한국심리학회지: 산업 및 조직
Korean Journal of Industrial and Organizational Psychology
2026. Vol. 39, No. 2, 317-342
https://doi.org/10.24230/kjiop.v39i2.317-342

평가센터의 구성개념 타당도: MTMM 상관행렬과 CFA 모형 비교

임 희 철　　이 주 일†

한림대학교
심리학과　　응용심리연구소

본 연구는 국내 A기업 책임급 승진 AC 자료(N = 324)를 활용하여, MTMM 상관행렬 패턴을 통해 수렴·변별 타당도 단서를 탐색하고, CFA 기반 점수 구조 모형(차원모형, 과제모형, 차원-과제 결합모형)을 비교하여 차원효과와 방법효과의 상대적 기여를 검토하였다. MTMM 상관분석 결과, 동일 과제 내 상이한 차원 간 상관이 높게 나타났으며 상이한 과제 간 동일 차원 상관이 비교적 낮게 나타나 방법효과의 개연성이 확인되었다. CFA 결과, 차원모형은 자료를 충분히 설명하지 못한 반면 과제모형은 제한적으로 수용 가능했으며, 차원-과제 결합모형이 가장 우수한 적합을 보였다. 본 연구는 국내 실제 운영 AC의 점수 구조를 제시하여, MTMM 상관 패턴만으로 구성개념 타당도를 단정하기보다 CFA(차원·과제·결합)와 병행하는 다중접근 해석이 필요함을 시사한다. 아울러 역량×상황 결합 관점에서, AC 간소화 국면에서도 중복측정 설계의 유지, 과제당 역량 탑재의 적정화, FoR 기반 평정 절차와 통합회의 운영의 표준화, 차원×과제 점수의 피드백·개발 활용 확대가 요구됨을 제안한다.

주요어 : 평가센터, 구성개념 타당도, 다속성-다방법(MTMM), 확인적 요인분석(CFA)

† 교신저자 : 이주일, 한림대학교 심리학과, (24252) 강원특별자치도 춘천시 한림대학길 1
Tel: 033-248-1723, E-mail: jilee@hallym.ac.kr

Copyright ⓒ 2026, The Korean Society for Industrial and Organizational Psychology. This is an open-access article distributed under the terms of the Creative Commons Attribution –NonCommercial Licenses (https://creativecommons.org/licenses/by-nc/4.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

서 론

평가센터(assessment center, 이하 AC)는 복수의 역량 차원과 평가기법을 기반으로, 실제 업무와 유사한 상황에서 피평가자의 행동을 관찰・기록・평정하는 상황기반 행동평가 도구이다. 일반적으로 AC는 직무분석을 토대로 도출된 핵심 역량을 과제(simulation)별 행동지표로 구체화하고, 복수의 평가자가 구조화된 절차에 따라 행동을 관찰・평정한 뒤 이를 차원별 점수와 최종 종합점수로 통합한다(International Task Force on Assessment Center Guidelines, 2009; International Taskforce on Assessment Center Guidelines, 2015).

국제적으로 AC의 준거관련 타당도는 다수 메타분석에서 반복적으로 보고되어 왔다. Gaugler 등(1987)은 종합평가점수(overall assessment rating)가 승진・직무성과 등 준거와 유의한 관련을 가진다고 보고하였고, Arthur 등(2003)은 차원점수도 상사평정 성과・승진 등 준거와 유의한 수준의 상관을 보인다고 정리하였다. Meriac 등(2008) 역시 AC 차원점수가 인지능력・성격 등과 함께 직무성과 예측에 기여할 수 있음을 제시하였다. 이러한 근거는 AC가 선발・승진뿐 아니라 개발・피드백 맥락에서도 널리 활용될 수 있음을 뒷받침한다.

국내에서도 2000년대 이후 역량 기반 인사관리 담론의 확산과 함께, 민간・공공부문 전반에서 AC 및 역량평가가 점차 도입・활용되어 왔다. 공공부문에서는 2006년 고위공무원단 후보자 역량평가 도입을 시작으로, 2015년 중앙부처 과장급 직위 임용을 위한 역량평가 의무화, 2017년 지방자치단체 국・과장급 후보자 역량평가 시행 등으로 적용 범위가 확대되어 왔다(이상준, 김미경, 2022). 또한 나라장터 공고 분석에 따르면 공공부문 역량평가 관련 용역은 2018년 34건, 2019년 40건, 2020년 45건, 2021년 50건으로 증가하였고, 평가대상을 분리한 기준으로도 같은 기간 47건, 55건, 62건, 68건으로 늘어나 총 169건(분리 기준 232건)이 확인되어 공공부문 역량평가 운영이 양적으로 확대되어 왔음을 보여준다(이상준, 김미경, 2022). 후속적으로 본 연구에서 2025년 기준 나라장터 공고를 추가 확인한 결과[1], 공공부문 역량평가 관련 용역은 총 73건, 평가대상 직급을 분리한 기준으로는 총 88건으로 집계되었다. 이는 선행연구에서 확인된 공공부문 역량평가의 양적 확대가 최근에도 지속되고 있음을 보여준다. 한편 민간부문에서도 H그룹, G그룹 등에서 평가센터가 활용되고 있으며, 실제 사례연구로는 IT기업 K사에서 평가센터를 활용한 직원 역량진단과 직무전문가 선발이 운영된 바 있다. 평가센터의 활용 영역도 선발・승진뿐 아니라 핵심인재 선발, 육성, 훈련 및 개발까지 확장되어 온 것으로 보고되었다(유기원, 2014; 송지은, 2020). 또한 민간부문은 공공부문에 비해 역량평가 결과를 선발・승진뿐 아니라 개인별 역량점수에 기반한 보직, 경력개발, 육성에 보다 적극적으로 활용하는 경향이 제시된 바 있다(권용만, 정장호, 2021). 그러나 실제 운영자료를 바탕으로 AC 점수 구조와 구성개념 타당도를 직접 검토한 국내 연구는 여전히 제한적이다.

AC의 확산과 별개로, “AC가 실제로 무엇을

1) 2025년 수치는 이상준・김미경(2022)의 나라장터 공고 분석 방식에 준하여 본 연구에서 추가 확인한 결과이다. 나라장터에서 공공부문 역량평가 관련 공고를 검색한 뒤, 중복 공고 및 역량평가와 직접 관련이 없는 사례를 제외하고 집계하였다.

측정하는가?"라는 구성개념 타당도 문제는 국제적으로 장기간 논쟁의 대상이었다. 초기 AC 연구와 운영 지침은 AC가 직무에서 요구되는 핵심 역량 차원을 과제 전반에서 구분·측정할 수 있다는 전제하에, 차원별 점수(dimension ratings)의 해석과 활용을 중요한 가치로 강조해 왔다(International Task Force on Assessment Center Guidelines, 2009; International Taskforce on Assessment Center Guidelines, 2015; Thornton & Rupp, 2006). 그러나 이후 MTMM 기반의 점수 구조 연구와 메타분석 결과는, 동일 차원의 과제 간 수렴이 기대만큼 강하지 않은 반면 동일 과제 내 차원 간 상관이 상대적으로 높게 나타나는 경향을 반복적으로 보고하면서, 차원효과와 더불어 과제/방법(exercise/method) 요인이 점수 분산에 상당한 비중으로 기여할 수 있음을 제시해 왔다(Lievens & Conway, 2001; Woehr & Arthur, 2003; Bowler & Woehr, 2006). 이러한 흐름에서 일부 연구자들은 AC 점수가 차원 간 변별적 정보라기보다 과제 특유 수행(exercise-specific performance)과 더불어 여러 지표에 공통으로 작동하는 전반적 수행요인(general performance factor)에 의해 더 강하게 설명될 가능성에 주의를 제기하며, 점수 구조 해석 및 차원점수 활용에 신중함이 필요하다고 논의해 왔다(Jackson et al., 2016; Jackson et al., 2022).

반면 AC가 차원을 "전혀" 측정하지 못한다고 결론내리기에는 근거가 충분히 누적되지 않았다는 반론도 꾸준히 제기되어 왔다(Hoffman et al., 2011; Kuncel & Sackett, 2014). 예컨대 AC 차원점수(특히 과제 간 통합된 overall dimension ratings)가 직무수행 준거를 예측하거나 인지능력·성격을 통제한 이후에도 추가적인 예측력을 보일 수 있다는 메타분석 결과는, 차원점수가 단순한 과제 인상만을 반영하는 것으로 환원되기 어렵다는 점을 시사한다(Meriac et al., 2008). 또한 차원점수의 의미를 외부 준거(예: 자기·상사·고객 평정 등 동일 차원에 대한 외부 평정)와 직접 연결하여 검증하려는 '외적 구성개념 타당도(external construct-related validity)' 접근도 제안·시도되어 왔으나, 결과는 일관되게 지지적이라기보다 오히려 차원요인보다 출처(source) 요인의 우세를 보고하는 등 논쟁을 확장시키는 양상도 관찰된다(Wirz et al., 2020). 한편, 일반화가능도 이론이나 분산분해 관점의 모형은 AC 점수가 단일 요인(차원 또는 과제)으로만 설명되기보다는, 차원·과제·평가자 및 오차 등 복수 분산원이 함께 기여하는 복합 구조임을 반복적으로 보여주었다(Putka & Hoffman, 2013; Jackson et al., 2016). 요컨대 최근의 보다 현실적인 합의는 "AC는 차원을 측정하지 못한다"라는 이분법적 결론이라기보다, 차원효과와 과제/방법효과가 함께 존재하는 결합(통합) 구조로 AC 점수 구조를 이해하는 쪽에 가깝다(Hoffman et al., 2011; Kuncel & Sackett, 2014).

국내 연구에서도 AC 점수 구조를 차원효과와 과제/방법효과의 관점에서 점검하려는 시도가 일부 이루어져 왔다. 허창구와 신강현(2010)은 내재설계 평가센터 자료에 일반화가능도 이론을 적용하여, 차원 관련 분산이 완전히 소거되지는 않으나 그 크기가 제한적이며 과제·평가자 및 오차 분산이 점수에 상당한 비중으로 기여할 수 있음을 보고하였다. 또한 허창구(2011)는 대학 선발 장면의 AC 자료를 대상으로 일반화가능도 분석과 MTMM 상관 패턴을 함께 검토하여, 동일 차원의 과제 간 수렴이 기대만큼 강하지 않고 동일 과제 내 차원 간 상관이 상대적으로 높게 나타나는 경향을 바탕으로, AC 점수 구조에서 차

원효과와 더불어 방법효과가 함께 작동할 가능성을 제시하였다. 다만 국내 연구들은 점수 구조를 점검하는 방법론을 일부 제시해 왔음에도, 동일한 중복측정 AC 자료에 대해 MTMM 상관 기반 해석과 CFA 기반 점수 구조 모형 비교를 병행 적용하고, 두 접근이 제공하는 정보의 공통점과 차이를 직접 연결해 검토한 연구는 보고되지 않았다.

요컨대 AC 구성개념 타당도 논쟁은 본질적으로 AC 점수 구조(score structure)를 어떻게 이해할 것인가의 문제다. 즉, 점수가 이론적 역량 차원에 의해 주로 구조화되는지, 방법 요인이 독립적인 분산원을 형성하는지, MTMM 접근과 CFA 접근이 어떤 조건에서 서로 다른 결론을 유도하는지에 대한 체계적 검토가 요구된다. 이러한 문제는 단순한 측정모형 간 비교에 그치지 않는다. 실제 AC 운영에서는 차원점수가 피드백 제공, 개발계획 수립, 후속 교육 연계, 승진·선발 의사결정의 근거로 활용되므로, 점수가 주로 차원을 반영하는지 아니면 과제 상황의 영향이 크게 개입된 결과인지를 구분하는 일은 결과 해석의 정당성과 직결된다. 따라서 AC 점수 구조에 대한 검토는 통계적 모형 비교 자체보다, 차원점수를 어느 수준까지 해석 및 활용할 수 있는지 판단하는 실무적 근거를 제공한다는 점에서 중요하다.

본 연구는 국내 기업에서 운영된 책임급 승진 AC 자료를 활용해, 중복측정 설계를 갖춘 AC의 점수 구조를 MTMM과 CFA로 동시에 분석하고 그 결과를 직접 비교하는 것을 1차 목적으로 한다. 특히 동일 자료에서 MTMM이 제공하는 수렴·변별 단서와 CFA 경쟁모형 비교가 제공하는 점수 구조 증거를 병렬적으로 제시하고 상호보완적으로 해석함으로써, 국내 AC 점수 해석의 근거를 보다 정교하게 제시하고자 한다. 이를 통해 국내 AC가 차원·방법 관점에서 어떤 구조적 특성을 갖는지 경험적 근거를 제시할 뿐 아니라, 실제 운영 장면에서 AC 결과점수의 해석과 활용 범위를 판단하는 데 필요한 근거를 제공하고자 한다.

연구문제 1. MTMM 상관행렬을 바탕으로, 동일 역량-이질 과제 상관과 이질 역량-동일 과제 상관을 포함한 상관 패턴이 어떻게 나타나는지를 확인하고, 이 패턴이 차원 간 수렴·변별 타당도 및 차원효과·과제효과에 대해 시사하는 바를 검토한다.

연구문제 2. 동일 자료에 대해 CFA 기반 점수 구조 모형(차원모형, 과제모형, 차원-과제 결합모형)을 각각 적합시킨 뒤, 모형 적합도 지표와 주요 모수를 비교하여 어떤 점수 구조 가설이 자료를 상대적으로 더 잘 설명하는지를 확인한다.

연구문제 3. MTMM 기반 해석과 CFA 기반 모형 비교 결과를 종합하여, 두 접근이 제시하는 결론이 어떤 점에서 일치하거나 차이를 보이는지를 정리하고, 이러한 차이가 국내 AC 설계·운영 및 결과 해석(특히 차원점수 활용)에 주는 시사점을 도출한다.

구성개념 타당도와 AC 점수 구조

구성개념 타당도(construct validity)는 측정도구가 의도한 심리적 구성개념을 얼마나 적절히 반영하는지에 대한 판단으로, 단일한 지표로 결정되기보다는 점수에 근거한 해석과 활용을 지지하는 다양한 증거의 축적을 통해 강

화되는 개념으로 이해된다(Cronbach & Meehl, 1955; Messick, 1995). Messick(1995)은 타당도를 내용, 구조, 준거, 결과 등 서로 다른 범주의 증거를 통합하여 점수 해석의 정당성을 평가하는 통합적(unified) 개념으로 제시하였으며, 이 관점에서 타당도는 검사 '자체'의 고정된 속성이라기보다 특정 점수 해석과 사용을 정당화하는 논증 과정에 가깝다. 또한 Standards for Educational and Psychological Testing은 타당도를 점수 해석과 활용을 뒷받침하는 증거와 이론의 축적을 통해 평가되는 것으로 정의하면서, 특히 해석과 사용의 맥락에 따라 요구되는 타당도 근거가 달라질 수 있음을 강조한다(American Educational Research Association et al., 2014).

인사선발 · 승진과 같은 고위험(high-stakes) 의사결정 맥락에서 구성개념 타당도의 중요성은 더욱 크다. Binning과 Barrett(1989)은 인사평가 맥락에서 타당도를 "측정점수 → 추론 → 의사결정"으로 이어지는 연쇄적 고리로 개념화하며, 그 중심에 구성개념 타당도를 위치시켰다. 즉 점수는 단순한 수치가 아니라 특정 구성개념(역량)에 대한 추론을 매개하며, 그 추론은 승진 · 선발 · 배치와 같은 결정으로 연결된다. 따라서 "점수가 무엇을 의미하는가"에 대한 근거가 충분하지 않다면, 점수의 예측력이 일정 수준 확보되더라도 점수 해석의 정당성과 활용의 정합성은 약화될 수 있다(Messick, 1995; American Educational Research Association et al., 2014).

평가센터(assessment center, AC)에서 구성개념 타당도 문제는 "AC 점수가 실제로 무엇을 반영하는가"라는 질문으로 구체화된다. AC는 복수의 시뮬레이션 과제에서 관찰된 행동을 바탕으로 차원점수(dimension ratings)와 종합점수(overall assessment rating; OAR 또는 종합평정)를 산출하며, 실무에서는 차원점수를 근거로 피드백 제공, 개발계획 수립, 후속 교육 연계, 승진 의사결정의 근거 제시 등이 이루어진다. 이 과정은 차원점수가 의도한 역량 구성개념을 충분히 대표한다는 전제를 포함한다. 그러나 AC 연구에서는 차원점수의 변별성이 기대만큼 뚜렷하지 않고 방법 요인이 점수 구조에 크게 기여할 수 있다는 결과가 반복적으로 보고되어 왔다(Lievens & Conway, 2001; Woehr & Arthur, 2003; Bowler & Woehr, 2006; Lance, 2008). 따라서 AC 점수의 구성개념 타당도는 "차원점수의 의미를 얼마나 정당화할 수 있는가"라는 점수 구조(score structure) 문제로 수렴된다(Hoffman et al., 2011; Kuncel & Sackett, 2014).

Campbell과 Fiske(1959)는 구성개념 타당도의 핵심 논리로 수렴타당도(convergent validity)와 변별타당도(discriminant validity)를 제시하였다. 수렴타당도는 동일(또는 매우 유사한) 구성개념을 서로 다른 방법으로 측정했을 때 지표들이 충분히 높은 상관을 보이는지를 의미하며, 변별타당도는 서로 다른 구성개념들이 과도하게 높은 상관을 보이지 않고 구분되는지를 의미한다. AC 맥락에서 이는 동일 역량을 서로 다른 과제로 측정했을 때(동일 특성-이질 방법) 점수가 얼마나 일관되게 수렴하는지, 그리고 서로 다른 역량이 동일 과제 내에서 과도하게 함께 움직이지는 않는지(이질 특성-동일 방법)를 통해 수렴 · 변별의 단서를 얻는 방식으로 구현된다. 즉 AC의 점수 구조 분석은 "차원 중심 구조"가 관찰되는지, 혹은 "방법 중심 구조"가 관찰되는지, 또는 두 효과가 함께 작동하는 "결합 구조"가 관찰되는지에 대한 검토와 직결된다(Lievens & Conway, 2001;

Hoffman et al., 2011).

현대 심리측정에서는 잠재변인 모형에 기반한 점수 구조 분석(CFA/ESEM 등)이 구성개념 타당도 근거를 평가하는 대표적 방법으로 널리 활용된다(Brown, 2015; Marsh et al., 2014). 이러한 접근의 장점은, 상관 패턴에 근거하여 간접적으로 추론하던 차원효과와 방법효과를 잠재변인 수준에서 보다 명시적으로 분리 추정하고, 대안적 구조 가설을 모형 적합도를 통해 직접 비교할 수 있다는 점이다. Standards for Educational and Psychological Testing 역시 점수의 내적 구조가 의도한 구성개념과 부합하는지 여부를 중요한 타당도 근거로 제시하며(American Educational Research Association et al., 2014), AC처럼 복수 과제에서 복수 차원을 반복 평정하고 과제・평가자 등 복수 분산원이 동시에 기여할 수 있는 평가체계에서는, 내적 구조 분석이 차원점수 해석의 정당화를 위한 핵심 근거가 된다(Hoffman et al., 2011; Jackson et al., 2016; Putka & Hoffman, 2013).

다만 AC에서 "방법(method)"은 전통적 심리검사에서의 방법과 달리 단순한 측정 형식이 아니라, 서로 다른 상황・과업 요구・상호작용 맥락을 포함하는 복합 자극(behavioral sample)에 가깝다(Hoffman et al., 2011; Jackson et al., 2016). 동일 역량명이라 하더라도 발표 과제와 역할수행 과제는 요구되는 행동양식이 다르고, 그 결과 동일 역량-이질 과제 상관이 낮게 관찰되더라도 이를 곧바로 수렴타당도의 결여로 단정하기는 어렵다. 즉 AC 점수는 역량 자체뿐 아니라 역량이 발현되는 상황(과제)의 특수성과 결합되어 나타날 가능성이 크며, 이러한 역량×상황의 결합성은 MTMM의 고전적 "동일 특성" 가정과 같다고 하기 어렵다. 따라서 AC의 구성개념 타당도 논의는 단순히 "상관이 높아야 한다"는 기준을 반복하기보다, 점수 구조가 어떤 분산원(차원, 방법, 그 외 요인)에 의해 조직되는지를 다각도로 검토하는 방식으로 진행될 필요가 있다.

MTMM 접근과 한계

Campbell과 Fiske(1959)가 제안한 다특성-다방법(multitrait-multimethod, MTMM) 접근은 동일한 특성(trait)을 서로 다른 방법(method)으로 측정했을 때 나타나는 상관구조를 통해 구성개념 타당도의 두 핵심 요소인 수렴타당도와 변별타당도를 평가하는 고전적 방법이다. MTMM에서는 일반적으로 동일 특성-이질 방법(monotrait-heteromethod) 상관이 충분히 높게 나타나는지를 수렴타당도의 단서로, 이질 특성-동일 방법(heterotrait-monomethod) 상관이 과도하게 높지 않은지를 변별타당도의 단서로 해석한다. 또한 이질 특성-이질 방법(heterotrait-heteromethod) 상관은 비교 기준으로 활용되어, 이상적인 MTMM 패턴에서는 동일 특성-이질 방법 상관이 다른 두 유형의 상관보다 유의하게 높게 나타나는 것이 기대된다.

AC 연구에서 MTMM은 다수의 역량(차원)과 다수의 과제(방법)를 교차하여 평정한 점수 구조를 점검하는 데 오랫동안 활용되어 왔다. 그러나 축적된 AC MTMM 연구들은 많은 경우 동일 역량-이질 과제 상관이 기대만큼 높지 않고, 오히려 동일 과제 내 서로 다른 역량 간 상관이 상대적으로 높게 나타나는 경향을 보고해 왔다(Lievens & Conway, 2001; Woehr & Arthur, 2003). 이는 차원효과가 존재하더라도 방법효과가 점수 구조에 상당한 비중으로 기여할 수 있음을 시사하며, 차원점수의 해석 가능성에 대한 논쟁을 촉발하였다(Lance, 2008;

Hoffman et al., 2011; Kuncel & Sackett, 2014).

다만 AC 맥락에서 MTMM 해석은 몇 가지 이유로 단순하지 않다. 첫째, MTMM에서 '동일 특성'으로 분류되는 지표가 실제로는 과제에 따라 상이한 행동 표본을 포함할 수 있다. 예를 들어 동일한 역량명이라 하더라도 발표 과제는 메시지 구성과 설득적 전달을, 역할수행 과제는 상대방의 반응에 따라 질문하고 수용적으로 상대방의 의견을 경청하는 능력을 상대적으로 더 요구할 수 있다. 이러한 경우 동일 역량-이질 과제 상관이 낮게 관찰되더라도 이를 수렴타당도의 결여로 단정하기 어렵다. 즉 AC 점수는 역량 자체뿐 아니라 역량이 발현되는 상황(과제)의 특수성과 결합되어 나타날 가능성이 크며, 이러한 역량×상황 결합성은 MTMM의 고전적 가정과 긴장을 만든다(Hoffman et al., 2011; Jackson et al., 2016).

둘째, MTMM 상관행렬은 차원효과와 방법효과를 분산 수준에서 명시적으로 분리하지 못한다. 동일 과제 내 서로 다른 역량 점수들이 높은 상관을 보일 때, 그것이 평가자의 전반적 인상(halo) 때문인지, 과제 설계가 여러 역량에 공통적으로 요구하는 행동을 많이 포함하기 때문인지, 혹은 평정 절차(관찰-기록-분류-평정)가 충분히 분리되어 실행되지 않았기 때문인지를 상관행렬만으로는 구분하기 어렵다. 다시 말해 MTMM은 AC 점수 구조의 문제를 '발견'하는 데에는 유용하지만, 그 원인을 분해하여 설명하는 데에는 구조적 제약이 있다(Lance et al., 2007).

셋째, MTMM은 대안적 점수 구조 가설(차원 중심 구조, 과제 중심 구조, 차원-과제 결합 구조)을 통계모형 수준에서 엄밀하게 비교하는 데 제한적이다. 상관의 크기와 패턴 비교는 해석의 출발점이 될 수 있으나, 경쟁모형들 간의 적합도 비교나 분산 기여의 추정과 같은 모형 기반 결론을 제공하기는 어렵다. 특히 AC처럼 지표 수가 제한적이고 방법 분산이 큰 상황에서는 상관 패턴만으로는 "차원효과가 없다" 또는 "과제효과가 지배적이다"와 같은 결론을 안정적으로 도출하기 어렵다. 이러한 이유로 MTMM 자료에 확인적 요인분석(CFA)을 적용하여 특성요인과 방법요인을 잠재변인 수준에서 분리 추정하고 모형 간 적합도를 비교하려는 접근이 발전해 왔다(Geiser et al., 2008; Marsh et al., 2014).

요약하면, 본 연구에서 MTMM 상관행렬은 신뢰도 지표라기보다 AC 점수 구조에 대한 수렴 · 변별 타당도 단서를 제공하는 탐색적 진단 도구로 활용된다. MTMM 결과는 그 자체로 결론이 되기보다는, 이후 CFA 기반 점수 구조 모형 비교 결과와 함께 종합적으로 해석될 필요가 있다. 이러한 병행 접근은 MTMM이 제공하는 직관적 패턴 정보(예: 동일 과제 내 상관의 크기, 과제 간 상관 수준)를 유지하면서도, 모형 기반 분석을 통해 차원효과와 방법효과의 상대적 기여를 보다 명시적으로 검토할 수 있다는 점에서 실증적 · 방법론적 설득력을 높일 수 있다.

CFA 기반 점수 구조 모형

확인적 요인분석(confirmatory factor analysis, CFA)은 연구자가 이론에 근거하여 가정한 잠재요인 구조를 사전에 명시하고, 관찰자료가 그 구조에 얼마나 부합하는지를 검증하는 대표적 잠재변인 모형 기반 방법이다(Brown, 2015). 구성개념 타당도 관점에서 CFA는 단순 상관 패턴의 비교를 넘어, 측정점수의 분산이 어떤 잠재요인(예: 역량 차원, 방법 요인)에 의

해 설명되는지, 그리고 경쟁하는 점수 구조 가설 중 무엇이 더 설득력 있는지를 모형 적합도와 모수 추정을 통해 비교할 수 있다는 장점을 제공한다(Marsh et al., 2014). 특히 AC처럼 동일한 피평가자에게 복수 과제를 적용하고 각 과제에서 복수 역량을 평정하는 구조에서는 차원효과와 방법효과가 동시에 관찰될 가능성이 높기 때문에, 이들 효과를 분리·비교하기 위한 점수 구조 모형의 선택이 핵심 쟁점이 된다(Hoffman et al., 2011).

AC 점수 구조 논의에서 가장 기본이 되는 CFA 모형은 차원모형, 과제모형, 그리고 차원-과제 결합모형이다. 첫째, 차원모형(dimension-only model)은 AC가 본래 의도한 "역량 차원 기반 평가" 가설을 반영한다. 이 모형에서는 각 역량 차원을 잠재요인으로 설정하고, 서로 다른 과제에서 측정된 동일 역량 지표들이 해당 차원요인에 로딩하도록 제약한다. 차원모형이 자료에 잘 부합한다면, AC 점수가 주로 역량 차원에 의해 구조화되어 있고 차원 간 변별성이 일정 수준 확보되어 있다고 해석할 수 있다. 이 경우 차원점수의 활용(차원별 피드백, 차원별 의사결정 근거 제시)은 비교적 더 정당화될 수 있다.

둘째, 과제모형(exercise-only model)은 AC 점수가 역량보다는 과제 또는 상황에 의해 주로 조직된다는 가설을 반영한다. 이 모형에서는 각 과제를 잠재요인으로 설정하고, 해당 과제에서 관찰된 여러 역량 지표들이 동일 과제요인에 로딩하도록 설정한다. 과제모형이 차원모형보다 우수한 적합을 보인다면, AC 점수가 이론적으로 정의된 차원보다는 "각 과제에서의 전반적 수행" 또는 "과제 특유 수행"에 더 강하게 구조화되어 있을 가능성을 시사한다. 이러한 결과는 차원점수가 과제 수행의 인상이나 과제 특수성의 영향을 상대적으로 크게 받을 수 있음을 의미하며, 차원점수 해석과 차원기반 피드백은 보다 신중하게 이루어질 필요가 있음을 함의한다.

셋째, 차원-과제 결합모형[2](dimension-exercise hybrid (mixed) model)은 차원효과와 과제효과가 동시에 존재한다는 보다 현실적인 점수 구조 가설을 반영한다(Hoffman et al., 2011). 이 모형에서는 차원요인과 과제요인을 모두 포함하며, 각 관찰지표(예: 과제별 차원점수)는 해당 차원요인과 해당 과제요인에 동시에 로딩하도록 설정된다. 결합모형이 차원모형과 과제모형보다 우수한 적합을 보인다면, AC 점수 구조를 "차원과 과제가 함께 작동하는 결합 구조"로 이해하는 것이 타당하다는 근거가 된다. 다만 결합모형은 모형의 복잡도가 증가하고, 자료 구조(지표 수, 과제×차원 매트릭스 형태, 표본 크기)에 따라 식별 문제나 추정 불안정이 발생할 수 있으므로, 동일 자료에 대해 여러 대안 모형을 병행 비교하며 해석의 일관성을 점검하는 것이 중요하다(Marsh et al., 2014).

한편 MTMM 자료를 CFA로 해석하는 흐름에서는 전통적 MTMM 상관행렬의 한계를 보완하기 위해 방법효과를 보다 명시적으로 다루는 모형들이 제안되어 왔다. 대표적으로 CT-CU(correlated trait-correlated uniqueness) 모형은 특성요인(차원요인)을 설정한 뒤, 동일 방법(예: 동일 과제)에서 측정된 지표들의 고유성(오차/특수성)을 상관시키는 방식으로 방법

2) 본 연구에서는 '혼합모형(mixed model)'이라는 표현이 주는 의미상의 혼선을 줄이기 위해 '차원-과제 결합모형'을 기본 용어로 사용한다. 다만 선행연구에서는 동일 개념을 hybrid model 또는 mixed model 등으로 지칭하므로, 문헌 인용 시 원 용어를 병기한다.

효과를 반영한다(Geiser et al., 2008). CT-CU는 방법효과를 별도의 방법요인으로 직접 설정하지 않더라도, 동일 방법에서 공유되는 잔차 공분산을 통해 방법 관련 분산을 모델링할 수 있다는 점에서 유용하며, 특히 실무 자료에서 결합 모형이 식별 문제를 보이거나 과도한 복잡성으로 추정이 불안정할 때 현실적인 대안이 될 수 있다.

또 다른 접근인 CT-C(M-1) 모형은 하나의 기준 방법(reference method)을 설정하고, 나머지 방법효과를 그 기준에 대한 상대적 편차로 추정하는 방식이다(Eid et al., 2003). CT-C(M-1)은 결합모형에서 빈번히 제기되는 식별 및 해석의 불안정을 완화하고, 방법효과를 비교적 해석 가능한 형태로 제시할 수 있다는 장점을 가진다(Marsh et al., 2014). 예를 들어 특정 과제를 기준 방법으로 두면, 다른 과제에서의 방법효과는 기준 과제 대비 어떤 체계적 편차(방법 특수성)를 가지는지로 표현될 수 있으며, 이는 “방법효과가 존재한다”는 일반적 결론을 넘어 방법효과의 형태를 보다 구조적으로 설명하는 데 도움이 된다.

요약하면, AC 점수 구조를 분석할 때 CFA 기반 모형 비교는 (a) 차원 중심 구조와 과제 중심 구조를 경쟁 가설로 설정하여 직접 비교할 수 있고, (b) 차원효과와 방법효과가 결합된 구조의 가능성을 검증할 수 있으며, (c) 결합 모형의 변형 모형을 통해 방법효과를 보다 정교하게 반영함으로써 전통적 MTMM 상관 기반 해석의 한계를 보완할 수 있다는 장점을 제공한다. 특히 본 연구에서 CT-CU 및 CT-C(M-1) 모형까지 포함하였다. 이는 국내 기업 AC 자료와 같이 과제 수와 지표 수가 제한되고 방법효과가 상대적으로 크게 나타날 수 있는 상황에서 단일한 결합모형 결과만으로 ‘차원효과 vs 방법효과’에 대한 결론을 내리기보다, 방법효과를 서로 다른 방식으로 모델링한 대안 모형들에서도 결론이 일관되게 재현되는지를 확인함으로써 점수 구조 해석의 신뢰성과 설득력을 강화하기 위함이다. 이는 특정 모형의 가정에 의해 결과가 인위적으로 규정될 가능성을 낮추고, 방법효과의 과대 · 과소 추정 위험을 줄이는 방법론적 안전장치로 기능한다(Geiser et al., 2008; Marsh et al., 2014). 따라서 본 연구는 전통적 MTMM 상관행렬 분석을 수렴 · 변별 타당도 단서로 활용하는 동시에, 차원모형 · 과제모형 · 결합모형 및 CT-CU, CT-C(M-1)과 같은 변형 모형을 병행하여 점수 구조를 비교함으로써 국내 AC 자료에서 차원효과와 방법효과가 어떻게 공존하는지에 대한 보다 정교한 근거를 제시하고자 한다.

방 법

연구 대상 및 자료수집

본 연구는 국내 A기업에서 운영된 책임급 승진 평가센터(Assessment Center; AC) 자료를 활용하였다. 본 AC는 책임급 승진 심사의 평가 절차로 운영되었으며, 실제 직무 상황과 유사한 시뮬레이션 과제 수행 과정에서 나타나는 행동을 관찰 · 기록 · 평정하여 역량 수준을 평가했다(International Task Force on Assessment Center Guidelines, 2009; International Taskforce on Assessment Center Guidelines, 2015).

분석 대상은 2023~2024년에 실시된 책임급 승진 AC 응시자 중, 재응시(동일인이 2회 이상 응시한 사례)를 제외한 인원으로 구성하였

다. 재응시 자료는 동일 과제 기반 운영 특성상 연습효과 및 절차 익숙함에 의해 점수 수준이 체계적으로 변할 가능성이 있어, 본 연구의 점수 구조(차원/과제 효과) 검토에 혼입을 유발할 수 있다는 판단에 따라 제외하였다. 최종 분석 표본은 총 324명이다.

본 연구에서 활용한 자료는 평가 당일 생성된 과제별·역량별 평정점수(10개 지표)이며, 인사평가 자료의 특성상 결측치는 존재하지 않았다. 또한 본 연구는 연도별 차이를 연구문제로 설정하지 않았으므로, 연도는 표본의 운영 범위를 나타내는 배경 정보로만 보고하고 모형 분석에서는 통합 표본을 사용하였다.

평가센터 설계 및 평가 구조

AC 설계 개요

본 AC는 3개 과제(기법)에서 5개 역량 차원을 평가하는 중복측정 설계(overlapping design)로 구성되어 있다. 즉, 각 역량은 서로 다른 두 과제에서 총 2회 측정되며, 각 과제에서는 3~4개 역량을 평가한다. 이러한 중복측정 설계는 단일 과제 기반 내재설계(각 역량 1회 측정)에 비해, 동일 역량이 서로 다른 상황(과제)에서 어떻게 나타나는지에 대한 정보를 제공한다는 점에서, 구성개념 타당도(수렴/변별 단서) 및 점수 구조 분석에 보다 적합한 구조를 가진다(Campbell & Fiske, 1959; Hoffman et al., 2011).

본 연구의 분석 자료는 과제별로 산출된 역량 평정점수(과제×역량 셀 점수)이며, 총 10개 지표가 생성된다(5개 역량 × 2회 측정). 본문에서는 기업·과제·역량의 민감 정보를 최소화하기 위해 과제와 역량을 익명화하여 기술하였다. 과제는 각각 Structured Interview(I), Role Play(R), Group Discussion(D)으로, 역량은 1~5로 표기하였다. 이 중 Structured Interview는 면접관이 구조화된 질문을 제시하고, 피평가자가 자신의 경험이나 판단을 언어적으로 진술하는 형식의 과제이다. 반면 Role Play와 Group Discussion은 주어진 상황 내에서 타인과의 상호작용, 의사결정, 발언 및 행동 수행을 직접 요구하는 시뮬레이션 과제로, 과제 구조와 수행 방식에서 차이가 있다. 또한 독자의 이해를 돕기 위해 역량의 성격을 최소한의 범주 수준에서 제시하면, 역량 1~5는 각각 문제해결, 기획·관리, 의사소통, 협업, 구성원 지원과 관련된 역량으로 이해할 수 있다.

과제×역량 매트릭스

본 AC의 과제×역량 매트릭스(중복측정 구조)는 표 1과 같다. 표 1에서 각 셀은 “해당 과제에서 해당 역량이 평가됨”을 의미한다. 이 설계에서 각 역량이 정확히 2회씩 측정되도록 구성되어 있다. 표 1의 매트릭스에 따라 생성되는 관찰 지표는 총 10개이며, 예컨대 과제 I

표 1. 과제 X 역량 매트릭스(중복측정 설계)

	역량 1	역량 2	역량 3	역량 4	역량 5
과제 I	●			●	●
과제 R		●	●	●	●
과제 D	●	●	●		

에서의 역량1 평정은 I1, 과제 R에서의 역량2 평정은 R2, 과제 D에서의 역량3 평정은 D3과 같이 표기할 수 있다. 본 연구의 MTMM 상관행렬 및 CFA 모형은 이러한 10개 관찰 지표의 공분산 구조를 기반으로 수행되었다.

평가자 구성 및 교육 절차

평가자 구성 및 자격

본 AC의 평가자는 사전에 구성된 평가자 풀(pool)에서 일정별로 선발되어 참여하였다. 평가자 풀은 (1) 심리학 · 교육학 · 경영학 등 관련 분야의 전 · 현직 교수, (2) 관련 전공 박사학위 소지자 또는 동등한 학력 · 경력을 보유한 전문가, (3) 국내 AC 실무에서 평가자 역할 수행 경력 10년 이상의 인원 중 선별된 전문가로 구성되었다. 이는 AC 운영 가이드라인이 강조하는 평가자 전문성(훈련과 경험을 갖춘 평가자 확보) 원칙을 반영한 것이다 (International Task Force on Assessment Center Guidelines, 2009; International Taskforce on Assessment Center Guidelines, 2015).

평가자 교육 및 점수 산출 절차

평가자 교육은 (1) 풀 구성 시점의 기본 교육과 (2) 시행 회차별 교육으로 구분된다. 기본 교육은 최초 4시간으로 운영되었고, 회차별 교육은 시행 당일 1시간 내외로 실시되었다. 회차별 교육은 주로 과제 숙지 및 평가기준(행동지표) 재확인, 평가자 간 눈높이 조정(토론 중심)에 초점을 두었으며, 본 자료 운영에서는 비디오 기반 캘리브레이션은 포함되지 않았다. 다만 과제가 장기간(여러 회차) 동일하게 운영되는 특성상, 회차별 교육은 “새로운 과제 학습”보다는 “운영 절차 · 평정 기준의 재정렬” 기능을 수행하는 데 초점이 있었다.

본 AC는 과제별 개별 평정 후, 전체 과제가 종료되면 통합회의(integration discussion)를 통해 점수를 최종 확정하는 절차를 운영한다. 통합회의에서는 각 피평가자에 대해 과제별 평가자가 제시한 평정 결과를 함께 검토하고, 역량별 점수 차이가 크게 나타나는 경우 합의 기반 논의를 통해 평정점수를 조정하며, 조정 시 근거를 기록으로 남겼다. 운영 규칙상 5점 척도에서 동일 역량에 대해 평가자 간 점수 차이가 1.5점 이상인 경우 논의를 통해 조정하며, 1점 차이는 필요 시 논의 · 수정할 수 있도록 하였다. 조정은 높은 점수를 준 평가자 또는 낮은 점수를 준 평가자 중 어느 쪽이 수정할지는 토론을 통해 결정하였다. 본 연구의 분석 자료는 통합회의를 거친 최종 확정 점수이다.

측정변인 및 분석 절차

측정변인 구성 및 점수 산출

본 연구에서 분석 대상으로 삼은 역량은 총 5개이다. 본 역량모델은 승진평가 시행 이전에 A기업 책임급 직무에서 고성과자로 평가되는 현직자를 대상으로 성공사례 기반 인터뷰(예: BEI, FGI 등)를 통해 수행 행동을 수집하고, 이를 바탕으로 역량 정의와 행동지표를 구성한 뒤, 외부 전문가 및 기업 내부 전문가, 전 · 현직 교수 등 다수의 검토를 거쳐 최종 확정된 것이다. 이러한 절차는 AC가 직무분석/직무요건에 기반해 역량과 행동지표를 정교화해야 한다는 가이드라인의 원칙과 부합한다 (International Task Force on Assessment Center Guidelines, 2009; International Taskforce on

Assessment Center Guidelines, 2015).

분석 절차 개요

본 연구의 분석 절차는 앞서 논의한 바와 같이, (a) MTMM 상관행렬을 "수렴·변별 단서 제공을 위한 진단 도구"로 활용하고, (b) CFA 기반 점수 구조 모형 비교를 통해 차원 효과와 방법효과의 상대적 기여를 구조적으로 검증하는 흐름으로 구성된다(Campbell & Fiske, 1959; Hoffman et al., 2011; Marsh et al., 2014).

분석은 크게 세 단계로 수행하였다. 첫째, 기초 통계(평균, 표준편차, 왜도, 첨도)를 산출하여 자료의 분포 특성을 점검하였다. 둘째, 10개 관찰 지표 간 상관행렬을 산출하고, MTMM 논리에 따라 상관을 유형별(동일 역량-이질 과제 vs 이질 역량-동일 과제 등)로 분류하여 수렴·변별 단서를 기술적으로 검토하였다. 셋째, 동일 자료에 대해 다수의 CFA 기반 점수 구조 모형을 적합시켜 모형 적합도와 파라미터 추정을 비교함으로써, 자료가 차원모형, 과제모형, 차원-과제 결합모형 중 어떤 가설에 더 부합하는지 검증하였다.

MTMM 상관행렬 분석

MTMM 분석에서는 10개 관찰 지표(과제×역량 셀 점수) 간 피어슨 상관을 산출하였다. 이후 상관을 MTMM의 기본 논리에 따라 다음과 같이 분류하여 해석하였다. 첫째, 동일 역량-이질 과제 상관(monotrait-heteromethod)은 동일한 역량이 서로 다른 과제에서 측정된 두 지표 간 상관(예: 역량1의 I1-D1)이다. 이는 AC 맥락에서 "동일 역량이 과제 간 얼마나 일관되게 나타나는지"에 대한 수렴 단서를 제공한다. 둘째, 이질 역량-동일 과제 상관(heterotrait-monomethod)은 동일 과제 내에서 서로 다른 역량 지표들 간 상관(예: 과제 R 내 R2-R3)이다. 이는 방법효과 등으로 인해 역량 간 변별이 약화되는 양상을 시사할 수 있다. 셋째, 이질 역량-이질 과제 상관(heterotrait-heteromethod)은 서로 다른 역량이 서로 다른 과제에서 측정된 지표들 간 상관이다. 이는 전반적 수행 수준의 공유 또는 일반적 수행 요인을 시사할 수 있다.

본 연구에서 MTMM 결과는 '신뢰도 지표'로 해석하기보다는, 앞서 논의한 한계(역량×상황 결합성, 분산원 분리의 어려움)를 고려하여 점수 구조에 대한 초기 진단 및 CFA 결과 해석의 보조 근거로 활용하였다(Lance et al., 2007; Marsh et al., 2014).

CFA 기반 점수 구조 모형 비교

CFA는 AMOS 23을 사용하여 최우도추정 방식으로 수행하였다. 모형 적합도 평가는 카이제곱(χ^2)과 함께 CFI, TLI, RMSEA, SRMR 지수를 종합적으로 고려하였다(Brown, 2015; Hu & Bentler, 1999).

본 연구는 앞서 정리한 점수 구조 논쟁의 핵심을 반영하여 다음 모형들을 비교하였다. 첫째, 차원모형(dimension-only model)은 5개 역량을 잠재요인으로 설정하고, 각 역량요인에 해당 역량의 두 관찰 지표(서로 다른 과제에서 측정된 점수)가 로딩하도록 설정하였다. 둘째, 과제모형(exercise-only model)은 3개 과제를 잠재요인으로 설정하고, 각 과제요인에 해당 과제에서 산출된 여러 역량 지표들이 로딩하도록 설정하였다. 셋째, 차원-과제 결합모형(dimension-exercise hybrid model)은 역량요인과 과제요인을 동시에 포함하고, 각 관찰 지표가 해당 역량요인과 해당 과제요인에 동시에 로딩하도록 설정하였다(Hoffman et al., 2011).

넷째, CT-CU 모형(correlated trait-correlated uniqueness)은 5개 역량요인을 설정한 후, 동일 과제에서 측정된 지표들의 잔차를 상관시키는 방식으로 방법 관련 분산을 반영하였다(Geiser et al., 2008). 이는 방법요인을 직접 세우지 않더라도 동일 방법에서 공유되는 특수성을 모델링할 수 있다는 장점이 있다. 마지막으로, CT-C(M-1) 모형(correlated trait-correlated method minus one)은 하나의 과제를 기준방법(reference method)으로 설정하고, 나머지 과제의 방법효과를 기준방법 대비 편차로 추정하는 방식이다(Eid et al., 2003). 본 연구에서는 Structured Interview 과제를 기준방법으로 설정하였다. Structured Interview는 면접관이 구조화된 질문을 제시하고 피평가자가 자신의 경험이나 판단을 언어적으로 진술하는 형식의 과제인 반면, Role Play와 Group Discussion은 주어진 상황에서 타인과 실시간으로 상호작용하며 행동을 수행하는 시뮬레이션 과제라는 점에서 과제 구조가 상이하다. 따라서 Structured Interview를 기준방법으로 설정하면, 나머지 두 과제에서 나타나는 방법효과를 시뮬레이션 수행 맥락에 따른 상대적 편차로 보다 해석 가능하게 제시할 수 있다고 보았다. CT-C(M-1) 모형은 결합모형에서 빈번히 제기되는 식별 문제 및 해석 불안정성을 완화할 수 있는 대안으로 제시되어 왔다(Marsh et al., 2014).

결 과

MTMM 상관행렬 결과

10개 관찰지표 간 상관행렬은 표 2에 제시

표 2. 주요 변인 간 상관관계 및 기술통계

		1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1.	과제I - 역량1	-									
2.	과제I - 역량4	.65**	-								
3.	과제I - 역량5	.64**	.74**	-							
4.	과제R - 역량2	.27**	.32**	.28**	-						
5.	과제R - 역량3	.31**	.37**	.30**	.82**	-					
6.	과제R - 역량4	.28**	.30**	.22**	.66**	.75**	-				
7.	과제R - 역량5	.28**	.32**	.28**	.66**	.76**	.82**	-			
8.	과제D - 역량1	.23**	.23**	.23**	.36**	.36**	.26**	.27**	-		
9.	과제D - 역량2	.22**	.21**	.25**	.34**	.35**	.22**	.23**	.81**	-	
10.	과제D - 역량3	.25**	.27**	.28**	.36**	.35**	.28**	.26**	.81**	.81**	-
	M	2.79	2.64	2.68	2.48	2.55	2.49	2.54	2.55	2.49	2.59
	SD	0.32	0.30	0.28	0.39	0.40	0.37	0.37	0.38	0.39	0.40

주. $^{**}p < .01$, 역량1 = 문제해결, 역량2 = 기획·관리, 역량3 = 의사소통, 역량4 = 협업, 역량5 = 구성원 지원, 과제I = Structured Interview, 과제R = Role Play, 과제D = Group Discussion

하였다. MTMM 해석 논리에 따라 본 연구는 상관을 (1) 동일 역량-이질 과제(수렴/차원효과), (2) 이질 역량-동일 과제(변별/방법효과), (3) 이질 역량-이질 과제로 구분하여 기술적으로 검토하였다. 이하에서 I, R, D는 각각 과제 I, 과제 R, 과제 D를 의미하며, 숫자 1~5는 익명화된 역량을 의미한다. 예를 들어 I1은 과제 I에서 측정된 역량 1, D1은 과제 D에서 측정된 역량 1을 의미한다.

첫째, 동일 역량-이질 과제 상관은 .23~.35 범위에서 관찰되었다. 역량 1의 과제 간 상관(I1-D1)은 .23이었고, 역량 2와 역량 3의 과제 간 상관(R2-D2, R3-D3)은 각각 .34, .35로 나타났다. 또한 역량 4와 역량 5의 과제 간 상관(I4-R4, I5-R5)은 각각 .30, .28로 나타났다. 이는 동일 역량이 과제 간 완전히 독립적인 상황에서도 일정 수준의 일관성을 보인다는 점을 시사하나, 상관의 크기 자체는 높은 편이라 보기 어렵다. AC에서 과제는 단순한 '방법'이 아니라 서로 다른 행동표본과 과업 요구를 포함하는 복합 자극이므로, 동일 역량-이질 과제 상관이 높지 않다는 사실만으로 수렴타당도의 결여를 단정하기는 어렵다. 다만 동일 역량-이질 과제 상관이 이후에 제시할 동일 과제 내 이질 역량 상관보다 현저히 낮게 나타난다는 점은, 점수 구조에서 방법 관련 분산이 상당한 비중으로 작동할 가능성을 시사한다.

둘째, 이질 역량-동일 과제 상관은 높은 수준으로 관찰되었다. 과제 R 내에서는 역량 2와 역량 3 간 상관(R2-R3)이 .82, 역량 4와 역량 5 간 상관(R4-R5)이 .82, 역량 3과 역량 5 간 상관(R3-R5)이 .76로 나타났다. 과제 D 내에서도 역량 1과 역량 2 간 상관(D1-D2)이 .81, 역량 1과 역량 3 간 상관(D1-D3)이 .81, 역량 2와 역량 3 간 상관(D2-D3)이 .81로 매우 높았다. 과제 I 내 상관 역시 역량 1과 역량 4 간 상관(I1-I4)이 .65, 역량 4와 역량 5 간 상관(I4-I5)이 .74, 역량 1과 역량 5 간 상관(I1-I5)이 .64 수준으로 높게 나타났다. 이러한 패턴은 동일 과제 상황에서 역량 간 변별이 충분히 확보되지 않을 가능성, 또는 과제 특유 수행/전반적 인상(halo) 등 방법 관련 요인이 점수 구조에 강하게 기여할 가능성을 시사한다.

셋째, 이질 역량-이질 과제 상관은 .21~.37 범위에서 관찰되었다. 흥미롭게도 일부 이질 역량-이질 과제 상관은 동일 역량-이질 과제 상관과 유사한 수준을 보이기도 하였는데, 예를 들어 과제 I의 역량 4와 과제 R의 역량 3 간 상관(I4-R3)은 .37이었고, 과제 R의 역량 2와 과제 D의 역량 1 간 상관(R2-D1)은 .36으로 나타났다. 이는 MTMM 관점에서 이상적으로 기대되는 "동일 역량-이질 과제 상관이 다른 상관 유형보다 두드러지게 높음"이라는 패턴이 명확히 나타나지 않았음을 의미하며, 이와 같은 경향은 방법효과뿐 아니라 과제와 차원을 넘어 비교적 공통적으로 작동하는 피평가자의 전반적 수행특성 또는 일반수행요인이 여러 점수에 반영되었을 가능성과도 관련될 수 있다. 다만 MTMM 상관행렬만으로는 이러한 공통 분산의 원천을 직접 분리·확인하기 어렵다. 점수 구조 해석에서는 차원효과의 존재 여부를 단선적으로 결론내리기보다, 이후 CFA 기반 모형 비교를 통해 차원효과와 방법효과의 상대적 기여를 구조적으로 점검할 필요가 있음을 뒷받침한다.

CFA 기반 점수 구조 모형 비교 결과

동일 자료에 대해 차원모형, 과제모형, 차원

표 3. 측정모형의 확인적 요인분석 결과

	χ^2	df	χ^2/df	CFI	TLI	RMSEA	SRMR
차원 모형	978.25	25	39.13	.645	.361	.332	.187
과제 모형	133.10	32	4.16	.962	.947	.096	.033
결합 모형	21.98	13	1.69	.997	.988	.045	.046
결합 모형 CT-CU	13.53	13	1.04	1.00	.999	.011	.016
결합 모형 CT-C(M-1)	29.46	18	1.64	.996	.989	.043	.030

주. N = 324, CFI = Comparative fit index, TLI = Tucker-Lewis index, RMSEA = Root-mean square error of approximation, SRMR = Standardized root-mean square residual.

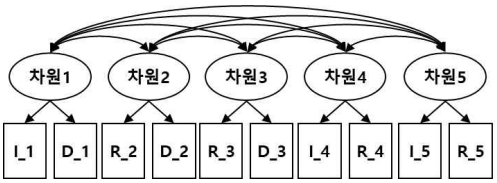


그림 1. 차원 모형

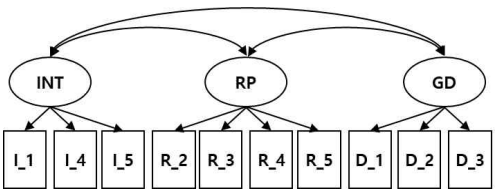


그림 2. 과제 모형

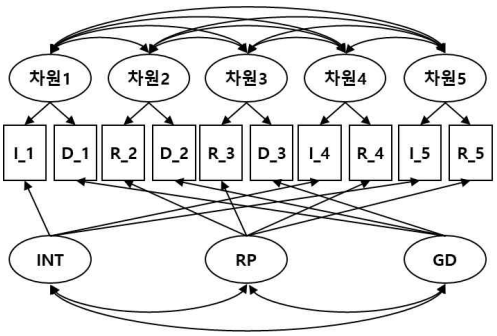


그림 3. 결합 모형

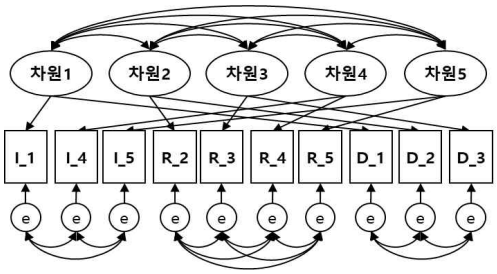


그림 4. 결합 모형 CT-CU

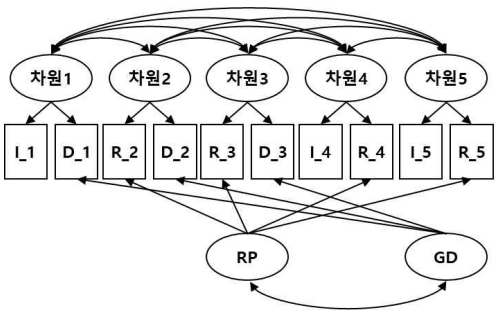


그림 5. 결합 모형 CT-C(M-1)

-과제 결합모형, CT-CU 모형, CT-C(M-1) 모형을 적합시켜 점수 구조 가설의 상대적 타당성을 검토하였다. 각 모형의 적합도 지수는 표 3에 제시하였다.

첫째, 차원모형(dimension-only model)의 경우

적합도가 매우 저조하였다(χ^2=978.25, df=25, χ^2/df=39.13, CFI=.645, TLI=.361, RMSEA=.332, SRMR=.187). 이는 본 자료의 공분산 구조를 "역량 차원 중심"으로만 설명하기에는 한계가 크며, 방법 관련 분산을 고려하지 않은 단순 차원 구조는 자료 적합을 충분히 확보하지 못함을 의미한다.

둘째, 과제모형(exercise-only model)은 차원 모형 대비 상당히 개선된 적합도를 보였으나(χ^2 =133.10, df=32, χ^2/df=4.16, CFI=.962, TLI=.947, RMSEA=.096, SRMR=.033), 일반적 기준에서 완전히 만족스러운 수준으로 보기에는 다소 제한이 있다. 그럼에도 과제모형이 차원모형보다 현저히 우수한 적합을 보였다는 점은, MTMM 상관행렬에서 관찰된 "동일 과제 내 상관의 과도한 상승" 패턴과 일관되게, 점수 구조에서 방법 관련 요인의 기여가 크다는 해석을 뒷받침한다.

셋째, 차원-과제 결합모형(dimension-exercise hybrid (mixed) model)은 전반적으로 양호한 적합도를 보였다(χ^2=21.98, df=13, χ^2/ df=1.69, CFI=.997, TLI=.988, RMSEA=.045, SRMR=.046). 이는 점수 구조를 차원효과 또는 과제효과 "하나"로만 환원하기보다, 두 효과가 함께 존재하는 결합 구조로 이해하는 것이 본 자료에 더 부합할 가능성을 시사한다.

넷째, CT-CU 모형은 본 비교 모형 중 가장 우수한 적합도를 나타냈다(χ^2=13.53, df=13, χ^2/df=1.04, CFI=1.000, TLI=.999, RMSEA=.011, SRMR=.016). CT-CU는 동일 과제에서 측정된 지표들의 잔차를 상관시키는 방식으로 방법 관련 분산을 반영하므로, 방법효과를 '방법요인'으로 직접 세우지 않더라도 공분산 구조에서 방법 관련 공유성을 효과적으로 흡수할 수 있다. 본 자료에서 CT-CU의 적합도가 가장 우수했다는 점은, 점수 구조에서 방법 관련 공유성이 상당히 크며, 이를 적절히 모델링할 때 자료 설명력이 크게 향상될 수 있음을 시사한다.

마지막으로, CT-C(M-1) 모형은 매우 우수한 적합도를 보였다(χ^2=29.46, df=18, χ^2/df=1.64, CFI=.996, TLI=.989, RMSEA=.043, SRMR=.030). CT-C(M-1)은 하나의 기준 방법을 설정하고 나머지 방법효과를 상대적 편차로 추정하는 방식이므로, 방법효과를 보다 해석 가능한 형태로 제시하면서도 식별 및 추정 불안정을 완화할 수 있다는 점에서 실무 자료에도 적용 가능성이 높은 대안이다.

추가로, 결합모형의 세부 모수를 검토하기 위해 지표별 차원요인 및 방법요인 부하량을 표 4에 제시하였다. 차원요인의 표준화 요인부하량은 .362~.907 범위로 나타났으며, 일부 역할수행 지표(R2=.905, R3=.907, R4=.786, R5=.849)에서는 비교적 높은 부하량이 관찰되었다. 반면 인터뷰 및 집단토론에서 측정된 차원지표들(I1=.372, I4=.404, I5=.362, D1=.458, D2=.413, D3=.438)은 상대적으로 낮거나 중간 수준의 부하량을 보였다. 방법요인의 경우, 인터뷰(I1=.659, I4=.759, I5=.795)와 집단토론(D1=.802, D2=.812, D3=.789)은 전반적으로 비교적 높은 표준화 부하량과 유의한 C.R. 값을 나타냈다. 반면 역할수행 방법요인은 표준화 부하량이 -.092~.658 범위로 나타났고, 관련 경로의 C.R. 값은 모두 유의하지 않았다. 이러한 결과는 방법효과의 상대적 크기가 모든 기법에서 균등하게 나타나는 것이 아니라, 인터뷰와 집단토론에서는 공통된 방법 관련 분산이 비교적 뚜렷하게 나타난 반면 역할수행에서는 단일한 방법요인으로 안정적으로 추정되지 않았을 가능성을 시사한다. 동시에 일

표 4. 결합모형의 지표별 차원요인 및 방법요인 부하량

지표	요인유형	요인	표준화 요인부하량	비표준화 계수	S.E.	C.R.
I1	차원	1	.372	1.000		
	방법	I	.659	0.210	0.015	13.566***
D1	차원	1	.458	1.481	0.383	3.862***
	방법	D	.802	0.307	0.019	15.935***
R2	차원	2	.905	1.000		
	방법	R	-.092	-0.036	0.221	-0.162
D2	차원	2	.413	0.457	0.067	6.805***
	방법	D	.812	0.315	0.019	16.542***
R3	차원	3	.907	1.000		
	방법	R	.027	0.011	0.227	0.048
D3	차원	3	.438	0.478	0.089	5.360***
	방법	D	.789	0.315	0.021	15.179***
I4	차원	4	.404	1.000		
	방법	I	.759	0.225	0.014	16.148***
R4	차원	4	.786	2.433	0.570	4.272***
	방법	R	.305	0.113	0.192	0.588
I5	차원	5	.362	1.000		
	방법	I	.795	0.224	0.014	16.514***
R5	차원	5	.849	3.080	1.457	2.114*
	방법	R	.658	0.244	0.195	1.250

주. 지표 표기에서 I, D, R는 각각 Structured Interview, Group Discussion, Role Play를 의미하고, 숫자(1~5)는 해당 역량을 의미한다. 예를 들어 I1은 Structured Interview에서 측정한 역량 1, R2는 Role Play에서 측정한 역량 2를 의미한다. $^{*}p < .05$, $^{***}p < .001$.

부 역할수행 지표에서 차원요인 부하량이 상대적으로 높게 나타난 점은, 결합모형 내 차원효과와 방법효과의 상대적 기여가 과제와 지표에 따라 상이할 수 있음을 보여준다.

종합하면, 본 연구의 CFA 결과와 결합모형의 세부 부하량은 차원모형이 자료를 충분히 설명하지 못한 반면, 과제모형은 상대적으로 양호한 적합을 보였고, 차원과 방법효과를 함께 고려한 결합모형 및 변형 모형(CT-CU, CT-C(M-1))이 가장 우수한 적합도를 나타냈음을 보여준다. 특히 MTMM 상관행렬에서 동일 과제 내 상관이 높게 나타난 패턴과 결합모형

의 세부 부하량을 함께 고려할 때, 본 자료의 점수 구조는 차원효과 또는 과제효과 중 하나로 환원되기보다 차원효과와 방법효과가 함께 작동하는 결합 구조로 이해하는 것이 보다 타당하다. 다만 방법효과의 상대적 크기와 안정성은 기법별로 동일하지 않았으며, 이러한 점은 AC 점수 구조를 해석할 때 기법별 특수성을 함께 고려할 필요가 있음을 시사한다.

논 의

본 연구는 국내 A기업 책임급 승진 평가센터(AC) 자료를 활용하여, 중복측정 설계에서 산출되는 과제×역량 셀 점수의 공분산 구조를 MTMM 상관행렬과 CFA 기반 점수 구조 모형 비교를 통해 분석하였다. MTMM 결과에서는 동일 과제 내 이질 역량 간 상관(heterotrait-monomethod)이 전반적으로 높게 관찰된 반면, 동일 역량의 과제 간 상관(monotrait-heteromethod)은 그에 비해 낮은 수준으로 나타나, MTMM 관점에서 이상적으로 기대되는 “동일 역량-이질 과제 상관이 다른 상관 유형보다 두드러지게 높음”이라는 패턴이 명확히 재현되지 않았다(Campbell & Fiske, 1959). 이는 본 자료에서 AC 점수 구조가 단순한 ‘역량(차원) 중심’으로만 조직되기보다는, 방법 특수성이 점수 변동에 유의미하게 기여할 가능성을 시사한다.

이러한 상관 패턴은 MTMM만으로 차원효과의 존재 여부를 일도양단으로 판단하기 어렵다는 점을 보여주며(Lance et al., 2007), 점수 구조 해석의 타당성을 높이기 위해서는 잠재변인 모형에 기반한 다중 접근이 요구된다. 본 연구에서 수행한 CFA 기반 모형 비교(차원모형, 과제모형, 차원-과제 결합모형, CT-CU, CT-C(M-1))는 이러한 문제의식에 근거한 것으로, AC 점수 구조가 “차원효과와 방법효과가 함께 작동하는 결합 구조”로 이해될 여지가 큼을 확인하고자 한 시도이다(Hoffman et al., 2011; Marsh et al., 2014). 요컨대, 본 자료의 결과는 “AC가 역량을 전혀 측정하지 못한다”는 결론보다는, 역량(차원) 신호 위에 방법 신호가 비교적 크게 얹히는 방식으로 점수 구조가 형성될 수 있음을 보여주는 것으로 해석하는 것이 적절하다.

또한 결합모형의 세부 부하량을 함께 고려하면, 방법효과의 상대적 크기와 안정성은 본 연구에 포함된 기법들 간에 동일하지 않았다. 인터뷰와 집단토론에서는 비교적 뚜렷한 방법 관련 부하량이 관찰된 반면, 본 자료의 역할수행 과제에서는 방법요인이 단일한 공통요인으로 안정적으로 추정되지 않았다. 다만 이러한 결과를 근거로 역할수행이라는 기법 일반이 일관되게 약한 방법효과를 가진다고 해석하는 것은 신중할 필요가 있다. 역할수행은 측정하고자 하는 역량, 상호작용 상대, 과제의 맥락과 주제, 행동 요구의 구조에 따라 과제 특성이 크게 달라질 수 있기 때문이다. 따라서 본 연구의 결과는 역할수행이라는 기법 일반에 대한 결론이라기보다, 본 자료에서 사용된 역할수행 과제 구성에서는 단일한 공통 방법요인이 상대적으로 약하게 포착되었을 가능성을 보여주는 것으로 이해하는 것이 적절하다. 이러한 점은 AC의 방법효과가 기법명 자체에 의해 일률적으로 결정된다기보다, 각 기법 내 과제 설계와 측정 역량의 구성 방식에 따라 다르게 나타날 수 있음을 시사한다.

아울러 본 연구에서 확인된 점수 구조는 과제 자체의 특성뿐 아니라 실제 AC 운영 과정

에서 이루어지는 점수 통합 절차의 영향도 함께 반영한 결과일 가능성이 있다. 특히 통합회의는 개별 평가자의 관찰과 판단을 조정하여 최종 점수를 산출하는 과정이라는 점에서, 특정 과제에서 형성된 수행 인상이 여러 차원 점수에 공통적으로 유지 · 강화되거나 반대로 조정될 수 있다. 또한 본 AC는 과제당 1명의 평가자가 1차 평정을 수행하는 구조이므로, 본 연구에서 관찰된 방법효과에는 과제 특성뿐 아니라 평가자 관련 분산이 함께 반영되었을 가능성도 있다. 따라서 본 연구에서 나타난 방법효과는 과제 설계의 영향과 더불어 통합회의를 포함한 운영 절차의 영향 가능성까지 함께 고려하여 해석될 필요가 있다. 실무적으로는 과제-역량 매트릭스의 정교화뿐 아니라, 통합회의에서 차원별 행동증거를 명시적으로 검토하고 전반적 인상보다 행동 근거를 중심으로 점수를 통합하는 절차를 구조화할 필요가 있다. 연구결과를 바탕으로 도출된 학술적 시사점은 다음과 같다.

첫째, 본 연구는 AC 구성개념 타당도 해석에서 MTMM 단독 접근의 한계를 국내 기업 실제 자료로 재확인하고, MTMM과 CFA를 병행하는 다중 접근의 필요성을 구체적으로 제시한다. MTMM 상관행렬은 수렴 · 변별 타당도 단서를 직관적으로 제공하는 장점이 있으나, 차원효과와 방법효과를 분산 수준에서 명시적으로 분리하지 못하며, 상관 패턴의 해석이 과제 설계 · 평정 절차 · 자극 특성 등 다수 요인에 의해 영향을 받는다는 제약이 존재한다(Lance et al., 2007). 따라서 MTMM 결과는 “구성개념 타당도의 결론”이라기보다 점수 구조의 문제를 진단하는 출발점으로 활용하고, 이후 잠재변인 모형 기반 점수 구조 분석과 함께 종합적으로 해석하는 것이 타당도 논증의 설득력을 높인다(Brown, 2015; Marsh et al., 2014).

둘째, 본 연구의 결과는 왜 AC 점수 구조가 차원-과제 결합 구조로 관찰되는지에 대해, “역량이 상황과 결합하여 행동으로 발현된다”는 관점에서 해석 가능함을 시사한다. 성격 및 행동 연구 전통에서 인간의 행동은 개인의 특성과 상황의 상호작용으로 설명될 수 있다는 명제가 반복적으로 논의되어 왔으며(예: Lewin의 B=f(P,E) 관점; Mischel, 1968; Mischel & Shoda, 1995), AC는 본질적으로 “상황화된 행동 표본”을 수집하는 평가체계라는 점에서 이러한 관점과 구조적으로 맞닿아 있다(Hoffman et al., 2011; Jackson et al., 2016). 예컨대 ‘의사소통’이라는 동일 역량이라 하더라도, 상사 · 경영진을 대상으로 제한 시간 내 논리를 구성해 전달하는 발표(PT) 상황과, 부하직원과의 상호작용에서 경청 · 질문 · 정서적 조율이 요구되는 면담/역할수행 상황은 행동 요구가 다를 수 있다. 이때 과제 간 점수의 완전한 일치를 기대하기보다, 어떤 역량이 어떤 상황에서 일관되게 나타나고(수렴), 어떤 상황에서는 차이를 보이는지(상황특수성)를 체계적으로 파악하는 것이 AC의 중요한 가치가 될 수 있다. 즉, 결합 구조의 관점은 차원효과의 약화를 ‘결함’으로만 보지 않고, 역량이 상황과 결합해 나타나는 행동의 현실성을 반영한 결과로 해석할 여지를 제공한다.

셋째, 국내 연구 맥락에서 본 연구는 동일한 중복측정 AC 자료에 MTMM과 CFA를 동시에 적용하고 그 결과를 직접 비교하려는 시도를 명시적으로 수행한다는 점에서 의의가 있다. 기존 국내 연구에서도 AC 점수 구조를 차원효과와 방법효과 관점에서 점검하려는 시도가 일부 이루어진 바 있다. 예를 들어 허창구

와 신강현(2010)은 일반화가능도 이론을 통해 분산원을 추정하였고, 확인적 요인분석에서는 평가과제모형이나 평가차원모형보다 평가과제와 평가차원을 함께 고려한 혼합모형의 적합도가 더 우수함을 보고하였다. 또한 허창구(2011)는 MTMM 상관 패턴을 통해 AC 점수 구조에서 차원효과와 방법효과가 함께 작동할 가능성을 검토하였다. 이러한 점에서 국내 선행연구들은 AC 점수 구조를 단일 차원효과 또는 단일 과제효과로 환원하기보다, 두 효과를 함께 고려하는 결합적 해석이 필요함을 시사한다는 점에서 본 연구와 맥을 같이한다. 다만 허창구와 신강현(2010)의 연구는 동일 역량이 여러 과제에서 반복 측정되지 않는 내재 설계 자료를 사용하였고, 허창구(2011) 역시 본 연구처럼 동일 자료에서 MTMM 상관 기반 해석과 CFA 기반 점수 구조 모형 비교를 병행한 연구는 아니라는 점에서, 두 연구를 본 연구와 일대일로 대응시켜 비교하기에는 한계가 있다. 따라서 본 연구의 차별성은 중복측정 설계를 갖춘 동일 자료에 MTMM과 CFA를 병행 적용하여 두 접근이 제공하는 점수 구조 해석을 직접 비교·보완하였다는 데 있으며, 이는 국내 기업 AC 운영 자료에서 다중 접근에 기반한 실증적 근거를 제시한다는 점에서 학술적 기여를 갖는다.

실무적 시사점은 다음과 같다. 첫째, 실무에서 “결합 구조”에 대한 이해는 과제 설계와 역량 매트릭스 구성의 핵심 전제가 되어야 한다. 점수 구조가 결합적으로 나타날 가능성이 높은 상황에서, 단순히 ‘일반적인 역량 행동’만을 상정해 과제를 개발하면, 과제 특수성에 의해 역량 간 변별이 약화되거나(동일 과제 내 상관 증가), 특정 과제에서 유리한 행동양식이 여러 역량 점수에 광범위하게 투사되는 문제가 강화될 수 있다. 따라서 각 과제는 “그 과제 맥락에서 그 역량이 어떻게 행동으로 구현되는지”를 전제로 설계되어야 하며, 매트릭스는 과제별 행동 요구와 역량 정의의 대응관계를 명확히 드러내는 방식으로 구성될 필요가 있다(International Task Force on Assessment Center Guidelines, 2009; International Taskforce on Assessment Center Guidelines, 2015). 나아가 본 연구의 세부 부하량 결과는 동일한 기법 범주 안에서도 과제 설계에 따라 방법효과의 크기와 형태가 달라질 수 있음을 시사한다. 따라서 실무에서는 인터뷰, 집단토론, 역할수행과 같은 기법명을 기준으로 일률적인 가정을 하기보다, 각 과제가 실제로 어떤 행동 표본을 요구하고 어떤 역량을 얼마나 중첩적으로 자극하는지를 점검하면서 과제별 역량 수와 행동지표의 배치를 조정할 필요가 있다. 특히 실무에서 AC 간소화가 진행될 때, 1개 과제에서 4개 이상 역량을 과도하게 측정하도록 “구겨 넣는” 방식은 과제 내 공통인상(또는 과제 특수성)이 여러 차원 점수에 동시 반영될 위험을 키울 수 있으므로 주의가 필요하다.

둘째, 차원효과를 확보하기 위한 평정의 질 관리(특히 FoR 관점의 기준 공유와 눈높이 정렬)는 여전히 중요하다. 다만 본 연구의 논의는 “평가자 훈련만 강화하면 차원 변별이 자동으로 확보된다”는 단순 처방이라기보다, 과제-역량 설계의 정교화, 관찰-기록-분류-평정의 절차 준수, 평가자 간 공통 준거(frame of reference)의 확보가 함께 작동해야 한다는 통합적 관점에 가깝다(International Task Force on Assessment Center Guidelines, 2009; International Taskforce on Assessment Center Guidelines, 2015). 실제 실무에서는 회차별 교육이 운영 중심으로 축소되는 경우가 많고, 비디오 캘리브레이

션과 같은 훈련 요소가 생략되기도 한다. 이러한 현실을 고려할 때, 최소한 과제별 핵심 행동지표의 해석을 정렬하고, “어떤 행동이 어느 수준의 평정으로 연결되는지”에 대한 공통 준거를 유지 · 갱신하는 장치가 필요하다. 다만 선행연구에서 평가자 훈련이 점수의 질에 기여하더라도 차원 변별을 완전히 해결하지 못하는 경우가 보고되어 왔다는 점을 고려하면, 실무적 개선은 평가자 훈련 단일 요인에 과도하게 귀속되기보다 설계 · 평정절차 · 운영 전반의 조합으로 논의될 필요가 있다(Hoffman et al., 2011; Lance, 2008).

셋째, 결과 활용 측면에서 OAR(종합평정) 중심의 단선적 활용을 넘어, 차원×과제 점수의 활용 전략을 적극적으로 강구할 필요가 있다. 특히 개발 · 육성 목적(교육/코칭/피드백)에서는 ‘전체적으로 잘함/못함’이라는 단일 지표만으로는 행동 변화의 방향성을 제공하기 어렵다. 결합 구조의 관점에서 보면, 동일 역량이라도 과제(상황)에 따라 강점 · 개선점이 달라질 수 있으므로, “어떤 상황에서 어떤 역량 행동이 강/약하게 나타났는지”를 피드백의 핵심 단위로 삼는 것이 실무적으로 더 유용할 수 있다. 이는 AC가 단지 선발 · 승진의 ‘선별 도구’가 아니라, 상황화된 행동 정보를 제공하는 ‘진단 도구’로 기능할 수 있음을 의미한다. 아울러 OAR을 별도로 평정하는 운영 관행이 존재하는 경우, OAR을 먼저 형성한 뒤 차원 점수를 그에 맞춰 정렬하는 방식은 전반적 인상(halo)의 영향을 강화할 여지가 있다. 가이드라인이 강조하는 관찰-기록-분류-평정의 절차 준수 원칙을 고려할 때, 가능한 한 차원별 행동근거에 기반해 차원 점수를 형성하고, OAR은 차원 점수의 종합(또는 통합회의에서의 종합적 판단)으로 후행적으로 다루는 방식이 점수 해석의 정합성을 높이는 데 유리할 수 있다(International Task Force on Assessment Center Guidelines, 2009; International Taskforce on Assessment Center Guidelines, 2015). 이는 “OAR을 쓰지 말자”가 아니라, OAR이 차원 평정의 근거를 잠식하지 않도록 운영 흐름을 점검할 필요가 있음을 뜻한다.

넷째, 국내 실무 환경에서 AC 간소화 압력이 커지는 현실을 고려하면, “간소화하더라도 중복측정은 유지되어야 한다”는 시사점이 중요하다. 중복측정 설계는 동일 역량이 서로 다른 상황에서 어떻게 나타나는지에 대한 정보를 제공하며, 이는 수렴 · 변별 단서 및 점수 구조 분석 가능성을 실질적으로 보장한다. 반대로 내재설계(각 역량 1회 측정)로의 축소는 운영 효율을 높일 수 있으나, 차원점수 해석의 근거를 약화시키고 과제특수성의 영향이 상대적으로 과대 반영될 가능성이 있다. 따라서 현실적 제약 속에서도 최소한의 중복측정(역량당 2회 이상)을 유지하는 설계 원칙이 논의될 필요가 있다.

본 연구가 갖는 제한점과 추후 연구를 위한 제언은 다음과 같다. 첫째, 본 연구는 단일 기업 · 단일 직급(책임급) 맥락의 자료에 기반하므로, 다른 조직 · 직무 · 직급으로의 일반화에는 제한이 있다. 향후에는 공공부문 및 다양한 산업군, 상이한 직무군을 대상으로 점수 구조가 재현되는지 비교할 필요가 있다.

둘째, 본 연구는 구성개념 타당도(점수 구조) 논증에 초점을 두었으며, 자료 제약상 신뢰도 추정(예: 검사-재검사, 평가자 간 일치도, 일반화가능도 계수)을 본 분석의 중심 결과로 제시하지 않았다. 또한 본 연구의 분석 자료는 통합회의를 거쳐 최종 확정된 점수이므로, 본 연구에서 확인된 점수 구조는 과제 자체의

특성뿐 아니라 점수 통합 절차의 영향 가능성도 함께 반영한 결과일 수 있다. 그러나 본 연구 자료에는 통합회의 전후 점수 변화, 점수 조정 비율, 평가자 간 일치도와 같은 정보가 포함되어 있지 않아, 통합회의가 방법효과와 차원효과의 해석에 미친 영향을 직접 검증하지는 못하였다. 향후 연구에서는 운영 설계가 허용되는 범위에서 재응시 자료, 평가자 교차 설계, 또는 일반화가능도 이론 기반 분산원 추정을 통해 점수의 일관성(신뢰도) 측면을 보완적으로 검토할 필요가 있다. 아울러 개별 평가자의 최초 평정점수와 통합회의 이후의 최종 점수를 함께 수집하고, 점수 조정 양상 및 평가자 간 일치도 지표를 분석함으로써 통합회의가 점수 구조에 미치는 영향을 보다 정교하게 검토할 필요가 있다. 또한, 서론에서 제기한 전반적 수행요인(g) 가능성은 평가센터 점수 해석에서 중요한 논점이지만, 본 연구는 동일 자료에 대해 차원효과와 과제/방법효과의 상대적 기여를 비교·검토하는 목적에 맞춰 차원모형, 과제모형, 그리고 차원-과제 결합모형(및 관련 변형)을 중심으로 경쟁모형을 구성하였다. 또한 본 자료의 지표 수와 모형 식별 제약을 고려할 때, 상위요인(g) 또는 바이팩터 계열 모형은 직접 비교 가능한 경쟁모형으로 포함하기에 제한이 있어 후속 연구에서 확장적으로 다루고자 한다. 한편 본 연구에서 제시한 결합모형의 세부 부하량은 기법별로 상이한 양상을 보였으며, 일부 방법요인은 다른 기법에 비해 안정적으로 추정되지 않았다. 따라서 이러한 세부 모수는 특정 기법 일반에 대한 결론으로 곧바로 일반화하기보다, 본 자료에서 사용된 과제 구성과 측정 역량의 조합, 그리고 운영 맥락 속에서 해석할 필요가 있다. 향후 연구에서는 동일한 기법 범주 내에서도 과제 설계의 차이에 따라 방법효과의 크기와 형태가 어떻게 달라지는지를 보다 체계적으로 비교할 필요가 있다.

셋째, 결합 구조의 해석을 더 강화하기 위해서는 AC 내부의 과제 간 상관 구조를 넘어 외부 준거와의 관계를 확장하는 연구가 요구된다. 예컨대 다면평가, 상사평정, 성과지표 등 외부 도구와의 연결을 통해 외적 구성개념 타당도(차원 대응성)를 점검하거나(Wirz et al., 2020), 차원모형/과제모형/결합모형에서 산출되는 점수(차원점수, 과제점수, OAR)가 준거를 얼마나 잘 예측하는지(준거관련 타당도, 증분 타당도)를 비교하는 연구가 필요하다(Meriac et al., 2008). 특히 “어떤 점수(차원, 과제, OAR)가 어떤 준거에 더 강하게 연결되는지”는 결합 구조의 실질적 의미를 판단하는 핵심 근거가 될 수 있다.

본 연구는 국내 기업 책임급 승진 AC 자료에서 MTMM과 CFA 기반 점수 구조 분석을 병행하여, AC 점수 구조가 단순 차원 중심이라기보다 차원효과와 방법효과가 함께 작동하는 결합적 성격을 가질 수 있음을 시사하였다. 이는 MTMM 결과를 단독으로 해석하기보다 잠재변인 모형 기반 접근과 결합한 다중 방법론적 논증이 필요함을 보여주며, 실무적으로는 과제-역량 매트릭스 설계의 정교화, 중복측정 유지, 평정의 질 관리(FoR 및 절차 준수), 그리고 차원×과제 점수의 적극적 활용 전략이 중요함을 함의한다.

참고문헌

권용만, 정장호 (2021). 공공부문 역량평가제도의 활성화 방안에 대한 연구: 민간부분의

운영방식과의 비교 연구. 벤처혁신연구, 4(1), 51-65.
https://doi.org/10.22788/4.1.4

송지은 (2020). 평가센터(Assessment Center)를 활용한 직무전문가 선발: K사의 사례를 중심으로. 경영컨설팅연구, 20(3), 237-249.

유기원 (2014). 승진자 선발을 위한 평가센터의 개발과 활용: K공단 사례를 중심으로. 경영컨설팅연구, 14(2), 425-445.

이상준, 김미경 (2022). 공공부문 역량평가 현황분석과 정책적 함의: 최근 4년 조달청 사업공고 내용을 중심으로. 한국인사행정학회보, 21(3), 169-199.
https://doi.org/10.23243/kosppa.21.3.202209.169

허창구, 신강현 (2010). 내재설계 평가센터의 신뢰도 및 타당도. 한국심리학회지: 산업 및 조직, 23(2), 225-249.
https://doi.org/10.24230/ksiop.23.2.201005.225

허창구 (2011). 대학선발장면에서 평가센터의 신뢰도 및 구성개념 타당도 분석. 한국심리학회지: 산업 및 조직, 24(4), 785-807.
https://doi.org/10.24230/ksiop.24.4.201111.785

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* American Educational Research Association.

Arthur, W., Jr., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology, 56*(1), 125-154. https://doi.org/10.1111/j.1744-6570.2003.tb00146.x

Binning, J. F., & Barrett, G. V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology, 74*(3), 478-494. https://doi.org/10.1037/0021-9010.74.3.478

Bowler, M. C., & Woehr, D. J. (2006). A meta-analytic evaluation of the impact of dimension and exercise factors on assessment center ratings. *Journal of Applied Psychology, 91*(5), 1114-1124. https://doi.org/10.1037/0021-9010.91.5.1114

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81-105. https://doi.org/10.1037/h0046016

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281-302. https://doi.org/10.1037/h0040957

Eid, M., Lischetzke, T., Nussbeck, F. W., & Trierweiler, L. I. (2003). Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple-indicator CT-C(M-1) model. *Psychological Methods, 8*(1), 38-60. https://doi.org/10.1037/1082-989X.8.1.38

Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., III, & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology, 72*(3), 493-511. https://doi.org/10.1037/0021-9010.72.3.493

Geiser, C., Eid, M., & Nussbeck, F. W. (2008). On the meaning of the latent variables in the CT-C(M-1) model: A comment on Maydeu-

Olivares & Coffman (2006). *Psychological Methods, 13*(1), 49-57. https://doi.org/10.1037/1082-989X.13.1.49

Hoffman, B. J., Melchers, K. G., Blair, C. A., Kleinmann, M., & Ladd, R. T. (2011). Exercises and dimensions are the currency of assessment centers. *Personnel Psychology, 64*(2), 351-395. https://doi.org/10.1111/j.1744-6570.2011.01213.x

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: *A Multidisciplinary Journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

International Task Force on Assessment Center Guidelines. (2009). Guidelines and ethical considerations for assessment center operations. *International Journal of Selection and Assessment, 17*(3), 243-253. https://doi.org/10.1111/j.1468-2389.2009.00467.x

International Taskforce on Assessment Center Guidelines. (2015). Guidelines and ethical considerations for assessment center operations. *Journal of Management, 41*(4), 1244-1273. https://doi.org/10.1177/0149206314567780

Jackson, D. J. R., Michaelides, G., Dewberry, C., & Kim, Y. J. (2016). Everything that you have ever been told about assessment center ratings is confounded. *Journal of Applied Psychology, 101*(7), 976-994. https://doi.org/10.1037/apl0000102

Jackson, D. J. R., Michaelides, G., Dewberry, C., Nelson, J., & Stephens, C. (2022). Reliability in assessment centres depends on general and exercise performance, but not on dimensions. *Journal of Occupational and Organizational Psychology, 95*(4), 739-757. https://doi.org/10.1111/joop.12398

Kuncel, N. R., & Sackett, P. R. (2014). Resolving the assessment center construct validity problem as we know it. *Journal of Applied Psychology, 99*(1), 38-47. https://doi.org/10.1037/a0034147

Lance, C. E. (2008). Why assessment centers do not work the way they are supposed to. *Industrial and Organizational Psychology, 1*(1), 84-97. https://doi.org/10.1111/j.1754-9434.2007.00017.x

Lance, C. E., Woehr, D. J., & Meade, A. W. (2007). Case study: A Monte Carlo investigation of assessment center construct validity models. *Organizational Research Methods, 10*(3), 430-448. https://doi.org/10.1177/1094428106289395

Lievens, F., & Conway, J. M. (2001). Dimension and exercise variance in assessment center scores: A large-scale evaluation of multitrait-multimethod studies. *Journal of Applied Psychology, 86*(6), 1202-1222. https://doi.org/10.1037/0021-9010.86.6.1202

Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology, 10*, 85-110. https://doi.org/10.1146/annurev-clinpsy-032813-153700

Meriac, J. P., Hoffman, B. J., Woehr, D. J., &

Fleisher, M. S. (2008). Further evidence for the validity of assessment center dimensions: A meta-analysis of the incremental criterion-related validity of dimension ratings. *Journal of Applied Psychology, 93*(5), 1042-1052. https://doi.org/10.1037/0021-9010.93.5.1042

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749. https://doi.org/10.1037/0003-066X.50.9.741

Mischel, W. (1968). *Personality and assessment.* Wiley.

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review, 102*(2), 246-268. https://doi.org/10.1037/0033-295X.102.2.246

Putka, D. J., & Hoffman, B. J. (2013). Clarifying the contribution of assessee-, dimension-, exercise-, and assessor-related effects to reliable and unreliable variance in assessment center ratings. *Journal of Applied Psychology, 98*(1), 114-133. https://doi.org/10.1037/a0030887

Thornton, G. C., III, & Rupp, D. E. (2006). *Assessment centers in human resource management: Strategies for prediction, diagnosis, and development.* Lawrence Erlbaum Associates.

Wirz, A., Melchers, K. G., Kleinmann, M., Lievens, F., Annen, H., Blum, U., & Ingold, P. V. (2020). Do overall dimension ratings from assessment centres show external construct-related validity? *European Journal of Work and Organizational Psychology, 29*(3), 405-420. https://doi.org/10.1080/1359432X.2020.1714593

Woehr, D. J., & Arthur, W., Jr. (2003). The construct-related validity of assessment center ratings: A review and meta-analysis of the role of methodological factors. *Journal of Management, 29*(2), 231-258. https://doi.org/10.1177/014920630302900206

투고일자: 2026-03-09
수정일자: 2026-04-05
게재일자: 2026-04-17

한국심리학회지: 산업 및 조직
Korean Journal of Industrial and Organizational Psychology
2026. Vol. 39, No. 2, 317-342
https://doi.org/10.24230/kjiop.v39i2.317-342

Construct Validity of Assessment Centers: Comparing the MTMM Correlation Matrix and CFA Models

Lim Hee Chul
Department of Psychology
Hallym University

Ju Il Rie
Applied Psychology Institute
Hallym University

This study used promotion assessment center (AC) data from a large Korean corporation (Company A) for managerial-level promotion candidates (N = 324). We examined convergent and discriminant validity cues from multitrait-multimethod (MTMM) correlation patterns and compared CFA score-structure models (dimension, exercise, and dimension-exercise mixed) to assess the relative contributions of dimension and method effects. MTMM results showed high correlations among different dimensions within the same exercise but relatively low correlations for the same dimension across exercises, indicating plausible method effects. CFA results showed poor fit for the dimension model, marginally acceptable fit for the exercise model, and the best fit for the mixed model. These findings suggest that construct validity should not be inferred from MTMM patterns alone and support a multi-approach interpretation that combines MTMM and CFA model comparisons. From a competency × situation perspective, we recommend maintaining an overlapping measurement design in which each competency is assessed in two exercises, calibrating the number of competencies per exercise, standardizing Frame-of-Reference (FoR)-based ratings and integration meetings, and expanding the use of dimension × exercise scores for feedback and development.

Key words : Assessment center; Construct validity; Multitrait-multimethod (MTMM); Confirmatory factor analysis (CFA)