한국심리학회지: 산업 및 조직
Korean Journal of Industrial and Organizational Psychology
2005. Vol. 18, No. 3, 547-565

지시문과 채점방식에 따른 상황판단검사의 타당도 비교*

강 민 우† 윤 창 영 이 순 묵
GIIR HR팀 대구대학교 성균관대학교

본 연구는 직무관련 상황을 묻는 상황판단검사에서 지시문과 채점방식에 따라 타당도 즉, 구성개념타당도와 준거관련 타당도 그리고 증분타당도가 달라질 수 있다는 가설을 가지고 출발하였다. 결과는 지시문과 채점방식에 따라 구성개념에 있어 차이가 나타날 수 있는 가능성을 보여주었다. 또한 준거관련 타당도도 준거의 속성에 따라 즉, 과업수행에는 지식형 지시문이, 맥락수행에는 행동경향 지시문이 더 의미있는 결과를 나타내고 있다. 마지막으로 증분타당도 역시 각 준거에 따라 의미있는 값을 보여주고 있다. 마지막으로 각 지시문에 따라 결과의 차이가 명백하지 않은 이유를 논의에서 살펴보도록 한다.

주요어 : 상황판단검사, 지식형 지시문, 행동경향 지시문, 채점방식, 과업수행, 맥락수행

* 이 연구를 위해 2005년도 Brain Korea-21, 아동교육연구 프로젝트로부터 부분적 지원을 받았음.
† 교신저자 : 강민우, GIIR HR팀, analysis78@naver.com

최근 들어, 상황판단검사(situational judgment tests)에 대한 관심이 더욱 늘어가고 있다(예, Chan & Schmitt, 2002; Clevenger, Pereira, Wiechmann, Schmitt, & Harvey, 2001; McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001; McDaniel & Nguyen, 2001; Motowidlo, Dunnette, & Carter, 1990; Ployhart & Ehrhart, 2003). 저충실도 시뮬레이션(low-fidelity simulation)으로 알려진(Motowidlo et al., 1990) 상황판단검사는 직무 관련 상황을 지원자에게 제시하고 각 상황에 대해 그들이 어떻게 반응하는가를 묻는 상황검사의 일종이다(이순묵, 2004).

최근의 연구는 상위분석(meta-analysis)을 통한 상황판단검사의 준거관련 타당도(ρ)를 .34로 추정하였다(McDaniel et al., 2001). 뿐만 아니라 인지 능력검사보다 불리효과(adverse impact)가 적은 것으로 나타났으며(Pulakos & Schmitt, 1996; Motowidlo & Tippins, 1993), 동시에 비교적 높은 안면타당도를 가지고 있다(Motowidlo et al., 1990; Weekley & Jones, 1999; McDaniel et al., 2001). 그럼에도 불구하고 구성개념에 관해서는 덜 분명하였다(이상철, 이순묵, 조영일, 2003). 상위분석을 통한 상황판단검사의 구성개념은 주로 성격 5요인 검사(Big 5)에서의 하위 차원인 성실성(conscientiousness), 정서 안정성(emotional stability), 원만성(agreeableness)(McDaniel, Hartman, & Grubb, 2003; McDaniel & Nguyen, 2001)과, 인지 능력의 국면(McDaniel et al., 2001; McDaniel et al., 2003)을 측정하는 것으로 알려져 있다. 이와 더불어 몇몇 연구자들은 상황판단검사가 종업원 면접이나 직무 지식 검사처럼 다양한 구성개념을 측정하는 측정도구라고 주장하거나(Chan & Schmitt, 1997; McDaniel et al., 2001; Weekley & Jones, 1999), 상황판단검사의 구성개념을 여전히 미해결된 문제라고 주장하였다(Weekley & Jones, 1997; Chan & Schmitt, 2002). 이와 같이 구성개념이 확립되지 않은 상황에서 최근 McDaniel 등(2003)은 상황판단검사에서의 지시문에 따라 구성개념이 다르게 측정됨을 보였다. 즉, 지시문의 두 가지 형태인, 행동경향 지시문과 지식형 지시문에 따라 상황판단검사 점수는 인지능력, 성격 등 여러 변수들과 상이한 관련성을 보였다.

이 연구에서는 McDaniel 등(2003)의 연구의 반복 및 확대를 목표로 한다. 즉, 지시문의 두 가지 형태인, 행동경향 지시문과 지식형 지시문에 따라 구성개념, 준거관련타당도 그리고 증분타당도가 달라질 수 있음을 보이고자 한다. 가장 널리 쓰이는 지시문들 중 행동경향 지시문은 "할 것 같은 행동/하지 않을 것 같은 행동(most/least likely)"을 묻는 것으로서 실제로 피험자 혹은 지원자로 하여금 어떻게 행동할 것인지를 묻고 있기 때문에 의도와 행동의 일관성 즉, 행동경향(behavior tendency)과 관련이 크다(McDaniel & Nguyen, 2001; Motowidlo et al., 1990). 반면에 다른 방식인, 최선/최악(best/worst)을 선택하도록 하는 지시문은 어떤 것이 최선인지 최악인지를 구분하는 적절성에 대한 판단 및 지식 즉, 인지적 능력을 요구함으로 인지능력 국면(aspects)과 관련되어 있다는 주장이 있다(McDaniel & Nguyen, 2001). 그러므로 지시문에 따라서 수집된 결과를 가지고 인지와 성격 변인과의 상관을 통해 구성개념이 다를 수 있다는 것을 반복검증하고, 나아가 지시문과 채점방식에 따라 차별적인 준거관련 타당도와 증분타당도를 알아봄으로써 지시문과 채점방식의 조합을 비교하고자 한다.

지시문 효과와 유형

상황판단검사에 주로 사용되는 지시문의 유형은 크게 직무상황에서 어떠한 행동 대안이 가장 적절한가의 최선/최악을 묻는 지식형(knowledge) 지시문과 그러한 상황 속에서 지원자나 검사 수행자가 어떠한 행동을 할 것인가를 묻는 행동경향(behavioral tendency) 지시문으로 나눌 수 있다.

최근의 연구들은 이와 같이 다양한 지시문의 형태에 따라 상황판단검사의 구성개념과 타당도가 달라질 수 있다고 설명하고 있다. McDaniel과 Nguyen(2001)는 최선/최악(best/worst) 방식, 즉 지식형의 지시문이 응답자들에게 가장 좋은 대안을 선택하거나 각 응답에 효율성을 평정하도록 지시하기 때문에 행동경향 지시문에 비해 위장반응의 문제가 덜 발생할 것으로 가설화하였다. 또한 Ployhart와 Ehrhart(2003)는 행동경향을 나타내는 "가장 할 것 같은 혹은 하지 않을 것 같은(would do)" 지시문들이 "어떤 것을 해야 하는가(should do)"의 지시문들에 비해 범주 내 지시문간에 더 높은 상호상관을 보이고 있음을(.76 vs .36) 제시하였다. 또한 두 지시문간의 상관을 보면 .22로 두 범주가 서로 다른 구성개념을 보이고 있음을 알 수 있다. 준거관련 타당도는 행동경향인 'would do' 지시문들이 지식형인 'should do'의 지시문들에 비하여 높은 것으로 나타났다.

McDaniel 등(2003)은 상위분석(meat-analysis)을 이용하여, 반응에 대한 지시문이 측정되는 구성개념에 영향을 미친다는 것을 보여줬다. 즉, 인지능력과의 관련을 보면 지식형 지시문이 행동경향 지시문보다 높은 상관을 보이고(.43 vs .23), 성격개념과의 관련을 보면 행동경향 지시문이 지식형 지시문에 비해 몇 개의 성격에서 높은 상관을 보이는 것으로 나타났다: 성실성(.51 vs .33), 원만성(.53 vs .20), 정서 안정성(.51 vs .11). 끝으로 준거와의 관련을 보면 지식형 지시문이 행동경향 지시문에 비해 높은 타당도를 보인다고 보고하였다.

그렇다면 지시문에 따라 구성개념은 물론이고 준거관련 타당도 역시 달라질 수 있을 것

표 1. 상황판단검사 응답 지시문의 비교

응답 지시문 형식	내 용	관련 연구들
Ⅰ. 행동경향 지시문		
가장 할 것 / 하지 않을 것 (Most/Least likely)	업무 상황에서 나타난 행동들에 대해 각 대안 중 가장 할 것 같은 대안과 가장 하지 않을 것 같은 대안을 모두 고르게 하는 방식.	· Motowidlo, Dunnett, & Carter(1990) · Motowidlo & Tippins(1993) · Pulakos & Schmitt(1996) · Smith & McDaniel(1998)
Ⅱ. 지식형 지시문		
최선의 반응(Best)	하나의 최선의 답을 묻는 방식.	· Reynolds, Winter, & Scott(1999) · Weekley & Jones(1997)
최선 / 최악의 반응 (Best/Worst)	각 시나리오에서 가장 최선이라고 생각되는 반응과 최악이라고 생각되는 반응을 모두 고르게 하는 방식.	· Weekley & Jones(1999)

이고, 나아가서 각 지시문이 증분적으로 기여하는 타당도 역시 다를 것이다.

채점방식의 유형과 영향

지시문과 더불어 점수화 방법에 따라 타당도와 신뢰도의 값이 변할 수 있다(Knapp, Campbell, Borman, Pulakos, & Hanson, 2001; Waugh, 2002). 두가지 채점방식이 표 2에 제시되어 있다.

먼저, 전형적으로 가장 많이 사용되는 점수화의 방법은 시나리오 당 점수가 부여되는 시나리오 채점방식이다. 우선, 전문가와 지원자에게 시나리오와 행동대안을 제시하고 가장 할 것 같은 행동(Most likely do)과 가장 하지 않을 것 같은 행동(Least likely do)을 선택하도록 한다. 그런 다음, 전문가에 의해 선택된 혹은 합의된 가장 할 것 같은 행동을 지원자가 가장 할 것 같은 행동으로 선택하면 1점을, 다른 행동 대안을 선택하면 0점을, 전문가가 가장 하지 않을 것 같은 행동 대안에 대해 지원자가 가장 할 것 같은 행동대안으로 선택하면 -1점을 받게 된다. 같은 방식으로 가장 하지 않을 것 같은 행동대안의 선택에도 적용된다(Motowidlo, et. al., 1990).

시나리오 채점 방식의 두 번째는 첫 번째 방식과 달리, 각 대안별로 주제전문가로부터 평균평정을 얻어 이 평균평정을 각 대안의 점수로 설정해 놓는다. 그런 다음, 지원자가 최선이라고 고른 대안과, 최악이라고 고른 대안에 부여된 주제전문가의 평정값들의 차이값을 지원자 점수로 하는 방식이다. 이 방법은 지원자들이 첫 번째 방법에서 얻을 수 있는 점수범위(-2～+2)보다 더 큰 점수범위를 가지게 된다. 그러므로 지원자간의 차이를 최대로 할 수 있는 장점이 있다(Knapp et al., 2001).

위에서 설명한 시나리오 채점방식이 산업심리학 연구에서 주로 사용되는 지시문에 결합된 방식이라면, Wagner와 Sternberg(1985)의 연구와 McDaniel 등(2004)의 최근 상황판단검사에서 사용되고 있는 방식은 반응대안 채점방식이다. 이 방식에서 응답자는 각 행동대안에 주어진 평정척도에 응답하도록 되어 있다. 이 방식에서도 지시문은 지식형일 수도 있고 행동경향일 수도 있다. 즉, 응답자가 보기에 적

표 2. 채점 방식의 비교

채점방식	점 수 화
시나리오 채점	만약 가장 할 것 같은 행동이 사전에 전문가에 의해 합의된 가장 할 것 같은 행동과 일치하면 1점, 가장 할 것 같은 행동이 전문가에 의해 합의된 가장 하지 않을 것 같은 행동과 일치하면 -1점. 총 점수의 분포는 -2～2점
	각 대안에 대해 전문가들의 평균평정을 얻고, 지원자에게 최선의 대안과 최악의 대안을 선택하도록 한 다음, 두 대안에 부여된 전문가의 평균평정을 빼서 그 차이를 점수로 함.
반응대안 채점	전문가와 지원자의 평정값 차이의 절대값
	전문가와 지원자 평정값 편차의 제곱(deviation-squared)이나 차이제곱평균의 제곱근(root mean squared deviation)

절한 정도(지식형) 또는 할 것 같은 정도(행동경향)의 관점에서 평정척도상(예를 들면, 7점 척도상)에 응답을 요구하게 된다. 채점방식은 전문가들이 응답한 평정을 해답으로 정해놓고, 지원자(혹은 초심자)의 응답한 평정과의 유사성이나 거리에 따라 응답자의 반응을 채점한다. 이 때의 채점을 보다 구체적으로 보면, 행동대안별로 전문가 응답평균과 응답자의 평정간 차이의 절대값(Waugh, 2002) 혹은 개별전문가들과 응답자간 반응차이들에 대한 차이제곱평균의 제곱근(root mean squared deviation)을 점수로 쓰기도 한다(Wagner, 1987). 이 방식은 지원자와 전문가 집단간의 의미있는 비교를 허용한다는 장점을 가지고 있다.

이와 같이 다양한 채점 즉, 점수화 방식에 따라 상황판단검사의 신뢰도와 타당도는 다른 결과를 보여주고 있다.

Knapp 등(2001)은 5가지 방식의 점수화 방법을 비교하였다. 요약하자면, 전문가들(experts)이 각 대안에 평정을 하여 대안마다 평균평정값이 부여된 상황에서 피험자가 가장 효과적(most effective)이라고 뽑은 대안의 평균평정과 가장 비효과적(least effective)이라고 뽑은 대안의 평균평정간의 차이를 점수화 하는 것이 다른 방식에 비해 좀 더 정보적이기 때문에 가장 큰 신뢰도를 얻었다고 설명하였다. 즉, 가장 효율적인 것과 가장 비효율적인 것을 모두 고르게 하는 것이 좀 더 정보적이라는 것이다.

Waugh(2002)는 미 육군의 미래의 수행에 대한 예측을 위해 지식형 상황판단검사를 실시하였다. 이 연구에서 6가지 점수화 방식이 비교되어 사용되었는데 첫 번째는 시나리오 채점방식에서 최선의 대안만을 선택하게 하여 이때 1점을 주는 방식으로, 두 번째는 최선과 최악의 대안을 모두 제대로 선택하면 1점을 주는 방식으로, 세 번째는 Weekly와 Jones (1999)에서 사용된 것과 같이 최선과 최악의 선택 각각에 1점씩을 부과하여 -2에서 2점의 점수 범위를 갖는 방식으로, 네 번째는 각 행동대안에 대해 각각 평정하도록 하여 전문가의 반응평정(해답)과 지원자의 반응평정의 차이의 절대값을 점수로 하는 방식을, 다섯 번째는 네 번째 방식으로 하되, 전문가에 의해 최선과 최악의 대안으로 정해진 대안에 대해서만 점수 계산을 하였다. 마지막으로 여섯 번째는 각 대안에 평균평정을 사전에 정하고, 가장 효과적인 대안과 가장 비효과적인 대안간을 지원자로 선택하게 한 다음 최선의 대안과 최악의 대안에 부여된 평정값 차이를 점수로 하였다. 결과는 1, 2, 3번째 점수화 방식간에 높은 상관을, 4번째와 5번째 점수화 방식간에 높은 상관을 보여주고 있었다. 6번째 점수화 방식은 모든 방식과 어느 정도의 상관을 보여주며 준거관련 타당도가 가장 크다고 보고하였다. 이 결과에 대해 Waugh는 점수화의 방식이 동등(equivalent)하지 않으므로, 점수화의 방식에 따라 신뢰도와 타당도에 영향을 준다고 설명하였다.

연구가설 및 연구문제

연구가설

본 연구에서는 상황판단검사의 구성개념이 모호한 것의 원인을 지시문의 영향이라고 본다. 즉 행동경향지시문은 행동경향성을 측정하는 것이고, 지식형 지시문은 상황에 관련된 실무적 지식 혹은 인지적 국면(aspects)과 관련이 있다는 주장(McDaniel & Nguyen, 2001;

McDaniel et al., 2003; Nguyen & McDaniel, 2003; Ployhart & Ehrhart, 2003)이 옳다면, 두 형태의 지시문은 재고자 하는 속성, 즉 구성 개념이 다를 수밖에 없는 것이다. 이 두 가지 다른 지시문이 그 목적에 관계없이 여러 연구에서 사용되어 왔다. 뿐만 아니라 어떠한 지시문이 더 효율적인가에 대해서 일관된 결과가 나타나지 않고 있다. 그러므로 본 연구에서는 상황판단검사의 지시문에 따라서 다른 특성을 나타내는지를 간접적으로 재확인하고자, 상이한 지시문이 사용된 상황판단검사 점수가 인지능력 및 성격과 어떤 관계를 가지는지를 살펴보고자 한다. 그런 다음 지시문에 따른 준거관련타당도와 증분타당도를 비교해 보고자 한다.

먼저, 지식형 지시문은 어떠한 행동 대안이 가장 적절한가에 대한 직무지식(Motowidlo et al., 1997) 혹은 인지능력의 국면(aspects)(McDaniel et al., 2001)과 관련되어 있기 때문에 인지능력과 높은 상관을 나타낸다는 것과 행동 경향 지시문이 행동일관성 원리(Motowidle et al., 1990)를 적용하고, 행동의도(Latham, 1989)를 측정한다는 가정 아래, 성격검사와 관련이 높다는 것을 재확인한다. 그러므로 다음과 같은 가설을 제안한다.

가설 1. 지식형 지시문의 상황판단검사는 행동경향성 지시문의 상황판단검사에 비해 인지능력과 상관이 높을 것이다.

가설 2. 행동경향 지시문의 상황판단검사는 지식형 지시문의 상황판단검사에 비해 성격변인과 상관이 높을 것이다.

다음으로 이러한 차이를 바탕으로 준거관련 타당도 역시 다를 것으로 기대된다. 그러나 두 연구에서 서로 상반되는 연구결과를 보고하고 있다. Ployhart와 Ehrhart(2003)가 대학생을 대상으로 지시문의 효과를 알아본 연구에서는 행동경향 지시문이 지식형 지시문에 비해 높은 준거관련 타당도를 보이고 있다고 주장한다. 반면에 McDaniel 등(2003)의 메타분석에서는 인지능력과의 높은 상관과 위장반응의 문제가 덜 발생한다는 이유로 지식형 지시문이 행동경향 지시문에 비해 높은 준거관련 타당도를 보이고 있다고 주장한다. 이러한 상반된 결과는 각 지시문이 사용된 검사를 어떠한 준거와 관련시켰는지를 분명히하지 않은데서 오는 결과일 수 있다. 즉, 준거가 전문과업수행이라면 지식수준이 중요 예측변수가 되고, 맥락수행이라면 행동경향이 중요한 예측변수가 될 것인데 연구자들은 이 점을 간과하였다. 즉, 지시문의 형태와 준거의 성격에 따라 상이한 타당도를 가질 것이다.

기존의 연구들에서 수행의 준거를 과업수행(task performance)에만 초점을 맞추었다면, 본 연구에서는 준거의 영역을 확대하여 과업수행과 맥락수행(contextual performance)으로 나누어 살펴보도록 한다. 여러 연구들(Borman & Motowidlo, 1993; Motowidlo, Borman & Schmit, 1997; Motowidlo & Van Scotter, 1994)에서 수행의 차원을 과업수행(task performance)과 맥락수행(contextual performance)으로 나누고 있다. 과업수행을 정의하자면, 조직의 핵심기능(technical core)에 직접적으로 기술적인(technical) 과정을 이행하거나, 간접적으로 재화나 용역을 제공함으로써 기여하는 활동들인 반면에, 맥락수행은 과업수행의 범주 안에는 속하지 않으나 여전히 조직의 효율성을 위해 중요한 활동들을 일컫는다(Borman & Motowidlo, 1993). 예를 들면, 조직시민행동, 친조직성향, 친사회성과

같은 행동들이다.

따라서 각 지시문의 형태(지식형 vs 행동경향)와 준거의 성격(과업수행 vs 맥락수행)에 따라 상황판단검사는 상이한 타당도를 가질 것이다. 준거가 과업수행에 대한 것이면 지식형 지시문이, 맥락수행에 관한 것이면 행동경향 지시문이 더 관련이 높을 것이다. 그러므로 다음과 같은 가설들이 설정된다.

가설 3. 상이한 지시문에 의한 상황판단검사는 준거의 성격에 따라 상이한 타당도를 가질 것이다.

가설 3a. 행동경향 지시문의 상황판단검사는 맥락수행 준거에 대해서 지식형 지시문의 상황판단검사에 비해 높은 타당도를 가질 것이다.

가설 3b. 지식형 지시문의 상황판단검사는 과업수행 준거에 대해서 행동경향 지시문의 상황판단검사에 비해 높은 타당도를 가질 것이다.

연구문제

가설이 제시될 수준은 아니지만 지시문에 따라 증분타당도가 어떻게 나타나는가의 연구문제를 살펴보고자 한다. 상황판단검사에서의 증분타당도는 비교적 일관되게 나타나고 있다(Clevenger et al., 2001; Chan & Schmitt, 2002; McDaniel et al., 2004; Weekly & Jones, 1997, 1999). 본 연구에서는 두 형태의 지시문에 따라 증분타당도의 유의도 그리고 추가적인 설명량이 어떠한지 비교해 보고자 한다. 즉 '각 준거에 대해서 지시문에 따른 증분타당도가 유의한가'와 '준거에 따라 지시문에 따른 추가적인 설명량(ΔR^2)이 얼마나 상이한가'를 알아보고자 한다.

연구방법

피험자

본 연구에서의 피험자는 서울 소재 3개 대학과 지방 소재 2개 대학에서 2005학년도 1학기 기준 심리학개론을 수강하는 1, 2, 3학년에 재학 중인 대학생 427명을 그 대상으로 하였다. 또한 피험자를 다양한 5개 대학에서 얻으므로 인지능력에 대한 범위제한(range restriction)의 효과를 줄였다. 대학입시가 주로 인지능력에 의해서 선발이 되므로, 일부학교만 할 경우 인지능력의 측정 범위가 제한될 수 있기 때문이다.

절차

본 연구는 지시문만 다른 동일한 상황판단검사의 실시에 따른 순서효과와 이월효과를 감소시키기 위해 일정 기간을 두고 두 차례로 실시하였다. 검사 시행 때는 아래와 같이 인지능력검사, 행동경향 상황판단검사, 성격검사와 맥락수행에 대한 자기평정, 그리고 지식형 상황판단검사를 학교별로 순서효과가 통제되도록 순서를 바꾸어 실시하였다. 두 시행 중간에는 퀴즈를 실시하여 과업수행 준거로 사용될 성적을 구하였다.

측정도구

상황판단검사

상황판단검사의 개발 절차는 Motowidlo 등

(1990)의 3단계 절차를 따른다. 첫 번째 단계는, 중요사건기법(critical incident technique) (Flanagan, 1954)을 이용하여 중요사건을 수집하는 것이다. 이를 위해 최송화(2000)의 연구에서 대학생 묵시지 검사의 개발을 위해 수집된, '대학생이 학업을 우수하게 수행하는데 필요한 비공식적인 지식' 문항들 91개와, 저자가 실시한 대학원 신입생 10명에 대한 면접과 다양한 지역의 30여명의 대학생들에 대한 설문을 통해 성공적인 대학생활을 위한 중요사건 50여개를 수집하였다.

두 번째 단계에서는, 시나리오를 제작하고 그에 대한 반응대안을 구하기 위해 3개 지역(서울·경기, 충청도, 경상도) 19개 대학의 40여명의 대학생들에게 보여주고, 각 시나리오에 대해 '당신이라면 이 상황에서 어떻게 행동할 것인가'를 물었다. 이를 통해 30개의 시나리오에 각각 5개의 반응대안들을 개발할 수 있었다.

세 번째로 정답(key)을 개발하기 위해 30개의 시나리오로 이루어진 상황판단검사를 가지고 성적이 우수한(4.5 만점에 4.0 이상) 4학년 학생 10명과 대학원 신입생 10명을 대상으로, 각 상황에 주어진 반응대안들이 얼마나 잘 대처하는 행동인지에 대해 대안별로 평정값을 받았다. 대안별로 그 평정치를 평균하여 각 시나리오내에서 대안들간 점수의 폭(range)이 3 미만이면 대안들간 변별이 어려운 것으로 보고 시나리오를 제외하기로 하였다. 이렇게 해서 시나리오 14개가 제외되었다. 또한 하나의 대안에 대해서 얻은 평정치간의 표준편차가 2 이상인(전문가의 합의가 이루어지지 않아 정답을 만들기 어려운 문항) 1개를 제거하고 최종적으로 15개의 시나리오에 대해서만 정답을 개발하였다.

지능검사

인지능력에 대한 측정은 이상철(2002)의 연구에서 사용되었던 것과 같이 한국교육개발원(KEDI) 집단지능검사 대학생 A형(박경숙, 현주, 이재분, 오승민, 1991) 문항 중 언어영역과 수리영역, 그리고 공간영역을 대상으로 홀수 문항만을 사용하였다. 총 58문항 중 검사집 구성시 언어영역에서 두 문항이 제외되어 총 56문항만을 사용하였다. 지능검사의 내적 일관성 신뢰도계수(α)는 한국교육개발원 집단지능검사 대학생 A형에서는 언어검사, 수검사, 공간검사에 대하여 각각 0.77, 0.85, 0.67이었으며 이상철(2002)의 연구에서는 0.69, 0.69, 0.58을 나타냈다. 본 연구에서는 0.74, 0.90, 0.51로 나타났다.

성격검사

성격의 측정은 성격 5요인 검사인 Goldberg와 Saucier(1996)의 IPIP(International Personality Item Pool)로 측정하였다.[1] 이 척도는 신경증, 외향성, 경험에 대한 개방성, 원만성, 성실성의 5요인으로 이루어져 있으며 각 요인별 10문항씩 총 50문항으로 구성되어 있다. 각 문항에 대한 응답은 5점 척도상에서('전혀 그렇지 않다'(1점) - '매우 그렇다'(5점)) 평정하도록 되어 있다. Goldberg(1999)의 타당도 연구에 따르면 NEO-PI-R와는 .73(교정 후 .94), 16PF (Conn & Rieke, 1994) 와는 .66(교정 후 .86)의 공시(concurrent) 타당도를 보여주고 있다. IPIP의 각 요인별 내적 일관성 신뢰도(α)는 각각 .87, .82, .79, .84, .86으로 보고되었다. 국내연구에서는 신현균과 장재윤(2003)의 연구에서 각각 .85, .75, .77, .75, .85로 보고되었다. 본

1) 이 척도에 대한 정보는 http:ipip.ori.org/ipip에서 얻을 수 있다.

연구에서는 검사 구성에서 신경증의 한 문항이 제외되어, 신경증에는 9문항을 나머지 요인들에는 각 10문항을 사용하였으며 소검사의 신뢰도는 각각 .86, .76, .79, .80, .77로 나타났다.

퀴즈(과업수행 준거)

본 연구에서는 준거 측정치를 통일하기 위해서 5개 학교 심리학개론 강의에서 내용과 형태가 동일한 시험지로 평가를 실시하여 그 성적을 과업수행 준거로 사용하였다. 이를 위해 사전에 강의 내용을 3개의 장으로 조정하였고, 각 학교의 강사 선생님들의 협조를 얻어 문항 검토 및 수정 보완의 절차를 통해 타당도를 확보하고 난이도 수준을 조절하여 최종적으로 4지선다형 24개 문항을 만들어 5개 학교의 퀴즈(quiz)로 사용하였다. 퀴즈의 내적 일관성 신뢰도(α)는 .81이었다.

맥락수행 측정(맥락수행 준거)

본 연구에서는 학교 장면에서의 맥락수행의 측정을 위해 먼저 맥락수행을 정의한다. 그런 다음 학교 장면에서의 맥락수행의 정의를 내린 후, 이를 측정하기 위한 문항을 개발하는 과정을 진행하였다. 수행을 과업수행과 맥락수행으로 나눌 때, 과업수행은 조직의 핵심기능(technical core)에 대하여 직접적으로 기술적인 과정을 이행하거나, 간접적으로 재화나 용역을 제공함으로써 기여하는 활동들이라고 정의되는 반면, 맥락수행은 과업수행의 범주 안에는 속하지 않으나 여전히 조직의 효율성을 위해 중요한 활동들을 일컫는다(Borman & Motowidlo, 1993).

맥락수행을 측정하기 위한 도구로는, 맥락수행에서 조직시민행동을 중요 개념으로 보고(Motowidlo 2000), Smith, Organ, 그리고 Near (1983)에서 조직시민행동(OCB) 측정을 위해 사용한 16문항을 사용하도록 한다. 이를 위해 먼저 척도를 학교 상황에 적합하도록 번역하였다. 그런 다음, 그 번역이 학교 상황과 원래 의미에 충실하였는가는 산업 및 조직심리학 전공 대학원생 5명의 합의를 통해 그 내용적 타당도를 확보하였다. 또한 앞서 언급한 바와 같이 맥락수행 중 '임무를 성공적으로 완수하기 위해 노력을 지속적으로 하는 것'을 제외하였다. 이와 같은 과정을 통해 최종적으로 16문항 중 9문항만을 척도로 사용하도록 하였다.

이 문항들이 하나의 구성개념을 지향하는지를 확인하기 위해 요인분석을 실시하였다. 이를 위해 SAS 8.01의 PROC FACTOR를 사용하였고 공통요인모형에 단일주축분해를 사용하였으며, 사전공통분추정치는 다중상관제곱치(SMC)를 사용하였다. 고유치에 대한 검토를 할 때, 1번 요인 이후에 고유치가 현저하게 감소하며 누적분산비율이 1요인에서만도 86%가 되므로 하나의 요인으로 묶여지는 것을 알 수 있다. 문항들을 대학생들에게 자기 보고식으로 5점 척도상에서 평정하도록 한 결과였다. 이 때 내적 일관성 신뢰도(α)는 .78로 나타났다.

분석

먼저 상황판단검사의 구성개념 탐색을 위해서 의존성있는 상관계수간 비교방법이 사용되었다(Steiger, 1980). 예로서, "성격"이라는 하나의 개념이 두 종류의 상황판단검사 중 어느 지시문에 따른 검사와의 상관이 더 큰지를 비교해야 하는데, 각 상관계수에는 "성격"이 공통으로 들어 있어서 의존성있는 상관계수간 비교가 된다. 다음으로 준거관련 타당도와 증

분타당도는 SPSS WIN 12.0판이 사용되었다. 준거관련 타당도를 얻기 위해서 상관분석을 실시하였고, 각 검사의 신뢰도는 내적일관성 신뢰도 α를 구하였다. 마지막으로 증분타당도를 얻기 위해서 다중회귀분석을 실시하였다.

연구결과

상황판단검사의 구성개념 탐색

먼저, 행동경향지시문은 행동경향성을 측정하는 것이고, 지식형 지시문은 판단과 관련된 인지능력과 관련되어 있다는 주장(McDaniel & Nguyen, 2001; McDaniel et al., 2003; Nguyen & McDaniel, 2003; Ployhart & Ehrhart, 2003)을 간접적으로 확인하기 위해 영차상관(zero-order correlation)을 살펴보도록 한다. 즉, 행동경향지시문의 상황판단검사는 성격변인들(Big 5)의 측정치와 상관이 높을 것이고 지식형 지시문은 인지능력 측정치와 상관이 높을 것이라는 가설 1과 가설 2를 영차상관으로 살펴보았다. 먼저, 본 연구에서는 기존 연구들에서 많이 사용되었던 시나리오 채점방식으로 상황판단검사와 다른 변수들간의 관계를 살펴본 후 반응대안 채점에 대해서도 영차상관을 살펴보도록 한다. 여기에서는 특별히 성격이나 인지능력이라는 하나의 개념이 두 종류의 상황판단검사 중 어느 지시문에 따른 검사와의 상관이 더 큰지를 비교해야 하기 때문에, 상관계수간에 독립이 아닌 의존성(dependency)이 발생한다. 따라서 그 점을 감안하고 상관계수간 차이에 대한 유의도 검증이 실시되었다(Steiger, 1980의 Z 통계량 사용).

표 3에서 보는 바와 같이 먼저 인지능력과의 상관을 보면, 지식형 지시문의 상황판단검사와 행동경향 지시문의 상황판단검사 둘 다 유의하긴 하였지만 지식형 방식일 때 행동경향 방식일 때에 비해 상관값이 통계적으로 유의하게 크지는 않았다. 그러므로 가설 1인 '지

표 3. 상황판단검사, 성격검사, 인지능력 검사간의 영차상관(시나리오 채점)

	평균	표준편차	행동경향 SJT	지식형 SJT	외향성	원만성	성실성	신경증	개방성	인지능력
행동형 SJT	17.18	5.62	.75	.59	.10	.25	.40	.05	.07	.13
지식형 SJT	20.01	5.55	.45**	.80	.04	.16	.17	.03	.01	.17
외향성	3.2	.72	.08	.03	.85	.44	.15	.13	.22	.00
원만성	3.6	.53	.19**	.13**	.36**	.76	.32	.25	.17	.02
성실성	3.1	.62	.31**	.13**	.12**	.25**	.79	.18	.17	.11
신경증	2.9	.73	-.043	-.02	-.10*	-.19**	-.14**	.80	.16	.03
개방성	3.2	.55	.057	.00	.18**	.13**	.13**	.13**	.77	.22
인지능력	24.5	7.3	.10*	.14**	.00	-.01	-.09*	-.02	.17**	.82

주. ** 0.01수준, * 0.05수준.
대각선 아래는 상관, 대각선은 내적일관성신뢰도(α), 대각선 위는 비신뢰도에 대하여 교정된 상관값

식형 지시문의 상황판단검사는 행동경향성 지시문의 상황판단검사에 비해 인지능력과 상관이 높을 것이다'는 지지되지 못하였다.

반면, 성격요인들과의 관련성을 살펴보면, 행동경향 지시문의 상황판단검사와 지식형 지시문의 상황판단검사가 성격5요인의 변인들 중 성실성과 원만성에서 유의한 상관을 보이고 있다. 하지만 성실성 차원에서만 행동경향 상황판단검사가 지식형 상황판단검사에 비해 통계적으로 유의하게 큰 상관을 보여주었다(.31 vs .13, $p<.01$).

이상을 요약하자면, 시나리오 채점방식의 두 지시문의 상황판단검사에서는 행동경향 상황판단검사가 성격변인과 상관이 높다는 가설은 부분적으로(성실성에서만) 지지되었고 인지능력에서는 유의한 차이를 보여주지 못하였다.

다음으로 반응대안 채점방식에서 두 지시문에 의해서 상황판단검사의 구성개념이 달리 채점되는지를 탐색해 보았다.

표 4에서 보는 바와 같이 먼저 인지능력과의 상관을 보면, 지식형 지시문의 상황판단검사와 행동경향 지시문의 상황판단검사 둘 다 유의하긴 하였지만 지식형 상황판단검사가 행동경향 상황판단검사에 비해 인지능력과 더 큰 상관값을 보여주고 있다(-.23 vs -.15, $p<.05$). 그러므로 가설 1인 '지식형 지시문의 상황판단검사는 행동경향 지시문의 상황판단검사에 비해 인지능력과 상관이 높을 것이다'는 지지되었다.

성격요인들과의 관련성을 살펴보면, 앞의 시나리오 채점과 유사하게 유의한 상관이 단 하나밖에 없다(행동형 SJT와 성실성). 행동경향 지시문의 상황판단검사가 성격5요인의 변인들 중 성실성과 유의한 상관을 보이고 있다. 그러므로 '행동경향성 지시문의 상황판단검사는 지식형 지시문의 상황판단검사에 비해 성격변

표 4. 상황판단검사, 성격검사, 인지능력 검사간의 영차상관(반응대안채점)

	평균	표준편차	행동경향 SJT	지식형 SJT	외향성	원만성	성실성	신경증	개방성	인지능력
행동형 SJT	105.3	24.5	.75	.80	.07	-.08	-.18	.06	.02	-.19
지식형 SJT	92.7	22.4	.62*	.80	.03	-.02	-.04	.09	-.01	-.28
외향성	3.2	.72	.06	.03	.85	.44	.15	-.13	.22	.00
원만성	3.6	.53	-.06	-.02	.36**	.76	.32	-.25	.17	-.02
성실성	3.1	.62	-.14**	-.03	.12**	.25**	.79	.18	.17	-.11
신경증	2.9	.73	.05	.07	-.10*	-.19**	-.14**	.80	.16	-.03
개방성	3.2	.55	.01	-.01	.18**	.13**	.13**	.13**	.77	.22
인지능력	24.5	7.3	-.15**	-.23**	.00	-.01	-.09*	-.02	.17**	.82

주. ** 0.01수준, * 0.05수준.

대각선 아래는 상관, 대각선은 내적일관성신뢰도(α), 대각선 위는 비신뢰도에 대한 교정된 상관값,

인지능력 및 성실성이 상황판단검사와 음(-)의 상관을 가지는 이유는 채점시 전문가와의 차이를 점수로 하였기 때문

인과 상관이 높을 것이다'라는 가설 2는 부분적으로만 지지된다.

표 3과 표 4에서 살펴본 바와 같이 채점방식에 따라서 상황판단검사의 구성개념에 대한 가설들이 지지되지 않거나 지지되는 경우가 있어도 매우 부분적이다. 그러나 반응대안 채점이 인지능력과 더 큰 상관을 보여주고, 시나리오 채점은 성격변인들과 더 큰 상관을 보여주는 것으로 보아 지식형 상황판단검사를 사용할 때는 반응대안채점을, 행동경향 상황판단검사를 사용할 때는 시나리오 채점을 하는 것이 타당화에 도움이 될 수 있다.

상황판단검사의 준거관련 타당도

가설 3에서 저자는 두 지시문의 상황판단검사가 각기 다른 준거와 관련성이 클 것이라고 기대하였다. 즉 행동경향 지시문의 상황판단검사는 맥락수행에 대해 지식형 지시문의 상황판단검사에 비해 높은 타당도를 가질 것으로 예측하였고, 지식형 지시문의 상황판단검사는 과업수행에 대해 행동경향 지시문의 상황판단검사에 비해 높은 타당도를 가질 것으로 예상하였다. 먼저 시나리오 채점에서의 상황판단검사의 준거관련 타당도를 살펴본 다음, 반응대안 채점에서의 준거관련 타당도를 살펴보도록 한다.

시나리오 채점의 경우, 맥락수행에서는 행동경향 상황판단검사가 지식형 상황판단검사에 비해 높은 준거 관련 타당도를 보여주고 있다(.35 vs .17, p<.01). 반면 과업수행의 경우에는 지식형 상황판단검사와 행동경향 상황판단검사 모두와 유의한 상관이 있지만 두 지시문 방식간에 상관의 차이가 통계적으로 유의하지는 못하였다. 그러므로 가설 3a인 "*행동경향 지시문의 상황판단검사는 맥락수행 준거에 대해서 지식형 지시문의 상황판단검사에 비해 높은 타당도를 가질 것이다.*"는 지지되었지만 가설 3b는 지지되지 못하였다.

반응대안 채점의 경우, 맥락수행에 대해서는, 행동경향 상황판단검사가 지식형 상황판단검사에 비해 유의미한 상관을 보이고 있다(-.19 vs -.10, p<.05). 그러므로 가설 3a인 "*행동경향 지시문의 상황판단검사는 맥락수행 준거에 대해서 지식형 지시문의 상황판단검사에 비해 높은 타당도를 가질 것이다.*"는 지지되었다. 하지만 시나리오 채점과 마찬가지로 과업수행의 경우에는 지식형 상황판단검사가 행동경향 상황판단검사에 큰 상관의 패턴을 보이고 있지만 통계적으로 유의한 차이는 아니다.

표 5와 6에서 보는 바와 같이 지시문에 상관없이 두 상황판단검사가 준거에 유의한 타당도를 보이고 있다. 이는 이전 연구(Chan & Schmitt; 2002)의 결과와 같이 지시문과 채점방식에 상관없이 과업수행과 맥락수행 모두에 유의한 준거관련 타당도를 보인다는 것과 일치한다. 다만 맥락수행에 있어서만은 행동경향 상황판단검사가 비교적 큰 준거관련 타당도를 보여주고 있다.

연구문제 - 지시문에 따른 증분타당도 비교

다음은 연구문제로 지시문에 따른 증분타당도를 비교하고자 한다. 즉, 지식형 지시문이 사용될 때 과업수행에 대해 인지능력과 성격 5요인을 감안하고도 상황판단검사가 유의하게 추가적으로 설명하는 부분이 있는지, 그리고 행동경향 지시문이 사용될 경우 맥락수행에 대해 성격 5요인과 인지능력을 감안하고도 상황판단검사가 유의하게 추가적으로 설명하는

표 5. 상황판단검사의 준거관련 타당도(시나리오 채점)

	평균	표준편차	행동경향 SJT	지식형 SJT	맥락수행	과업수행
행동경향 SJT	17.18	5.62	.75	.59	.46	.23
지식형 SJT	20.01	5.55	.45**	.80	.22	.23
맥락수행	3.1	.6	.35**	.17**	.78	.04
과업수행(퀴즈)	14.7	4.8	.18**	.18**	.03	.81

주. ** 0.01수준.
대각선 아래는 상관, 대각선은 내적일관성신뢰도(α), 대각선 위는 비신뢰도에 대한 교정된 상관값

표 6. 상황판단검사의 준거관련 타당도(반응대안채점)

	평균	표준편차	행동경향 SJT	지식형 SJT	맥락수행	과업수행
행동경향 SJT	105.3	24.5	.75	.80	-.25	-.33
지식형 SJT	92.7	22.4	.62**	.80	-.12	-.36
맥락수행	3.1	.6	-.19**	-.10*	.78	.04
과업수행(퀴즈)	14.7	4.8	-.25**	-.29**	.03	.81

주. ** 0.01수준, * 0.05수준.
대각선 아래는 상관, 대각선은 내적일관성신뢰도(α), 대각선 위는 비신뢰도에 대한 교정된 상관값
상황판단검사와의 상관이 음(-)인 이유는 채점시 전문가와의 차이를 점수로 하였기 때문

표 7. 지식형 상황판단검사의 증분타당도(N=472)

모형	종속변수	모형	오차항 자유도	R^2	ΔR^2
모형 1	과업수행(퀴즈)	인지능력	466	.26**	
모형 2	과업수행(퀴즈)	모형1 + 성격5요인	465	.28**	.02**
모형 3	과업수행(퀴즈)	모형2 + 지식형SJT	464	.31**	.03**

주. ** 0.01수준, * 0.05수준

부분이 있는지를 알아보고자 위계적 회귀분석을 실시하였다. 첫 번째로, 지식형 지시문의 상황판단검사의 경우(표 7 참조) 과업수행을 준거로 하여 모형 1에서는 먼저 인지능력검사를 회귀식에 넣고, 다음으로 모형 2에서는 모형 1에다가 성격변인들을 추가하였다. 마지막으로 모형 3에서는 모형 2에 상황판단검사를 추가하여 추가적인 설명량(ΔR^2)을 확인하였다.

표 7의 결과에서 보는 바와 같이 과업수행에 대해 인지능력만으로도 유의미한 설명력을 나타내었고(.26), 인지능력에 성격 5요인을 추가한 모형 역시 과업수행을 잘 설명하였다(.28). 마지막으로 인지능력과 성격 5요인 그리고 상황판단검사가 들어간 모형 역시 과업수

표 8. 행동경향 상황판단검사의 증분타당도(N=472)

모형	종속변수	모형	오차항 자유도	R^2	ΔR^2
모형 1	맥락수행	성격5요인	470	.29**	
모형 2	맥락수행	모형1 + 인지능력	465	.29**	.00
모형 3	맥락수행	모형2 + 행동경향SJT	464	.32**	.02**

주. ** 0.01수준, * 0.05수준

행을 잘 설명하였다(.31). 더불어 인지능력을 감안하고도 성격 5요인이 추가적으로 설명하는 부분이 유의미하였고(.02), 인지능력과 성격 5요인을 감안하고도 지식형 상황판단검사가 .03에 해당하는 추가적인 설명량을 가짐을 알 수 있다.

두 번째로 행동경향 지시문의 상황판단검사의 경우(표 8 참조), 맥락수행을 준거로 하여 먼저 모형 1에서는 성격5요인을 회귀식에 넣었고, 다음으로 모형 2에서는 모형 1에 인지능력을 넣었다. 마지막으로 모형 3에서는 모형 2에 행동경향 지시문의 상황판단검사를 추가하여 추가적인 설명량을 확인하였다.

표 8에서 보는 바와 같이 맥락수행에 대해 성격 5요인이 설명하는 부분은 유의미하게 나타났다(.29). 그러나 성격 5요인에 인지능력을 추가한 모형은 맥락수행에 대해 유의미한 설명력을 보였으나, 인지능력에 의해 추가적으로 설명하는 부분은 유의미하지 않았다. 마지막으로 성격 5요인과 인지능력에 행동경향 지시문의 상황판단검사를 추가한 모형 3 역시 맥락수행에 대해 유의미한 설명을 하였고, 또한 인지능력과 성격 5요인을 감안하고도 행동경향 상황판단검사가 유의미하게 추가적으로 설명력을 보였다(.02). 이 정도의 크기는 기존 연구의 결과에 대응하는 크기라고 할 수 있다.

논의 및 제언

연구의 논의

본 연구를 통해 적어도 상황판단검사는 제시되는 지시문에 따라 다른 반응을 산출한다고 할 수 있다는 가능성을 보여주었다고 할 수 있다. 상황판단검사를 지식형 지시문으로 제시하였을 경우, 참여자가 상황과 그에 따른 행동대안들을 인지적으로 판단할 수 있을 것으로 예상된다. 즉, 주어진 상황을 해결하기 위해 어떠한 행동전략이 효율적인가를 인지적으로 판단하여 제시된 대안 반응 내에서 선택하게 될 것으로 예상된다. 행동경향 지시문이 제시되었을 경우, 참여자는 인지적인 판단보다는 과거의 행동에 비추어 자신의 행동이 어떤 대안으로 행동할지를 예상하고 그러한 방향으로 대안을 결정할 것이다. 즉 습관이나 과거의 행동, 성격과 같은 부분이 앞으로의 행동에 영향을 미칠 것이다.

본 연구에서 제시된 준거관련 타당도는 기존의 연구와 비교할 때 만족할만한 수준이라고 할 수 있다. 준거관련타당도의 경우 Motowidlo와 Tippins(1993)의 연구에서는 수행평정과의 상관이 .28로 Pulakos와 Schmitt(1996)의 연구에는 두 표본에서 .20과 .14로, Weekley와 Jones(1999)의 연구에서는 두 표본에서 .16, 19로 나

타나고 있다. 본 연구에서는 시나리오 채점의 경우 행동경향 지시문과 맥락수행간의 상관이 .35로, 반응대안 채점에서 지식형 지시문과 과업수행간의 상관이 .29를 나타내고 있다.

하지만 본 연구에서는 아쉽게도 지시문 간의 명백한 차이를 보여주지 못하였다. 여기서는 이러한 원인을 살펴보고자 한다.

두 지시문간에 상관값의 차이를 보여주지 못한 이유는 지시문의 효과를 충분히 주지 못한 점이다. 지시문의 효과가 잘 나타나지 못한 이유는 첫 번째로 3주에서 4주에 걸쳐서 기간을 두고 두 차례 검사가 실시되었지만 여전히 피험자들은 자신이 평정한 반응을 기억하고 지시문에 상관없이 평정하였을 가능성이 있다. 즉, 단일 검사를 하는 것보다 한 피험자에게 동일한 내용의 검사를 지시문만 달리 시행함으로써 기억의 효과를 배제하기 힘들었을 것이다. 두 검사의 상관값(.62)이 이러한 가능성을 보여준다.

두 번째로는 상황판단검사를 수행하는 피험자들의 자세이다. 학교장면이라는 점이 실제 수행자들과 달리 피험자들의 동기에 영향을 미쳤을 것이다. 학교 장면에서 퀴즈는 성적에 기여하는 부분이 크고 그 점수가 드러나는 반면에, 상황판단검사의 수행은 성적에 참여점수로 동일한 점수(실험참여 점수)로 들어가고 그 점수에 대한 중요성이 크지 않기 때문에 반복되는 설문과 검사에 그 참여의 동기와 열정이 지속되기 어려웠다는 점을 들 수 있다. 일반적으로 채용시 적성검사의 일부로 상황판단검사로 사용되는 것과 달리 대학생을 대상으로 하다보니 그러한 효과가 감소되었다고 할 수 있을 것이다.

세 번째로는 위와 같은 맥락에서 지시문의 차이에 대한 피험자의 주의가 부족했을 것이다. 지시문이 거의 동일한 형태로 주어지면서 단지 뒷부분에 가서 바뀌기 때문에 충분히 읽지 않거나 지시를 제대로 듣지 못한 피험자들은 동일한 검사로 오인을 하거나 기억력 테스트로 오인하여 이러한 결과를 산출할 수 있을 것이다.

세 번째로 4개의 검사와 퀴즈를 1-2개월이라는 기간동안 연속으로 측정함에 따라 검사수행의 이탈자들이 많았다.

이러한 원인들로 인하여 두 검사간의 지시의 차이를 충분히 나타내지 못했을 것으로 기대된다.

증분타당도의 경우에는 기존의 연구에서는 준거의 속성을 구분하지 않고, 상황판단검사가 일반정신능력(general mental ability)를 감안하고도 수행에 대한 증분타당도가 있는가에 초점을 두고 이루어졌다(예, Clevenger et al., 2001; Chan & Schmitt, 2002; Weekly & Jones, 1997, 1999) 하지만 본 연구에서는 각 지시문에 따라 과업수행과 맥락수행으로 준거를 구분하였고 각각 인지능력과 성격변인을 감안하고도 증분타당도가 있다는 것을 보고하고 있다.

더불어 본 연구에서는 위의 결과들의 가능성을 바탕으로 상황판단검사 제작 및 사용에 관한 다음과 같은 제안을 하고자 한다.

첫째, 상황판단검사를 사용하는 목적에 따라 지시문뿐만 아니라 채점방식을 달리할 것을 제안한다. 즉 과업수행을 예측하는 경우라면 전문가와의 차이가 현저하게 드러날 수 있는 문항별 채점을 사용할 것을 권하고, 맥락수행을 예측하는 경우라면 보고자 하는 행동 즉 최선과 최악의 대안에만 초점을 맞춰서 채점하는 시나리오별 채점을 권고하는 바이다.

둘째, 응용장면에서 지시문에 따른 채점방

식 사용에 대한 제안을 해본다. 상황판단검사로 인지능력과 성격과는 다른, 개인의 상황판단능력을 보고자 한다면 지식형 지시문의 상황판단검사를 사용하여 회사 내 전문가와의 차이를 보는 방식으로 채점하여 신입사원 선발시 이용하도록 권한다. 한편, 개인 특유의 행동특성을 성격과 달리 조직 내에서 발생 가능한 상황이라는 틀 내에서, 해당 조직의 특성과 가치 문화와 얼마나 유사한가를 보고자 한다면, 행동경향 지시문의 상황판단검사를 시나리오 채점방식을 이용하여 사용할 것을 권한다. 즉 모집이나 신입사원 선발시 초기 선별(screening)의 목적으로 사용할 것을 권하는 바이다.

참고문헌

박경숙, 현주, 이재분, 오승민 (1991). KEDI 집단지능검사 개발연구. 한국교육개발원.

신현균, 장재윤 (2003). 대학 4학년생의 성격특성과 성별에 따른 취업 스트레스 및 정신건강. 한국심리학회지: 임상, 22(4), 815-827.

이상철 (2003). 보험설계사 영역에서 지필형 상황판단검사의 구성개념 탐색 및 확인. 성균관대학교 대학원 심리학과 석사학위 논문.

이상철, 이순묵, 조영일 (2003). 지필형 상황판단검사에 대한 비평적 고찰. 한국심리학회지 : 산업 및 조직, 16(3), 129-154.

이순묵 (2004). 상황역량 측정에서 상황에 대한 두 관점: 측정오차인가 해석되어야 할 환경요인인가? 한국심리학회지: 산업 및 조직, 17(2), 243-263.

Borman, W. C., White, L. A., Pulakos, E. D., & Oppler, S. H. (1991). Models of supervisory job performance ratings. *Journal of Applied Psychology, 76,* 863-872.

Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W. Borman(Eds.), *Personnel selection in organizations*(pp. 71-98). New York: Jossey-Bass.

Chan, D., & Schmitt, N. (1997). Video-baded versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology, 82,* 143-159

Chan, D., & Schmitt, N. (2002). Situational Judgment and Job Performance. *Human Performance, 15(3),* 233-254

Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Harvey, V. S. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology, 86(3),* 410-417.

Conn, S. R., & Rieke, M. L. (1994). The 16PF fifth edition technical manual. Champaign, IL: Institute for personality and ability Testing

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51,* 327-358

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 classfication. Mahwah, N. J.: Lawrence Erlbaum Associates.

Latham, G. P. (1989). The reliability, validity, and practicality of the situational interview. In R. W. Eder G. R. Ferris(Eds.), The employment interview: theory, research, and practice(pp. 169-182). Newbury Park, CA: Sage.

McDaniel, M. A., Hartman, N. S., & Grubb Ⅲ, W. L. (2003). Situational Judgment Tests, Knowledge, Behavioral Tendency, and Validity: A Meta-Analysis. Paper presented at the 18th Annual Conference of the Society for Industrial and Oraganizational Psychology. Orlando, FL. April, 2003.

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: a clarification of the literature. *Journal of Applied Psychology. 86(4),* 730-740.

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., Yost, A. P., Ludwick, M. H., Hense, R. L., & Hartman, N. S. (2004). Paper presented at the 19th Annual Conference of the Society for Industrial and Oraganizational Psychology. Chicago. April, 2004.

Motowidlo, S. J.(2000). Some basic issues related to contextual performance and organizational citizenship behavior in human resource management. *Human Resource Management Review, 10,* 115-126

Motowidlo, S.. J., Borman, W., & Schmit, M. (1997). A theory of individual differences in task and contextual performance. *Human Performance, 10,* 71-83.

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative seleciton 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 & G. R. Wheaton(Eds.), *Applied measurement methods in industrial psychology.* Palo Alto, CA: Davies-Black Publishing.

Motowidlo, S. J., & Tippins, N. (1993). Further studies of the low-fidelity simulation in the form of a situation inventory. *Journal of Occupational and Organizational Psychology, 66,* 337-344.

Motowidlo, S. J., & Van Scotter, J. R. (1994) Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology. 79(4),* 475-480.

Ployhart, R. E., & Ehrhart, M. G., (2003). Be careful what you ask for: effects of response instructions on the construct validity and reliability of situational judgment tests. *International Journal of Selection and Assessment. 11(1),* 1-16

Pulakos, E. D., & Schmitt, N. (1996). An evaluation of two strategies for reducing adverse impact and their effects on criterion-related validity. *Human Performance, 9,* 241-258

Reynolds, D. H., Sydell, E. J., Scott, D. R., &

Winter, J. L. (2000). Factors affecting situational judgment test characteristics. Paper presented at the 15th Annual Conference of the Society for Industrial and Organizational Psychology. New Orleans, LA. April.

Smith, K. C., & McDaniel, M. A. (1998). Criterion and construct validity evidence for a situational judgment measure. Paper presented at the 13th Annual Conference of the Society for Industrial and Organizational Psychology. Dallas, TX.

Smith, C. A., Organ, D. W., & Near, J. P. (1993). Organizational citizenship behavior: Its nature and antecedents. *Journal of Applied Psychology, 68,* 653-663.

Steiger, J. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87,* 245-251.

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 judgment 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.

Wagner, R. K., & Sternberg, R. J. (1985). Practical intelligence in real-world pursuits: the role of tacit knowledge. *Journal of Personality and Social Psychology, 49(2),* 436-458

Weekley, J. A., & Jones, C. (1997). Video-based situational testing. *Personnel Psychology, 50,* 25-49.

Weekley, J. A., & Jones, C. (1999). Further studies of situational tests. *Personnel Psychology, 52,* 679-700.

1차 원고접수 : 2005. 9. 7
2차 원고접수 : 2005. 11. 2
최종 게재결정 : 2005. 11. 8

한국심리학회지: 산업 및 조직
Korean Journal of Industrial and Organizational Psychology
2005. Vol. 18, No. 3, 547-565

Comparison of Validities for Different Response Instructions and Scoring Alternatives in Situational Judgement Tests

Min U Kang
GⅡR HR department

Chang-Young Youn
Daegu University

Soonmook Lee
Sungkyunkwan University

The present study starts with hypotheses that response instructions and scoring alternatives influence construct validities, criterion-related validities and incremental validities. The results show possibilities that validities could vary with response instructions and scoring scoring alternatives. In a situational judgment test there are two types of instructions: One is to draw knowledge responses and the other is to draw behavioral tendency responses. In scoring the tests a scenario can be the unit or a response alternative under a scenario can be the unit. First, construct validities can be different depending on response instructions. Second, knowledge response instructions are more correlated with task performance when a scenario is the scoring unit and behavioral tendency response instructions with contextual performance when response alternatives are scored. Third, meaningful incremental validities are observed for each of response instructions. Finally the reasons why effectiveness of response instructions were not clearly differentiated are discussed.

key words : situational judgment test, knowledge response instruction, behavior tendency response instruction, scoring alternatives, task performance, contextual performance