구자들에게 도움이 될 수 있는 내용을 정리하면 다음과 같다. 첫째, 타당도와 신뢰도가 확보된 다양한 심리적 변수에 대한 상황판단검사가 개발되어야 할 것이다. 현재 리커트형 중심으로 이루어진 심리적 변수 측정방법에서 벗어나 다양한 측정방법을 개발하는 것은 CMB를 제거할 수 있는 방안 중 하나이기 때문이다. 둘째, 상황판단검사 개발 중 채점방식을

주제관련 전문가의 효과성 점수를 적용할 경우 절대값-순서점수 방식을 추천한다. 본 연구결과 절대값-순서점수 방식이 타당도가 우수하였고, 거리점수(절대값, 제공근)만을 고려한 방식보다 거리점수와 순서점수를 함께 적용한 채점방식이 타당도에 긍정적 영향을 미쳤기 때문이다. 셋째, 측정방법분리에 따른 CMB 존재를 증명하는 연구가 진행되어야 할 것이다. 이론적으로 CMB를 해결할 수 있는 방안이 측정방법분리 이지만 이에 대한 실증적 연구가 이루어지지 못하였다. 따라서 측정방법분리와 그렇지 않을 경우 변수들 간 관계가 통계적으로 변화되는 양상을 살펴봄으로써 방법분리가 CMB제거의 효과성에 어느 정도 영향을 미치는지를 파악할 수 있을 것이다. 넷째, SJT와 일반 리커트형 질문지 간의 CMB의 영향 정도를 비교하는 연구가 진행되어야 할 것이다. SJT가 스포츠영역에서 대중적으로 적용되기 위해서는 CMB의 영향이 기존 리커트형 질문지보다 우수하다는 실증적 연구가 증명되어야 한다.

## 참고문헌

- 강민우, 윤창영, 이순목 (2005). 지시문과 채점방식에 따른 상황판단검사의 타당도 비교. 한국심리학회지: 산업 및 조직, 18(3), 547-565.
- 권성호, 전현수, 이근철 (2014). 한국형 운동선수 인성의 개념구조 및 척도개발. 한국스포츠심리학회, 25(2), 115-128.
- 박치홍 (2010). 섬유매체를 활용한 집단미술치료가 중학생의 정서지능에 미치는 영향. 미간행 석사학위논문. 영남대학교 환경보건대학원.
- 백순근, 임현수 (2006). 중학교 교사용 실천지능검사의 개발 및 타당화. 교육평가연구, 19(1), 223-240.
- 신명진 (2014). 청소년 운동선수 생활실천지능 구성개념 탐색 및 상황판단검사를 적용한 척도개발. 미간행 박사학위논문. 서울대학교 사범대학원.
- 신명진, 이근철 (2014). 내적동기와 외적동기의 위계적모형(HMIEM)에서 사회적 바람직성의 영향력 검증. 한국스포츠심리학회지, 25(3), 89-100.
- 이준석 (2013). 운동선수 자아탄력성 척도 개발 및 타당화. 한국스포츠심리학회지, 24(3), 65-86.
- 정현선, 박동건 (2009). 지식형 상황판단검사의 채점용 답 결정과 채점방식에 따른 준거관련 타당도 비교 연구. 한국심리학회지: 산업 및 조직, 22(1), 53-70.
- Anderson, L. & Wilson, S. (1997). Critical incident technique. In D. L. Whetzel and G.R. Wheaton (eds.), *Applied Measurement Methods in Industrial Psychology*, Palo Alto, CA: Davies-Black.
- Chan, D. & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance*, 15, 233-254.
- Cianciolo, A. T., Grigorenko, E. L., Jarvin, L., Gil, G., Drebort, M. E. & Sternberg, R. J. (2006). Practical intelligence and tacit knowledge: Advancements in the measurement of developing expertise. *Learning and Individual Differences*, 16, 235-253.
- Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N. & Schmidt Harvey V. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology*, 86, 410-417.
- Cote, J. A. & Buckley, R. (1987). Estimating trait, method, and error variance: Generalizing across 70 construct validation studies. *Journal of marketing Research*, 24, 315-318.
- Dalessio, A. T. (1994). Predicting insurance agent turnover using a video-based situational judgment test. *Journal of Business and Psychology*, 9, 23-32.
- Guilford, J. P. & Lacey, J. I. (1947). *Printed Classification Tests*. AAF Aviation Psychology Research Program, Report No. 5.D.C.: U.S. Government Printing Office.
- Hedlund, J., Forsythe, G. B., Horvath, J. A., Williams, W. M., Snock, S. & Sternberg, R. J. (2003). Assessing practical intelligence in business school admissions: A supplement to the graduate management admissions test. *Learning and Individual Differences*, 16(2), 101-127
- Jones, C. & DeCotiis, T. A. (1986). Video-assisted selection of hospitality employees. *The Cornell H.R.A. Quarterly*, 67-73.
- Kline, T. J. B., Sulsky, L. M. & Rever Moriyama, S. D. (2000). Common methods variance and specification errors: a practical approach to detection. *The Journal of Psychology*, 134(4), 401-421.
- Knapp, D. J., Campbell, C. H., Borman, W. C., Pulakos, E. D. & Hanson, M. A. (2001). Performance assessment for a population of jobs. In J. P. Campbell & D. J. Knapp(Eds.), *Exploring the limits in personnel selection and classification*. Mahwah, N. J.: Lawrence Erlbaum

- Associates.
- Latham, G. P., Saari, L. M., Pursell, E. & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology*, 65, 422-427.
- McDaniel, M. A. & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment*, 9, 103- 113.
- McDaniel, M. A. & Whetzel, D. L., (2005). Situational judgment test research: Informing the debate on practical intelligence theory. *Intelligence*, 33, 515-525.
- McDaniel, M. A., Hartma, N. S., Whetzel, D. L. & Grubb III, W. L., (2007). Situational judgement tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, 60, 63-91.
- McFarland, L. A. & Ryan, A. M. (2000) Variance in faking across non-cognitive measures. *Journal of Applied Psychology*, 85, 812-821.
- Motowidlo, S. J., Dunnette, M. D. & Carter, G. W. (1990). An alternative selection procedure: the low-fidelity situation. *Journal of Applied Psychology*, 75(6), 640-647.
- Motowidlo, S.J., Hanson, M.A. & Crafts, J.L. (1997). Low-fidelity simulations. In D.L. Whetzel and G.R. Wheaton (eds.), *Applied Measurement Methods in Industrial Psychology*. Palo Alto, CA: Davies-Black.
- Nguyen, N. T., Biderman, M. D. & McDaniel, M. A. (2005). Effects of response instructions on faking in a situational judgment test. *International Journal of Selection and Assessment*, 13, 250-260.
- Podsakoff, P. M. & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospect. *Journal of Management*, 12, 69-82.
- Podsakoff, P. M., MacKenzie, S. B., Lee, Y. J. & Podsakoff, N. P. (2003). Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *Journal of Applied Psychology*, 88(5), 879-903.
- Sharma, S., Gangopadhyay, M, Austin, E. & Mandal M. K., (2013). Development and validation of a situational judgement test of emotional intelligence, *International Journal of Selection and Assessment*, 21(1), 57-73.
- Sharma, S., Yetton, P. & Crawford, J. (2009). Estimating the effect of common method variance: the method-method pair technique with an illustration from TAM research, *MIS Quarterly*, 33(3), 473-490.
- Thornton, G.C. & Cleveland, J.N. (1990). Developing managerial talent through simulation. *American Psychologist*, 45, 190-99.
- Vealey (1986). Conceptualization of sport-confidence and competitive orientation: preliminary investigation and instrument development. *Journal of Sport Psychology*, 8, 221-246.
- Wagner, R. K. (1987). Tacit knowledge in everyday intelligent behavior. *Journal of Personality and Social Psychology*, 52, 1236-1247.
- Waugh, G. (2002). *Selecting response options and items for a situational judgement test*. Paper presented as part of the following symposium - Understanding and Predicting Performance in Future Jobs. 17th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto.
- Weekley, J. A. & Jones, C. (1999). Further studies of situational tests. *Personnel Psychology*, 52, 679-700.
- Weekley, J. A. & Ployhart, R. E. (2006). *Situational judgment tests: Theory, management, and application*. Mahwah, NJ: Erlbaum.
- Williams, L. J., Gavin, M. B. & Williams M. L. (1996). Measurement and nomeasurement processes with negative affectivity and employee attitudes. *Journal of Applied Psychology*, 81(1), 88-101.
- Zerbe, W. J. & Paulhus, D. L., (1987). Socially desirable responding in organizational behavior: A reconception. *The Academy of Management Review*, 12(2), 250-264.

# The Important of Situational Judgement Test on The Sport Psychology Research and The Validities of Different Combinations of Scoring Alternatives

**Myoungjin Shin**

*Seoul National University*

Common Method Biases(CMB) is not the matter of measuring tools but the various errors caused by measurement. One of the procedural remedies to overcome these errors is the separation of measurement. However, according to the analysis of the papers published in sports psychology academic journals during the last ten years, the papers used the separation of measurement were just 24 papers of total 197 papers and most measurement were the questionnaires of Likert scale. So this research introduces the Situational Judgment Test(SJT) which can measure the psychological variables using different method except for the questionnaires of Likert scale and describes the developing process and the existing research results. In addition to this, on the basis of the Situational Judgment Test(SJT) developed in the area of sports, it suggests that the scoring key methods which is applied both the distance score and order score show better the validity compare with methods using only the distance score.

**key words:** Common Method Bias, Situational Judgement Test, Scoring Key