

다차원 퍼기 모형과 자료의 특성: 순위자료와 평정자료의 비교

한 유 화[†]

충북대학교 심리학과 / 연세대학교 법학연구원

본 연구는 순위자료와 평정자료를 이용하여 판단자와 판단대상을 하나의 도면에 시각화하는 분석 기법인 퍼기 모형을 적용한 후 그 결과를 비교하였다. 퍼기 모형을 비롯한 다차원 척도법 모형들은 통계적 가정에서 비교적 자유로운 분석으로 소집단 연구나 개인차에 관심을 가지는 연구자들에게 유용한 도구가 될 수 있다. 분석에 사용된 자료는 33개 행동을 비난가능성 차원에서 측정하기 위해 수집된 37명의 자료로, 33개 행동에 대해 두 가지 방법으로 수집된 자료였다. 33개 행동을 두 개씩 짝비교한 자료는 순위자료로 변환하고, 평정자료는 원자료를 그대로 사용하였다. 순위자료와 평정자료의 관계를 파악하기 위하여 두 자료의 순위 상관계수를 산출하고, 두 자료에서 산출된 33개 행동들의 상관계수 행렬을 비교하였다. 그 결과 두 자료는 서로 관련이 없는 것으로 나타났으며, 상관계수 행렬도 매우 다른 양상으로 나타났다. 퍼기분석을 수행한 결과, 평정자료보다 순위자료를 이용하였을 때 판단대상들의 배치가 더 분산되었으며, 판단대상과 판단자의 관계를 시각적으로 파악하는 데 더 유용할 것으로 판단할 수 있었다. 본 연구는 판단대상과 판단자가 동일한 순위자료와 평정자료의 퍼기분석 결과가 다르다는 것을 보여줌으로써 개인차 연구에서 유용하게 활용될 수 있는 퍼기 모형을 적용하기 위해 연구자들이 고려해야 할 점을 확인하였다.

주제어 : 다차원 척도법, 퍼기 모형, 순위자료, 평정자료, 개인차 연구

[†] 교신저자: 충북대학교 심리학과 강사, 연세대학교 법학연구원 학술연구교수, E-mail: yuhwahan82@gmail.com



Copyright © 2023, The Korean Psychological Association. 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.

다차원 척도법(multi-dimensional scaling: MDS) 모형 중 하나인 펴기 모형(unfolding model, Coombs, 1950)은 직사각행렬을 분석하여 자료의 판단대상과 판단자를 척도화(scaling)하는 대표적인 반응중심 측정방법(response-centered scaling method)으로 그 척도화의 결과로 자료를 시각화(visualization)한다. 대표적인 직사각행렬 구조의 자료는 일련의 대상에 대한 선호순위/서열(preference rank) 또는 대상의 속성(대상에 대한 선호)을 평정(rating)한 자료이다. 순위자료(ranking data)와 평정자료(rating data)는 응답에 요구되는 인지적 과정이 다르기 때문에 측정치에 반영되는 속성이 다를 수 있고, 따라서 동일한 대상에 대한 측정 자료의 공변량 구조도 다를 수 있다. 이와 같은 자료의 속성에서 관찰되는 차이는 순위자료와 평정자료가 한 사람으로부터 얻어졌더라도 관찰될 수 있다. 즉, 참가자 내 설계를 통해 획득한 순위자료와 평정자료를 이용한 펴기분석의 결과가 다를 수 있다. 현재까지 순위자료와 평정자료의 비교는 주로 두 자료의 관련성과 기술통계치를 파악하는 방식으로, 참가자 간 설계를 통해 수집한 자료에 기초하여 이루어졌으며(예를 들면, Alwin & Krosnick, 1985; Russell & Gray, 1994), 그 비교를 위해 펴기분석을 적용한 결과는 찾기 어렵다. 본 연구는 순위자료와 평정자료를 이용하여 펴기분석을 수행하고, 그 적합도와 자료의 시각화 결과(도면)를 비교하였다.

다차원 척도법과 펴기 모형

다차원 척도법은 거리모형(distance model)에 기초하여 유사성(similarity) 또는 근접성(proximity) 자료에 내재된 심리적 속성들을 파악하도록

돕는 자료 시각화 방법이다(Kruskal, 1964; Torgerson, 1952). 다차원 척도법 모형은 도면에 배치된 대상(또는 판단자)과 자료의 관계에 대한 가정에 따라 계량적(metric) 모형과 비계량적(non-metric) 모형으로 구분할 수 있다. 비계량적 방법에는 개인차를 고려한 도면을 도출할 수 있는 개인차 척도법(individual difference scaling)과 판단대상과 판단자를 동시에 시각화할 수 있는 펴기 모형이 있다. 다차원 척도법은 자극(대상)들을 다차원 공간에 시각화하기 위해 차원을 찾고 그 차원에서 자극의 측정치를 찾는 측정법이므로 자극중심 측정 방법(stimulus-centered scaling method)으로 분류된다. 그러나 판단대상과 판단자를 하나의 공간에 시각화하는 펴기 모형은 자극과 판단자를 같은 차원에서 측정하는 반응중심 측정 방법이다. 두 방법은 모두 자료에 내재된 차원을 이용하여 자극 또는 자극과 판단자 모두를 척도화하는 방법이다.

일반적인 다차원 척도법 모형들은 대상들 사이의 (비)유사성 정도를 행렬의 요소로 하는 ‘대상 × 대상’의 정사각 대칭행렬을 분석하는 반면, 펴기 모형은 판단자들이 판단대상에 대해 평가한 순위(서열) 또는 평정치를 행렬의 요소로 하는 ‘참가자 × 대상’의 직사각행렬을 분석한다. 펴기 모형의 자료는 판단대상과 판단자의 근접성(proximity)으로 간주되므로 판단대상(또는 판단자) 간 근접성(거리) 정보가 누락된 다차원 척도법의 특수한 모형이라고 할 수 있다(D'Ambrosio et al., 2022).

펴기 모형은 다양하지만 분석의 기본적인 원리는 다차원 척도법과 동일하다(Borg et al., 2018). 다른 점이 있다면 사용되는 근접성 자료의 형태와 속성이다. 다차원 척도법의 과정은 다음과 같다: 1) (비)유사성 자료를 거리모

형을 적용할 수 있는 자료로 변환한다. 이 과정은 자료가 삼각비동등성(triangle inequality, a에서 b의 거리와 b에서 c의 거리 합은 a에서 c의 거리와 같거나 그보다 크다는 조건)을 충족하도록 변환하는 과정으로 임의의 숫자를 사용하여 자료를 선형변환한다. 2) 변환된 자료행렬을 구성하는 숫자를 상위(disparity)라고 하며, 이 행렬을 분해하여 차원과 좌표값에 해당하는 값들을 산출한다. 즉, 초기 도면(initial configuration)을 얻는다. 3) 알고리즘 방식을 사용하여 도면의 대상 간 (유클리드) 거리가 일정한 기준에 부합할 때까지 순환절차(iteration)를 반복한다. 다차원 척도법은 분석 자료의 행과 열이 같으므로 분석 후에 시각화를 위한 하나의 행렬(대상 × 차원)이 얻어지는 반면, 퍼기분석은 행과 열이 각각 판단자와 판단대상이므로 분석 후에 두 개의 행렬(‘판단자 × 차원’과 ‘대상 × 차원’)이 얻어지고, 이 두 행렬을 하나의 도면에 시각화함으로써 결합도면(joint configuration)을 산출한다.

퍼기 모형의 유용성

퍼기분석은 순위/평정 자료를 이용하여 하나의 도면에 판단대상과 판단자를 동시에 배치하기 때문에 심리학 영역에서 활용도가 높을 수 있다. 퍼기분석을 통하여 얻을 수 있는 판단대상과 판단자의 위치(좌표)는 곧 특정 차원에서 이들의 위치 즉, 측정치로 간주될 수 있다. 따라서 퍼기분석을 이용하는 경우, 상관계수를 산출하지 않고도 결합도면 상에서 판단대상과 판단자의 관계를 시각적으로 파악할 수 있을 뿐만 아니라 판단대상과 판단자의 도면상의 거리를 산출하여 분석에 활용하거나 특정 차원에서 개인차의 양을 수량화할 수 있

다. 특히, 자료의 중심경향성(central tendency)보다는 소표본을 대상으로 개인차에 관심을 가진 연구자라면 퍼기분석을 통하여 기존의 통계적 방법으로 파악할 수 없는 유용한 정보를 얻을 수도 있다(Borg et al., 2017; Fabbris, 2012). 최근에 국내에서 발표된 한 연구는 퍼기 모형을 적용하여 범죄 피해자의 피해자다움을 평가하는 데 중요하게 고려될 수 있는 네 개의 특성에 대한 순위자료를 분석하였다. 이 분석을 통하여 네 개의 특성 중 피해자다움에 대한 편견을 구성하는 피해자의 옷차림이나 피해자가 늦은 시간에 외출하는 것을 중요하게 생각하는 사람들은 소수에 불과함을 알 수 있었다. 연구자들은 퍼기분석의 결과를 바탕으로 피해자다움 평가에서 집단적인 경향성보다 개인차가 중요하게 고려되어야 한다고 주장하였다(Jin et al., 2023). 이와 같은 결과는 중심경향성을 분석하는 방법으로는 탐지해내기 어려우므로 퍼기분석의 유용성을 보여주는 사례라고 할 수 있다.

퍼기분석 도면의 해석

퍼기분석의 도면에서 판단대상과 판단자의 상대적 거리는 ‘선호’ 또는 ‘중요성’으로 해석될 수 있다. 이 경우 판단자의 위치는 심리적 차원에서 판단자가 가장 선호하고 중요하게 생각하는 지점 즉, 이상점(ideal point)으로 간주된다. 판단자의 위치는 원점으로부터 시작되어 판단자가 가장 선호하고 중요하게 생각하는 방향 즉, 벡터(vector)로 이해될 수도 있다. 이상점 모형을 적용하는 경우 판단자와 판단대상의 거리는 방향에 관계없이 해석되지만 벡터 모형을 적용하는 경우 판단자의 벡터를 중심으로 해석된다(Borg et al., 2018). 모형

의 선택은 연구 목적에 따라 퍼기분석의 도면을 어떻게 이해할 것인지에 따라 달라져야 할 것이다. 본 연구는 이상점 모형만을 다루고자 한다. 구체적인 퍼기분석 도면의 해석 예를 제시하기 위하여 R의 smacof 패키지에 포함된 공개 자료인 ‘아침식사 종류에 대한 선호도’ 자료(Green & Rao, 1972)에 대한 2차원 퍼기분석을 직접 수행하여 결과 도면을 그림 1에 제시하였다.

그림 1에서 숫자로 표시된 지점은 각 판단자 즉, 그들의 이상점을 나타낸다. 도면의 가운데에 배치된 24번 판단자는 파란색으로 표시된 다양한 아침식사 중 어느 것과도 가깝지 않다. 그러나 10번과 38번 판단자는 커피 케이크(cofcake)와 매우 가깝게 배치되었다. 이 도면을 통하여 연구자는 24번 판단자는 어떤 종류의 아침식사도 특별히 선호하지 않으며, 10번과 38번 판단자는 아침식사로 커피 케이크를 가장 선호한다는 것과 10번 판단자가 24번 판단자보다 아침식사로 커피 케이크를 선

호한다는 것을 알 수 있다. 이와 같이 퍼기분석의 이상점 모형에서 판단대상과 판단자의 상대적 거리는 개인의 선호도에서의 이상점과 대상의 근접성을 반영한다(Busing, 2006).

순위자료와 평정자료의 수집

순위자료는 판단자에게 가장 선호하는 것에 순서대로 숫자를 부여하도록 요구하는 방식으로 직접 얻을 수 있다. 판단대상의 순서를 결정하기 위해서는 판단자는 대상들을 상대적으로 비교해야 하므로 대상의 수가 적은 경우, 이 방법은 참가자에게 쉬운 과제이다. 그러나 순서를 결정해야 하는 대상의 수가 증가할수록 직접 순서를 부여하기 어려워진다(Feather, 1973; Rokeach, 1973). 순위자료는 짝 비교(paired comparison)를 통해 수집한 자료를 변환하여 얻을 수도 있다(Böckenholt, 2004). 이 변환은 판단대상들의 모든 가능한 쌍에서 각 대상이 선택된 빈도에 기초하여, 각 대상의

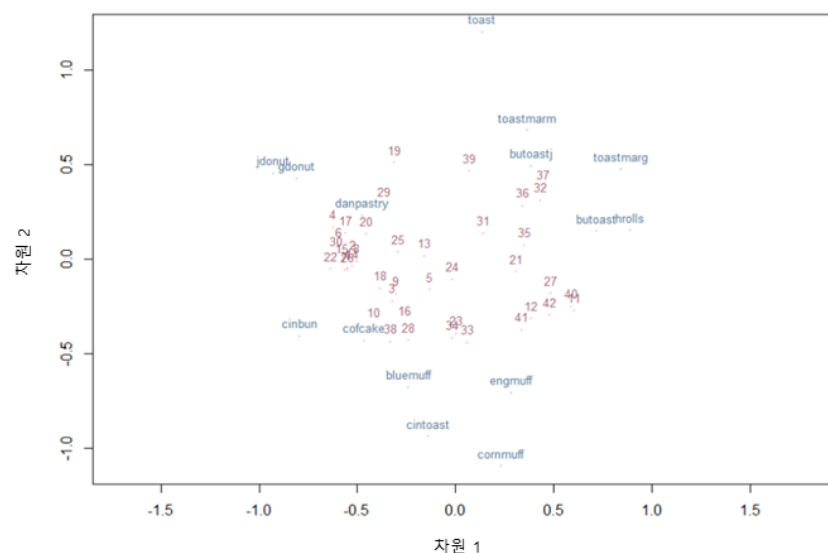


그림 1. 아침식사 종류에 대한 선호 순위의 퍼기분석 결함도면 예시

순위를 결정한다. 이렇게 하면 범주적 판단을 순위정보가 누락되지 않은 완전한 순위자료(fully expressed ranking)로 변환할 수 있지만, 이 방법 또한 자극의 개수가 많아지면 자료를 얻기 어려워진다는 단점이 있다. 다만, 판단자들이 각 시행에서 수행해야 하는 과제 자체의 난이도는 매우 쉽다는 장점이 있다. 순위자료는 수집하기 어렵지만, 신뢰도와 타당도가 평정자료보다 높다고 알려져 있다(Krosnick, 1999; Rankin & Grube, 1980; Reynolds & Jolly, 1980).

평정자료는 심리학 영역에서는 매우 흔하게 볼 수 있다. 리커트식 척도(Likert scale, Likert, 1932)를 사용하여 판단대상에 특정한 점수 범위(예를 들면, 1점은 매우 아니다, 9점은 매우 그렇다) 내의 숫자를 부여한 자료는 대표적인 평정자료이다. 평정자료는 각 대상에 대한 절대적인 판단이 반영된 자료이다. 예를 들어, 어떤 사진 속 고양이를 좋아하는 정도를 7점으로 평정하였다면, 이 고양이에 대한 선호 반응은 다른 고양이에 대한 선호 정도와 비교되지 않고, 독립적으로 이루어지며, 6점, 7점, 8점에 해당하는 반응범주의 경계와 비교되는 것으로도 이해할 수 있다(Torgerson, 1958). 평정자료의 장점은 판단대상들에 대한 상대적 선호도는 알 수 있지만 각 판단대상에 대한 선호 정도를 알 수 없는 순위자료와 달리 각 판단대상에 대한 선호 정도를 반영하는 독립적인 반응을 모두 얻을 수 있다는 것이다(Elig & Frieze, 1979; Groves & Kahn, 1979; Munson & McIntyre, 1979).

순위자료와 평정자료의 관계

순위자료와 평정자료는 집단 수준 즉, 통합

된 통계치(aggregated statistics)에서는 관련성이 높은 것으로 알려져 있고 개인 수준에서도 유사한 관련성이 있다. 순위자료와 평정자료의 기술통계치 간 상관에 관심을 가졌던 한 연구는 집단 수준에서 순위자료와 평정자료의 상관이 .85로 매우 높았지만 개인 수준에서는 25%의 참가자가 .80보다 낮은 상관을 가지는 것으로 나타났다고 보고하였다. 이 연구에서 어떤 참가자들의 두 자료는 부적 상관관계를 보여주기도 하였다(Russell & Gray, 1994). 집단 내 설계를 이용하여 수집된 순위자료와 평정자료의 상관관계를 확인한 또 다른 연구는 하위집단(국가)에 따라 .62에서 .77까지의 관련성을 보고하였다(van Herk & van de Velden, 2007). Rankin과 Grube(1980)는 참가자 내 설계를 이용하여 수집된 순위자료와 평정자료의 신뢰도와 타당도를 비교하였다. 이들의 연구 결과 타당도에서는 두 자료 모두 매우 높은 상관계수를 보여주어 자료 유형에 따른 차이를 관찰하지 못했지만, 평정자료보다 순위자료의 검사-재검사 신뢰도가 더 높았다. 이들은 이와 같은 결과를 순위자료의 내준적 속성(ipsative characteristic, Cattell, 1944)에 의한 것으로 판단하고 신뢰도에서도 차이가 없다고 주장하였다. 내준적 속성은 순위자료에서 각 판단대상에 부여된 숫자들이 판단대상의 수를 넘지 못하고, 이미 부여된 숫자들에 종속될 수밖에 없다는 것을 의미한다.

위의 연구 결과는 주로 순위자료와 평정자료의 평균을 비교하거나 상관계수를 확인함으로써 이루어졌다. 한 연구는 두 자료의 공변량 구조의 비교를 위해 두 자료로 수행한 요인분석 결과를 비교하였는데, 도출된 요인들은 유사하기는 하였지만 동일하지는 않았다(Alwin & Krosnick, 1985). 이 연구에서는 두 자

료의 준거관련 타당도 즉, 두 자료로 도출된 요인들이 이론적으로 관련된 변인들을 잘 예측하는지도 확인하였다. 그 결과 도출된 요인과 준거의 관련성이 평정자료보다 순위자료에서 더 컸고, 이 결과에 기초하여 저자들은 연구자들이 자료의 유형에 따라 다른 결론에 이를 수 있음에 주의해야 한다고 주장하였다. 보다 최근에 수행된 연구는 순위자료와 평정자료의 수집 방식에서 기인하는 편향(순위자료의 경우 부적편향, 평정자료의 경우 정적편향)을 교정한 후에도 두 자료로부터 도출된 요인들이 동일하지 않음을 보여주었다(Hino & Imai, 2019).

연구 문제

본 연구는 순위자료와 평정자료를 비교하여 개인차를 분석하는 유용한 방법인 퍼기분석을 중심으로 그 결과를 비교하였다. 퍼기분석은 순위자료와 평정자료에 공통적으로 적용될 수 있는 분석방법일 뿐만 아니라 척도화 방법이라는 점에서 두 자료의 변량 차이에 따른 척도(scale) 문제에 대처하기에도 적절한 도구가 될 수 있다. 두 자료의 퍼기분석 결과를 비교하는 것은 개인차에 관심 있는 연구자들이 자료의 특성에 대한 이해를 높이고, 퍼기모형의 활용 가능성을 모색하며 자료수집 방법을 결정하는 데 유용한 정보를 제공할 수 있을 것이다.

순위자료는 판단대상의 수가 많아지면 수집하기 어려워진다는 단점이 있지만, 평정자료에는 없는 순위자료의 내준적 속성이 오히려 개인차 분석에 유용하게 작용할 수 있다는 견해가 있는바(Fabbris, 2012), 순위자료와 평정자료의 퍼기분석 도면에 차이가 있을 가능성

이 있다. 순위자료의 내준적 속성이 실제로 개인차 분석에 유용하다면 평정자료보다 개인을 잘 구분할 것이므로, 순위자료를 이용한 퍼기분석의 도면에서 평정자료를 이용했을 때보다 개인들이 더 넓게 퍼져있을 것으로 기대할 수 있다. 또한, 순위자료는 판단대상에 동일한 순위가 부여되지 않아 자료에 동점(tied value)이 존재하지 않으므로 판단대상들도 평정자료를 이용하는 경우보다 더 잘 구분될 가능성이 있다. 동점이 존재하지 않는다는 순위자료의 특성은 두 자료의 변량 차이에 영향을 줄 수 있다. 즉, 두 자료는 척도가 다른 자료이다. 서론에 언급한 것처럼 퍼기분석은 반응 중심 측정방법으로서 자료를 그 내재된 차원에서 척도화하는 과정으로 이해될 수 있다. 척도화 과정은 자료에 새로운 척도를 부여하므로 퍼기분석은 척도가 다른 순위자료와 평정자료를 비교하기 위한 적절한 도구가 될 수 있다. 다만, 본 연구는 퍼기분석의 수행과정에서 자료의 동점을 그대로 유지하지 않고 다른 상위를 허용하는 변환 즉, 동점을 헤쳐놓는 일차적 접근(primary approach)을 사용하였다. 이 방법은 제약이 상대적으로 적으므로(De Leeuw & Mair, 2009) 퍼기분석 결과 도면에 동점의 영향을 감소시킬 수 있을 것으로 기대하였다. 두 자료가 척도만 다른 측정치라면 퍼기분석의 결과(척도화된 결과) 또한 매우 유사해야 할 것이다.

본 연구는 순위자료와 평정자료를 비교하기 위하여 두 자료를 비교한 선행연구 결과들이 재검증되는지 확인하고자 순위자료와 평정자료의 평균과 상관계수를 확인하고, 판단대상에 대한 두 자료의 상관계수 행렬을 검토하였다. 또한, 퍼기분석 결과의 비교를 위해서는 모형의 적합성 지수(STRESS)와 퍼기분석 도면

의 시각적 비교와 통계적 비교를 수행하였다.

방 법

분석자료

본 연구는 행동의 비난가능성을 연속적 간격법(method of successive interval)과 상대적 비교법(the method of comparative judgment)을 이용하여 측정하기 위해 기관심의위원회 승인을 받고 수행된 연구(CBNU-202108-HR-0130, 미출판)에서 수집된 자료를 재분석하였다. 원자료(original data)는 비난가능성이 있는 33개의 행동에 대해 두 가지 방법으로 평가한 두 개의 자료 세트와 판단자의 성별 및 연령 정보였다. 분석자료는 총 37명의 판단자의 자료였으며, 판단자의 평균 연령은 24.43세(표준편차: 2.27세, 범위: 21세~30세)였고 여성 판단자가 22명으로 전체의 59.46%를 차지하였다. 분석자료 중 첫 번째 자료는 판단자들에게 33개 행동을 두 개씩 짝지어 제시한 후 서로 비교하여 ‘더 비난받아야 한다고 생각하는 행동’을 선택하게 한 자료(짝비교 자료)였고, 두 번째 자료는 동일한 판단자들이 33개 행동에 대해 ‘비난받아야 하는 정도’를 1점(절대 비난받아서 안된다)에서 9점(반드시 비난받아야 한다) 사이의 리커트 척도상에 평정하게 한 자료(평정자료)였다. 평정자료의 응답을 검토한 결과 무성의하다고 판단할 수 있는 자료(예를 들면 모든 행동에 대해 같은 점수로 응답함)가 발견되지 않아 모든 자료를 분석에 사용하였다. 모든 판단자들은 두 방법 모두를 이용하여 33개 행동을 평가하였으므로 본 연구의 분석자료는 참가자 내 설계를 통해 수집

된 자료라고 할 수 있다. 판단자들은 리커트 척도에 먼저 응답하여 평정자료가 먼저 수집되었고 그 직후에 짝비교를 수행하였다. 짝비교를 위한 판단대상은 무선적인 순서로 짝지어 제시되었다.

본 연구의 목적은 순위자료와 평정자료의 퍼기분석 결과를 비교하는 것인바, 짝비교 자료는 순위자료로 변환하여 분석에 사용하였고, 평정자료는 그대로 사용하였다(37명이 33개 행동에 대해 1-9점으로 평정한 자료로 ‘판단자 수’ \times 33의 직사각행렬). 짝비교 자료를 순위자료로 변환하는 과정은 다음과 같다: 1) 33개 행동을 짝비교한 결과를 기록할 33 \times 33 행렬을 만든다. 2) 열과 행이 만나는 각 칸(cell)에 판단자가 열에 제시된 행동과 행에 제시된 행동을 비교한 결과를 기록하되, 열에 있는 행동을 선택하였으면 1, 행에 있는 행동을 선택하였으면 0을 기록한다. 3) 이 과정을 모든 판단자 자료에서 반복하여 판단자 수만큼의 33 \times 33 행렬을 만든다. 4) 각 판단자의 33 \times 33 행렬에서 각 열에 기록된 모든 행의 숫자들을 합한 값이 그 열이 나타내는 행동의 순위가 된다. 5) 모든 판단자의 33개 행동에 대한 순위 측정치를 이용하여 판단자를 행에, 33개 행동의 순위를 열에 배치하면 ‘판단자 수’ \times 33의 직사각행렬이 만들어진다. 이 자료는 모든 대상에 대한 순위가 평가된 완전한 순위자료(fully expressed rank data)이다. 원자료를 그대로 이용한 평정자료 역시 판단자가 행에 배치되고 행동이 열에 배치된 ‘판단자 수’ \times 33의 직사각행렬이지만, 순위자료는 행렬의 각 요소는 순위이고 평정자료의 각 요소는 행동의 비난 가능성에 대한 평정치이다. 순위자료에서는 각 행동은 짝비교에서 더 비난받아야 한다고 선택된 빈도가 높을수록 큰 숫자를

부여받게 되고 평정자료의 경우 더 비난받아야 한다고 평정될수록 높은 점수를 부여받게 되므로, 두 자료 모두에서 큰 숫자를 부여받을수록 비난 가능성이 높은 행동으로 간주된다.

분석방법

순위자료와 평정자료를 이용하여 퍼기분석을 수행한 결과를 비교하기 위하여 본 연구는 두 자료에서 각 행동별 평균과 표준편차 및 측정치 간 상관관계수(Spearman's r_{ho})를 비교하고(순위자료와 평정자료는 모두 서열척도 수준의 측정치이므로 Spearman의 상관관계수를 산출함), 두 자료의 퍼기 모형 적합도와 결과로 도출되는 결합도면을 비교하였다. 본 연구의 모든 분석은 R을 이용하여 수행되었으며, 퍼기분석을 위하여 smacof 패키지(De Leeuw & Mair, 2009)의 unfolding 함수와 permtest 함수, 결합도면의 비교를 위하여 같은 패키지의 Procrustes 함수를 사용하였다. Procrustes 변환은 다차원 척도법 도면을 비교하기 위해 자료의 시각화 과정에서 발생하는 부수적인 차이(회전, 변환, 확장 등 대상의 상대적 관계와 도면의 적합도에 영향을 주지 않는 의미 없는 차이)를 제거하는 데 사용되는 유사성 변환(similarity transformation) 방법(Borg et al., 2018)으로 본 연구에서는 순위자료 도면을 목표도면으로 사용하였다. 이 방법을 통해 자료에 의해 발생한 도면의 의미 있는 차이만을 비교할 수 있다. 순위자료와 평정자료는 모두 서열척도이므로 퍼기분석은 아무런 제약이 없는 서열적 모형(type = "ordinal")으로 수행하였다.

결 과

33개 행동의 순위 및 평정치 평균과 표준편차, 상관관계수

33개 행동에 대한 순위자료와 평정자료의 특성을 파악하기 위하여 통합된 측정치인 평균과 표준편차를 산출하여 표 1에 제시하되, 순위자료와 평정자료의 비교를 위해 통합된 측정치를 순위로 변경한 순위정보를 추가하였다.

표 1에 제시된 바와 같이 순위자료와 평정자료의 평균과 표준편차는 두 자료에 포함된 측정치의 가능한 점수 범위와 의미가 달라 직접적인 비교가 어렵다. 따라서 두 자료의 통합된 측정치인 평균에 기초하여 생성한 순위를 비교하였다. 두 자료의 평균을 이용한 순위는 작게는 1, 크게는 28까지 차이가 있었다. 가장 작은 차이가 났던 행동은 “재미로 친구에게 신체적 고통을 주기”였고, 가장 큰 차이가 났던 행동은 “도망갈 수 없는 상태의 동물을 때리거나 괴롭히기”였다. 이외에 “친구의 슬픔을 모른척 하기”와 “미성년자와 성관계”도 두 자료의 통합된 값으로 얻은 순위에 매우 큰 차이가 있는 것으로 나타났다. 두 자료의 평균 간 상관관계수는 .260이었으며 통계적으로 유의하지 않은 것으로 나타났다($p=.144$). 개인 수준에서의 상관관계수는 33개 행동에 대해 -.34(V11)에서 .38(V25)의 범위로 나타났다. 이 중 통계적으로 유의한 관계는 유의수준 .1에서 유의한 상관관계수를 포함하여 4개에 불과하였다.

33개 행동에 대한 순위자료와 평정자료의 상관관계수 행렬

33개 행동에 대한 순위자료와 평정자료의

표 1. 33개 행동의 순위 및 평정치 평균(표준편차)과 상관계수

| 변인 | 행동 | 순위자료 | | 평정자료 | | ρ |
|-----|--|-------------|----|------------|----|--------|
| | | $M(SD)$ | 순위 | $M(SD)$ | 순위 | |
| V1 | 에이즈에 감염된 것을 숨기고 성관계 | 32.22(1.20) | 33 | 8.14(1.18) | 31 | .21 |
| V2 | 다른 사람의 자동차를 재미로 훔쳐타고 여행 | 25.81(4.67) | 30 | 6.97(1.67) | 22 | .22 |
| V3 | 부당한 복지혜택을 달라고 주장 | 17.73(4.85) | 17 | 8.54(1.04) | 33 | .01 |
| V4 | 자신의 업무와 관련된 뇌물 받기 | 22.30(5.54) | 24 | 6.43(1.41) | 16 | .34† |
| V5 | 자신과 의견이 다른 사람을 위협하기 | 22.68(4.77) | 25 | 6.70(1.47) | 20 | -.1 |
| V6 | 시험에서 부정행위 하기 | 20.87(4.97) | 21 | 7.97(1.46) | 30 | -.06 |
| V7 | 아내(남편) 몰래 바람 피우기 | 27.27(5.51) | 31 | 7.70(1.05) | 29 | 0 |
| V8 | 도둑질한 물건이라는 것을 알면서도 구입하기 | 15.78(3.52) | 15 | 6.68(1.45) | 19 | .11 |
| V9 | 주차된 차량을 긁고 차 주인에게 알리지 않기 | 19.03(5.07) | 18 | 7.51(1.28) | 27 | -.01 |
| V10 | 탈세하기 | 24.87(4.67) | 28 | 7.49(1.24) | 26 | -.14 |
| V11 | 요금을 내지 않고 대중교통 이용하기 | 14.51(4.20) | 13 | 6.87(1.58) | 21 | -.34* |
| V12 | 성매매 | 24.35(8.88) | 26 | 4.30(1.98) | 11 | -.21 |
| V13 | 마약류 소지하기 | 22.16(7.38) | 23 | 6.62(1.93) | 17 | .14 |
| V14 | 경찰과 싸우기 | 15.89(7.46) | 16 | 5.68(1.89) | 14 | -.03 |
| V15 | 개인의 이익을 위해 거짓말하기 | 15.22(4.45) | 14 | 7.03(1.89) | 23 | -.16 |
| V16 | 동성애 | 5.03(6.66) | 3 | 5.60(1.67) | 13 | .24* |
| V17 | 미성년자와 성관계 | 24.78(8.46) | 27 | 1.32(0.53) | 2 | -.15 |
| V18 | 주운 돈을 그냥 갖기 | 9.19(2.91) | 7 | 2.41(1.94) | 5 | -.34 |
| V19 | 어려운 사람 돕기 | 1.87(0.89) | 1 | 2.03(1.67) | 3 | -.09 |
| V20 | 이혼 | 5.87(4.52) | 5 | 7.35(1.62) | 25 | .03 |
| V21 | 낙태 | 13.22(8.69) | 11 | 3.05(1.84) | 6 | -.11 |
| V22 | 정당방위에 의한 살인 | 13.14(9.06) | 10 | 4.08(1.53) | 8 | -.32 |
| V23 | 친구의 슬픔을 모른척 하기 | 10.27(4.75) | 8 | 8.19(0.97) | 32 | -.07 |
| V24 | 어린이나 노인의 짐을 들어주기 | 2.60(3.41) | 2 | 4.14(1.77) | 9 | .02 |
| V25 | 재미로 친구에게 신체적 고통을 주기 | 25.68(5.09) | 29 | 7.60(1.28) | 28 | .38 |
| V26 | 회사의 비리를 고발하기 | 5.81(3.53) | 4 | 5.49(1.71) | 12 | -.3 |
| V27 | 도망갈 수 없는 상태의 동물을 때리거나 괴롭히기 | 28.87(3.65) | 32 | 2.19(1.76) | 4 | .03 |
| V28 | 미운 사람에게 불행한 일이 있기를 기도하기 | 7.76(2.40) | 6 | 6.65(1.38) | 18 | .17 |
| V29 | 친구의 외모를 공개적으로 비난하기 | 21.92(5.11) | 22 | 1.24(0.55) | 1 | -.08 |
| V30 | 모르는 사람들과 하는 놀이에서 돈을 따기 위해 부정행위 하기 | 19.70(3.72) | 19 | 3.70(1.84) | 7 | -.31† |
| V31 | 내 부탁을 들어준 친구의 부탁을 거절하기 | 10.57(3.30) | 9 | 5.73(1.41) | 15 | .01 |
| V32 | 친분이 있는 사람에게 유리한 결정을 하여 모르는 사람에게 불이익 주기 | 20.38(4.49) | 20 | 7.16(1.44) | 24 | .04 |
| V33 | 친한 친구 헐뜯하기 | 13.51(5.76) | 12 | 4.27(2.08) | 10 | -.2 |

† $p < .1$, * $p < .05$

주. Spearman's ρ 의 유의성은 본 분석자료의 동점 문제로 정확하지 않을 수 있음.

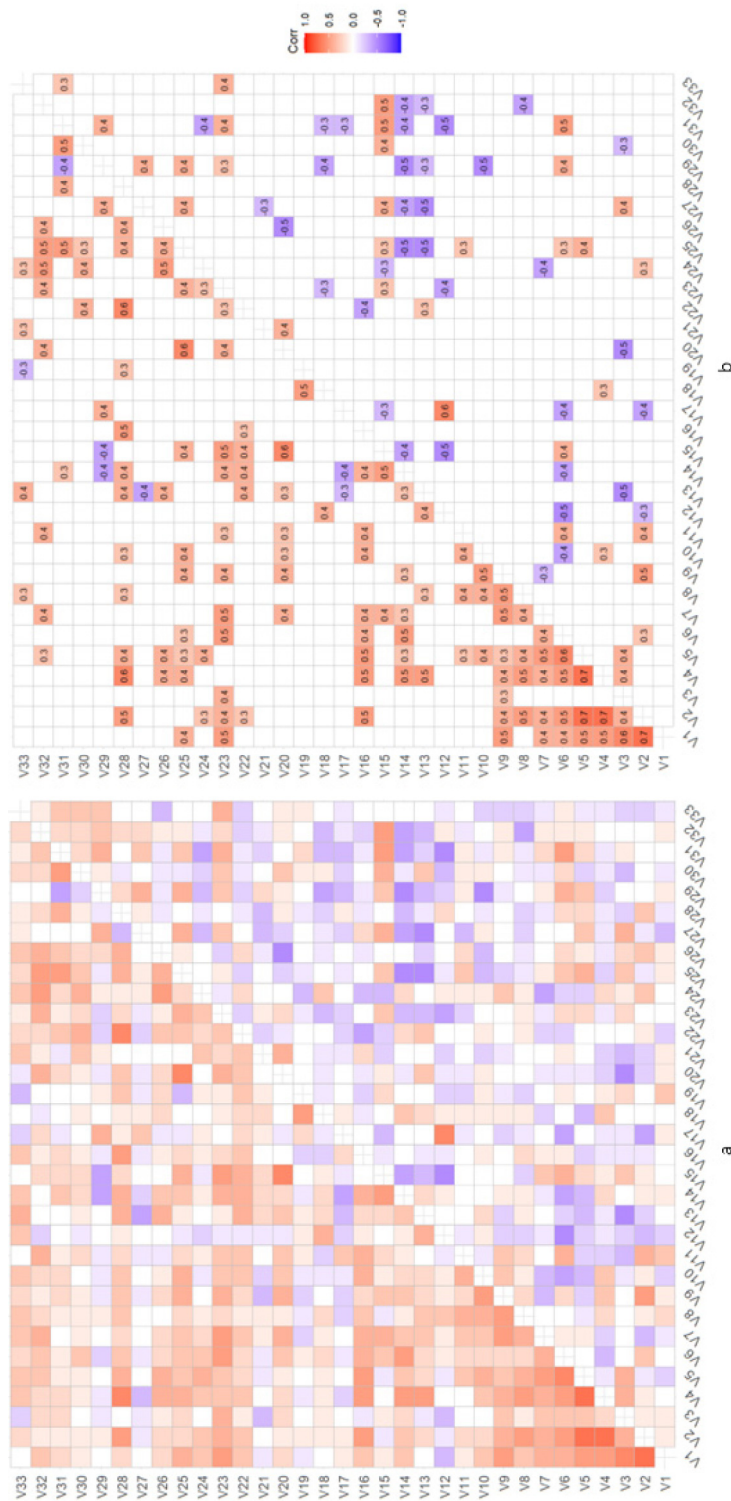


그림 2. 순위자료와 평정자료의 상관관계의 시각화

주. 33개 행동의 상관계수는 각 행동들에 대해 평가된 비난가능성의 유사성을 의미함. a는 통계적 유의성을 고려하지 않고 제시한 그림이며, b는 통계적으로 유의한($p < .05$) 상관관계만 제시한 그림임. 다만, Spearman's ρ 의 유의성은 본 분석자료의 동점 문제로 정확하지 않을 수 있음. 대각선 아래는 순위자료, 대각선 위는 평정자료를 이용하여 산출한 상관계수임.

상관계수 행렬이 유사한지 확인하기 위하여 순위자료와 평정자료 각각에서 33개 행동 간 상관계수를 산출하여 그림 2에 제시하였다. 대각선 아래는 순위자료, 대각선 위는 평정자료의 상관계수이다.

그림 2에서 확인할 수 있듯이, 33개 행동에 대한 측정치의 상관은 자료에 따라 다소 달랐다. 순위자료의 경우 33개 행동에 대한 측정치들의 관련성에 대한 방향과 크기에서의 경향성을 찾기 어려웠다. 그러나 평정자료의 경우 대부분의 행동에 대한 평정치들이 정적 관련성을 가지는 것을 확인할 수 있었고, 이 경향은 통계적으로 유의한($p < .05$) 관계만 표시한 그림 2b에서도 유지되는 것을 확인할 수 있었다.

퍼기 모형의 적합도 및 결합도면의 비교

순위자료와 평정자료를 이용한 아무런 제약이 없는 서열적 퍼기분석을 수행하여 2차원 결합도면을 도출한 결과, 적합도를 나타내는

STRESS 지수는 각각 0.244와 0.234로 나타났다. 이 지수가 무선 자료를 퍼기분석 하였을 때 얻을 수 있는 STRESS와 통계적으로 유의하게 다른지 확인하기 위한 순열검증(permutation test)을 수행하였다. 순열검증은 permutation 횟수(순열의 추출 수)를 100으로 지정하여 수행하였으며, 그 결과인 p -값은 반복적으로 추출된 STRESS 중 본 연구에서 산출된 STRESS보다 극단적인 값의 비율로 산출된다. 검증 결과, 무선자료의 STRESS는 대략 0.4 정도로 두 자료에 대한 퍼기분석 결과인 결합도면의 STRESS는 이보다 통계적으로 유의하게 낮았다($p < .001$). 즉, 순위자료와 평정자료의 퍼기 모형 적합도는 통계적으로 유의하여 두 자료의 결합도면은 두 자료를 잘 반영한다고 할 수 있었다.

두 자료의 퍼기분석 결합도면을 시각적으로 비교하기 위하여 순위자료와 평정자료의 결합도면을 그림 3에 제시하였다.

그림 3에 제시된 바와 같이 판단대상들은 평정자료의 도면에서보다 순위자료의 도면에

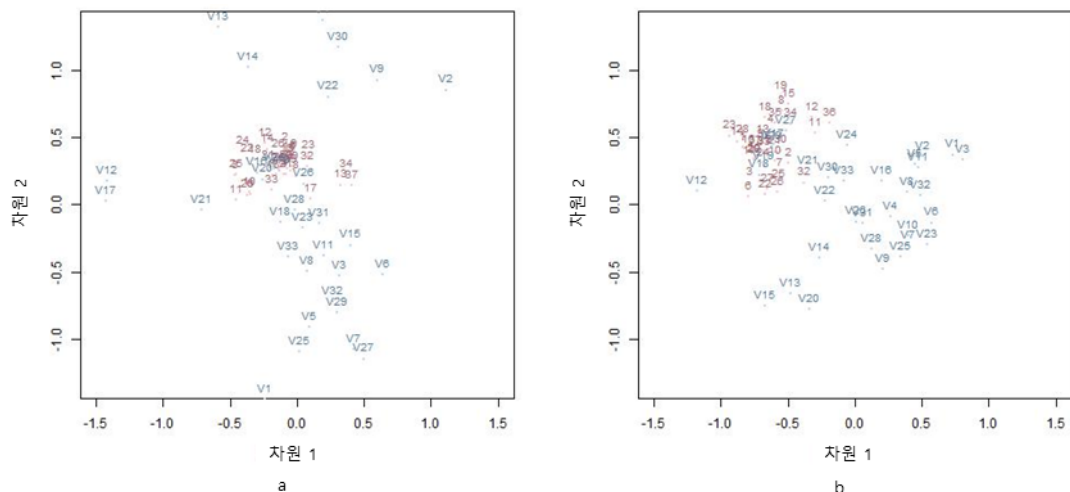


그림 3. 순위자료(a)와 평정자료(b)의 결합도면

서 더 넓게 퍼져있는 경향이 있었다. 이것은 순위자료를 이용하여 퍼기분석을 하는 경우 판단대상들이 더 잘 구분될 수 있음을 시사한다. 반면에 두 도면에 배치된 판단자들의 경우 군집화된 양상은 시각적으로 큰 차이가 없었다. 그러나 두 자료에서 동일한 판단대상과 판단자의 도면상 위치는 다소 다르다는 것을 확인할 수 있었다.

퍼기 모형의 판단자 도면과 판단대상 도면의 차이

순위자료와 평정자료의 퍼기분석 결합도면에서 관찰된 시각적인 차이가 의미 있는 차이인지 확인하기 위해 Procrustes 변환을 통하여 도면을 비교하였다. Procrustes 변환은 목표 행렬을 중심으로 비교하려는 행렬을 변환하여 두 행렬을 비교하는 방법이다(Mair et al., 2022). 이 방법은 차원이 동일한 두 개의 행렬만을 비교할 수 있다. 그러나 서론에 언급한 바와 같이 퍼기분석의 결합도면은 분석 결과

로 산출된 판단대상의 행렬(판단대상 × 차원)과 판단자 행렬(판단자 × 차원)을 결합하여 작성되는바, 퍼기분석 결과 두 개의 행렬이 도출된다. 따라서 결합도면을 직접 비교하는 것이 불가능하여 두 자료의 도면을 통계적으로 비교하기 위해 판단대상과 판단자의 행렬을 구분하여 비교하였다. 그림 4는 Procrustes 변환 후의 판단대상과 판단자의 도면을 보여준다.

그림 4a는 순위자료와 평정자료의 판단대상 배치를 동시에 보여주는데, 두 자료의 판단대상 도면이 매우 다르다는 것은 시각적으로도 확인할 수 있다. 이 도면에서 목표도면인 순위자료와 Procrustes 변환한 평정자료의 도면에 배치된 대응되는 판단대상들의 유사성을 나타내는 상관계수는 -.34로 비교의 기준(무선적인 도면에서 산출된 상관계수의 99%에 해당하는 임계치로 이 값보다 높은 상관계수는 두 도면이 무선적으로 기대되는 것보다 유사하다는 것을 의미하고 낮은 상관계수는 그렇지 않음을 의미함)이 되는 .41(산출 방법은 Borg et al.,

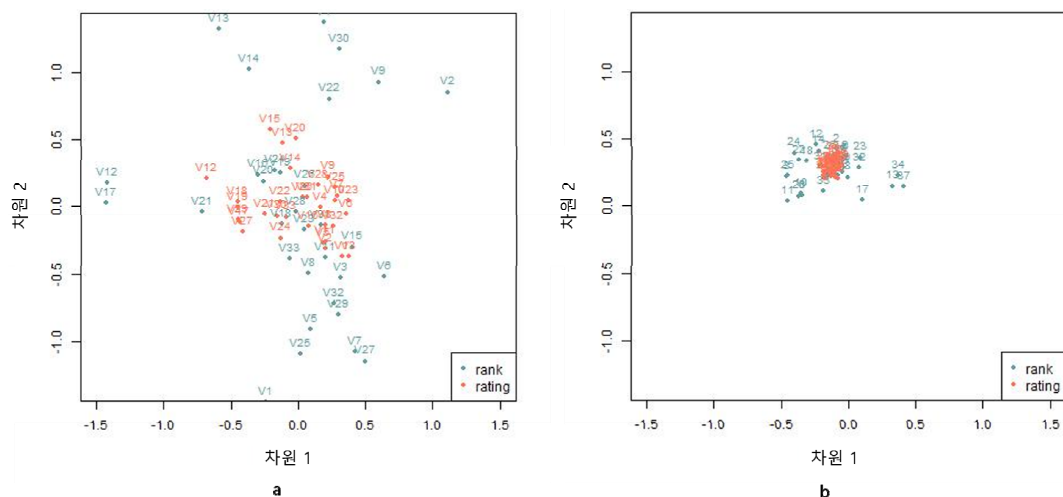


그림 4. Procrustes 변환된 순위자료와 평정자료의 판단대상(a)과 판단자(b) 배치 도면

2018, p.85의 R코드 참조)보다 낮을 뿐만 아니라 부적 상관관계를 보여준다. 이것은 순위자료에서 멀리 위치한 판단대상들이 평정자료에서는 가깝게 위치하는 것으로, 또는 그 반대의 양상으로 시각화될 수 있음을 의미한다. 순위자료의 판단대상들은 약 세 개로 군집화된 경향을 확인할 수 있다. 구체적으로 V12(성매매)와 V17(미성년자와의 성관계)은 ‘성범죄’로 군집화할 수 있고, V13(마약류 소지하기), V14(경찰과 싸우기), V22(정당방위에 의한 살인)와 V30(모르는 사람들과 하는 놀이에서 돈을 따기 위해 부정행위 하기) 등은 ‘범죄행동’으로, V15(개인의 이익을 위해 거짓말하기), V20(이혼) 및 V33(친한 친구 헐뜯하기) 등과 가까운 나머지 행동들은 ‘비도덕적 행동’으로 군집화할 수 있다.¹⁾

그림 4b는 순위자료와 평정자료의 판단자 배치에 다르다는 것을 보여준다. 다만, 목표도면인 순위자료의 도면과 Procrustes 변환한 평정자료 도면의 유사성을 나타내는 상관계수는 .94로 나타나 두 도면은 매우 유사하다고 할 수 있었다. 이 값은 무선적으로 기대할 수 있는 도면의 유사성(.38)보다 크기 때문에 이 시각적인 차이는 의미 있는 자료의 차이라고 결론 내리기 어렵고, 오히려 무선적으로 기대할 수 있는 것보다 더 유사하다고 할 수 있다. 그림 4b는 차원 축의 단위를 달리한다면 판단자들의 군집화 양상이 포착될 수 있겠으나, 도면의 차이가 통계적으로 유의하지 않으므로 그 양상은 해석하지 않았다.

1) 군집화 양상에 대한 판단은 본 연구자의 시각적 판단이며 본 연구의 목적이 두 자료의 도면을 비교하는 데 있으므로 통계적 검증은 수행하지 않았음.

논 의

본 연구는 순위자료와 평정자료를 비교하기 위하여 심리학 영역에서 활용 가능성이 높은 폐기분석을 사용하였다. 구체적으로 폐기분석을 간략히 소개하고 자료의 특성과 폐기분석 결과인 결합도면의 관계에 관한 지식을 제공하고자 순위자료와 평정자료의 폐기분석 결과를 비교하였다. 두 자료의 비교를 위해 출판되지 않은 기존 연구에서 수집된 자료를 재분석하였으며, 자료의 속성을 탐색하기 위해 순위자료와 평정자료의 기술통계치를 비교한 후, 두 자료의 폐기 모형 적합도 및 폐기분석 결과인 결합도면을 비교하였다.

순위자료와 평정자료의 비교

33개 행동의 비난가능성 순위와 평정치의 통합된 측정치인 평균을 산출하고, 이 평균에 기초한 순위가 얼마나 일치하는지 확인한 결과, 순위자료와 평정자료에 기초하여 산출된 각 행동의 순위는 행동이 무엇인지에 따라 크게 달라졌다. 또한, 순위 평균과 평정치 평균의 상관계수(집단 수준에서 두 자료의 관련성)는 .260으로 매우 낮을 뿐만 아니라 통계적으로도 유의하지 않았다. 이와 같은 결과는 선행연구의 결과와 매우 다른 것으로 대부분의 선행연구들에서 보고된 두 자료의 관련성은 집단수준에서는 .8 이상, 개인 수준에서도 .6 이상으로 높은 수준이었다(Russell & Gray, 1994; van Herk & van de Velden 2007).

선행연구와 본 연구의 상반된 결과는 자료가 수집된 과정에서의 차이일 수 있다. 선행연구들에서 순위자료는 판단자들로부터 순위를 직접 매기는 방식으로 수집된 반면, 본 연

구에서 사용된 자료는 짝비교를 수행한 후 순위자료로 변환한 자료이다. 즉, 본 연구의 순위자료는 모든 가능한 짝비교 결과에 기초한 순위이지만 선행연구들의 순위자료는 판단자가 한 대상에 대해 판단하기 위해 나머지 모든 대상과의 비교를 수행하여 순위를 결정하였는지가 불분명하다. 평정자료의 경우에는 본 연구의 분석자료와 선행연구의 자료수집 과정에서 차이가 없다는 점을 고려한다면, 순위자료를 어떻게 수집하는지에 따라 평정자료와 순위자료의 관련성이 낮거나 없을 수 있을 것으로 추론할 수 있다. 다만, 본 연구의 목적은 퍼기 모형의 적용이었던바, 소수($N=37$)의 판단자 자료를 사용하였기 때문에 극단치가 기술통계치의 비교에 영향을 주었을 가능성이 있다. 또한, 본 연구는 기존 연구자료를 재분석한 것으로 기존 연구에서 수집된 자료에만 의존해야 했는데 분석자료의 판단대상 수가 33개로 매우 많았다는 점도 고려해 볼 수 있다. 이와 같은 추론을 바탕으로 추후에는 순위자료와 평정자료의 관련성에 대한 타당한 경험적 근거를 확보하기 위해 순위자료 생성의 방법, 판단대상과 판단자 수에 따른 비교등이 이루어질 필요가 있다.

특히, 본 연구에서 개인 수준의 상관분석을 통해 확인한바, 순위자료와 평정자료의 상관계수가 낮다는 것은 한 개인이 같은 행동을 평가더라도 평가하는 방법에 따라 비난가능성 차원에서 그 행동의 상대적 위치가 크게 달라질 수 있음을 의미한다. 즉, 두 자료 간 상관계수가 낮은 행동의 경우 그 행동에 대한 개인의 평가가 평가방식의 영향을 크게 받을 수 있다. 개인의 평가가 평가방식의 영향을 받는다는 것은 그만큼 그 평가에 관여하는 내적 속성이 불안정하다는 것이므로, 연구자들은

평가하려는 판단대상과 그 평가에 관련된 개인의 속성이 안정적인 것인지를 고려하여 측정방법을 신중히 선택해야 할 것이다.

자료수집 방법이 연구 결과에 영향을 줄 수 있음은 본 연구의 상관계수 행렬을 비교한 결과를 통해서도 합리적으로 추론할 수 있다. 순위자료와 평정자료의 상관계수 행렬을 시각화한 그림은 두 방법으로 평가된 판단대상 간 상관관계의 방향이 반대인 경향이 있음을 직관적으로 보여준다. 순위자료와 달리 평정자료는 판단자의 응답 양식(response style)이 자료에 반영된다는 단점이 있고(Phillips, 1973), 이 단점은 평정자료의 상관계수 행렬이 정적으로 편향되게 한다(Alwin & Krosnick, 1985). 본 연구의 분석자료는 다양한 행동의 비난가능성을 판단한 것이고, 이 행동들은 비난가능성이 거의 없는 행동(예를 들면, 어린이나 노인의 짐을 들어주기)부터 범죄에 이르는 행동(예를 들면, 성매매)까지 다양하였으므로 이 행위들에 대한 비난가능성 측정치들 사이에 이론적으로 기대되는 일관된 관련성의 방향은 없다. 그럼에도 불구하고 평정자료에서 일정한 방향성이 관찰되었다는 것은 그 방향성을 유발하는 속성이 측정치에 반영되어 있음을 의미한다. 즉, 상관계수 행렬의 비교에서 관찰된 순위자료와 평정자료의 상관계수 행렬의 차이는 자료수집 방법에 내재된 인지적인 속성에 의한 것일 수 있다. 반대로 순위자료의 경우 측정에서 얻을 수 있는 숫자가 총점(판단대상의 개수)에 종속되어 있다는 점에서 내준적이고, 이 속성은 반응 오차의 부적 편향을 유발한다(Chan & Bentler, 1993; Vanleeuwen & Mandabach, 2002). 이와 같은 경향은 본 연구의 사례 수가 충분하지 않았음에도 관찰되었다. 두 자료의 편향을 교정한 후 요인분석 결

과를 비교한 Hino와 Imai(2019)의 연구에서 저자들은 두 자료는 이 편향을 교정한 후에도 동일한 요인들을 추출하지 못한다고 주장하였다. 본 연구에서 보고한 상관계수는 Spearman의 순위 상관계수이고 상관계수에 대한 추론을 하기에는 사례 수가 충분하지 않지만, 선행연구와 유사한 경향을 관찰하였다. 이 결과는 연구자가 검증하려는 모형이 공변량 구조에 의존하고 있다면 이 차이는 연구자를 다른 결론으로 이끌 수 있다는 것을 의미한다(Alwin & Krosnick, 1985). 따라서 연구자들은 자료수집 방법을 결정하기 전에 자료 분석방법을 고려해야 하겠다.

퍼기분석의 도면 비교

순위자료와 평정자료를 이용하여 퍼기분석을 시도한 결과 두 자료의 모형 적합도는 모두 수용 가능한 수준이었고, 결합도면은 시각적으로는 차이가 있었다. 통계적인 분석을 시도한 결과, 결합도면의 판단대상 배치는 순위자료와 평정자료에서 의미 있는 차이가 있다는 결론을 내릴 수 있었지만 판단자 배치는 의미 있는 차이가 있다는 결론을 내릴 수 없었다.

순위자료가 평정자료보다 판단대상을 잘 구분하는 것은 평정자료에는 동점이 많고 순위자료에는 동점이 없기 때문일 수 있다. 한 사람의 순위자료에서 모든 대상은 하나의 숫자만을 부여받는다. 반면에, 평정자료에서 각 대상이 부여받는 숫자는 평정척도의 범위에 따라 동점이 존재하는 것이 가능하다. 이것은 판단대상의 수가 평정척도의 범위 수보다 작은 경우에도 마찬가지이다. 이러한 문제를 우려하여 일부 연구에서는 평정척도의 범위를 0

에서 99까지로 넓게 지정하여 자료를 수집하기도 한다(Rankin & Grube, 1980). 그러나 많은 연구자들이 실제 연구에서 사용하는 평정척도의 범위는 5점에서 9점 정도이다. 따라서 100개의 평정범주로 측정된 평정자료를 이용하여 평가된 순위자료와 평정자료의 관련성을 보여주는 선행연구의 결과는 경험적 연구의 맥락에 일반화되기 어려울 것으로 생각된다. 자료에 동점이 존재할 가능성을 낮추면서 연구자들이 쉽게 이용할 수 있는 방법은 연속적 평정 척도(continuous rating scale, 다양한 유형에 대해서는 Chyung et al., 2018 참조)를 사용하는 것이다. 이 유형의 측정방법은 선택의 양극단값만을 제시하고 판단자들에게 자신의 상태나 동의하는 지점을 자유롭게 표시하도록 한다. 영점으로부터 판단자가 표시한 지점의 거리를 길이의 단위로 측정한 값이 판단자의 평정치가 되고, 이 값은 연속적인 값으로 간주된다. 이와 같은 방법을 사용하여 얻은 평정자료는 동점의 가능성이 매우 낮아지므로 순위자료와 평정자료의 결합도면이 본 연구에서 관찰한 것보다 유사할 가능성이 있다. 비록 본 연구가 동점을 해쳐놓는 방법을 적용하여 동점의 문제를 최소화하고자 하였지만, 이 방법이 평정자료의 동점을 어느 정도나 분산시키는지는 불분명하므로 자료수집 단계에서 동점을 최소화하는 방법과 해쳐놓기 방법의 효과는 후속연구를 통해 비교해 볼 필요가 있을 것으로 생각된다.

그밖에, 순위자료와 평정자료의 차이가 판단대상보다 판단자에게서 더 작았던 것은 본 연구에서 사용한 자료의 내용적 특성 때문일 가능성도 있다. 본 연구에서 사용한 판단대상인 비난가능성이 있는 행동들이 판단자들이 공유하고 중시하는 사회적 규범에 의해 평가

되었고, 그 차원에서의 개인차가 매우 작았다면 여기에 포함된 자료의 수집방법에 따른 차이도 매우 작을 수밖에 없다. 따라서 추후에는 본 연구의 결과가 다른 유형의 판단대상에 대해서도 관찰되는지 확인될 필요가 있다.

본 연구의 결과는 연구자가 폐기분석을 사용하는 목적에 따라 자료수집 방법을 달리해야 할 필요가 있다는 것을 시사한다. 판단대상 또는 판단자를 독립적으로 시각화한 도면(행렬)만을 제공하는 다른 다차원 척도법 모형들과 달리 폐기분석은 판단대상과 판단자를 동시에 시각화할 수 있도록 판단대상과 판단자의 도면을 위한 행렬을 모두 제공한다. 이 두 행렬을 결합하여 작성한 결합도면을 이용하면 판단대상과 판단자를 동일한 차원에서 이해할 수 있고, 판단대상과 판단자의 거리를 파악할 수 있으므로 판단대상이 각 판단자에게 얼마나 중요한지, 얼마나 선호되는지 등을 추론할 수 있다. 연구자는 위의 두 가지 목적을 달성하기 위해 폐기분석을 사용할 수 있다. 본 연구의 결과에 따르면, 연구자가 판단대상을 군집화하고, 군집화된 대상들 사이에서 개인의 위치를 파악하고자 하는 경우 순위자료가 평정자료보다 유용할 수 있다. 우수한 군집화는 유사한 속성을 가진 대상들은 가깝게, 다른 속성을 가진 대상들은 멀리 배치되었을 때 가능하다. 본 연구의 그림 3에서 확인된바, 평정자료보다는 순위자료로 분석하는 경우에 판단대상들이 잘 구분될 수 있다. 평정자료를 이용하는 경우, 판단대상들이 모두가 가깝게 배치되어 하나의 군집으로 판단될 가능성이 높다. 본 연구의 판단대상의 도면을 비교한 그림을 보면 순위자료의 도면에서는 약 세 개의 행동 군집(성범죄, 범죄, 비도덕적 행동)이 확인되는 반면, 평정자료의 경우에는

세부적인 군집을 확인하기 어렵다. 폐기분석의 중요한 목적은 판단대상과 개별 판단자의 관계(거리)를 파악하는 것이다. 본 연구의 결과는 이 경우에도 순위자료를 사용하는 것이 유용할 수 있음을 시사한다. 비록 순위자료와 평정자료가 개인들을 구분하는 정도는 유사하지만, 단순히 개인을 군집화할 목적이 아니라면 판단대상이 더 잘 구분될수록 판단자와 판단대상 간 거리의 변량이 커질 것이므로 그만큼 정보의 양이 증가할 것으로 기대할 수 있다.

한계 및 의의

다차원 척도법의 모형들은 통계적인 가정으로부터 비교적 자유롭다(Jaworska & Chupetlovska-Anastasova, 2009). 또한, 자료의 시각화 방법으로 많이 사용될 뿐만 아니라 표본의 크기가 커질수록 자료와 도면의 괴리를 나타내는 STRESS 지수는 커지는 경향이 있다(Spence & Young, 1978). 최근에는 표본 크기가 큰 자료에 다차원 척도법 모형들을 적용하는 경향이 있지만, 사례 수가 커야 분석 결과를 신뢰할 수 있기 때문이라기보다는 적합도의 유의성 판단이 사례 수에 영향을 받지 않는 방법(Mair et al., 2016)이 개발되었기 때문이다. 즉, 다차원 척도법은 연구의 목적에만 부합한다면 작은 사례 수가 결과와 결과의 해석에 큰 영향을 주지 않는다. 따라서 본 연구의 폐기분석 도면의 비교는 분석자료의 표본 크기가 작더라도 그 자체로 의미가 있다고 하겠다. 그러나 본 연구에서 부수적으로 시도한 상관계수 행렬의 비교나 기술통계치의 비교는 표본 크기와 극단치의 영향을 크게 받으므로, 후속 연구를 통하여 본 연구의 결과를 반드시

재검증해 볼 필요가 있다. 또한, 충분히 큰 사례 수를 확보할 수 있다면 자극(문항)과 판단자를 동시에 시각화하는 또 다른 방법인 잠재공간 문항반응모형(latent space item-response model, Jeon et al., 2021)을 적용해 볼 수도 있을 것이다. 다만, 문항반응이론은 통계적 가정에 민감하므로 이 점을 반드시 고려해야 한다.

본 연구는 순위자료와 평정자료를 비교하면서 심리학 영역에서 개인차 연구를 위해 활용 가능성이 높은 퍼기분석을 적용하여 그 결과가 자료의 특성에 따라 어떻게 달라질 수 있는지 확인함으로써 연구자들이 자료수집 단계에서 고려해야 하는 문제들이 무엇인지에 대한 정보를 제공한 것에 의의가 있다. 그러나 본 연구는 다른 목적으로 수집된 경험자료를 사용하여 수행된 탐색적 목적의 연구인바, 이 결과가 순위자료와 평정자료의 수집과정에서 발생하는 인지과정에 의한 것인지 자료 자체의 속성에 의한 것인지에 대한 추론은 제한적으로 이루어져야 할 것이다. 이 차이의 원인은 두 자료의 수집과정에 필요한 인지적 과정과 각 평가방식에 내재된 속성(측정치의 제한 또는 동점의 문제)을 조작하는 실험연구와 체계적인 시뮬레이션 연구의 반복적 검증을 통해 확인되어야 할 것으로 판단된다. 다만, 체계적인 시뮬레이션 연구를 위해서는 자료의 특성에 영향을 줄 수 있는 절차적인 요인이나 인지적 요인의 효과가 경험적으로 먼저 확인되어 자료생성을 위한 근거가 마련될 필요가 있을 것으로 판단된다. 본 연구는 두 자료에 의한 퍼기분석 결과 차이를 탐색함으로써 관련된 후속연구를 촉진하는 좋은 시작점이 될 수 있을 것이다. 본 연구는 사람들의 일반적인 특성에 관심을 가지는 심리학의 주류 연구

자들에게는 순위자료의 활용법을 소개하고, 개인차를 유발하는 요인과 소집단 연구에 관심을 가지는 연구자들에게는 개인과 판단의 대상의 관계를 동일한 차원에서 이해할 수 있는 방법을 소개했다는 점에서 심리학 연구자들에게 유용한 참고자료가 될 것으로 기대된다.

참고문헌

- Alwin, D. F. & Krosnick, J. A. (1985). The measurement of values in surveys_A comparison of ratings and rankings. *Public Opinion Quarterly*, 49(4), 535-552.
<https://doi.org/10.1086/268949>
- Borg, I., Bardi, A., & Schwartz, S. H. (2017). Does the value circle exist within persons or only across persons?. *Journal of Personality*, 85(2), 151-162.
<https://doi.org/10.1111/jopy.12228>
- Borg, I., Groenen, P. J. F., & Mair, P. (2018). *Applied Multidimensional Scaling and Unfolding*. Springer.
- Böckenholt, U. (2004). Comparative judgments as an alternative to ratings_identifying the scale origin. *Psychological methods*, 9(4), 453-465.
<https://doi.org/10.1037/1082-989X.9.4.453>
- Busing, F. M. (2006). Avoiding degeneracy in metric unfolding by penalizing the intercept. *British Journal of Mathematical and Statistical Psychology*, 59(2), 419-427.
<https://doi.org/10.1348/000711005X68507>
- Cattell, R. B. (1944). Psychological measurement: Normative, ipsative, interactive. *Psychological*

- Review*, 51(5) 292-303.
<https://doi.org/10.1037/h0057299>
- Chan, W. & Bentler, P. M. (1993). The covariance structure analysis of ipsative data. *Sociological Methods Research*, 22(2), 214-247.
<https://doi.org/10.1177/0049124193022002003>
- Chyung, S. Y., Swanson, I., Roberts, K., & Hankinson, A. (2018). Evidence based survey design: The use of continuous rating scales in surveys. *Performance Improvement*, 57(5), 38-48.
<https://doi.org/10.1002/pfi.21763>
- Coombs, C. H. (1950). Psychological scaling without a unit of measurement. *Psychological Review*, 57(3), 145-158.
<https://doi.org/10.1037/h0060984>
- D'Ambrosio, A., Vera, J. F., & Heiser, W. J. (2022). Avoiding degeneracies in ordinal unfolding using Kemeny-equivalent dissimilarities for two-way two-mode preference rank data. *Multivariate Behavioral Research*, 57(4), 679-699.
<https://doi.org/10.1080/00273171.2021.1899892>
- De Leeuw, J. & Mair, P. (2009). Multidimensional scaling using majorization: SMACOF in R. *Journal of statistical software*, 31(3), 1-30.
<https://doi.org/10.18637/jss.v031.i03>
- Elig, T. W. & Frieze, I. H. (1979). Measuring causal attributions for success and failure. *Journal of Personality and Social Psychology*, 37(4), 621-634.
<https://doi.org/10.1037/0022-3514.37.4.621>
- Fabbris, L. (2012). Measurement scales for scoring or ranking sets of interrelated items. In *Survey data collection and integration* (pp. 21-43).
https://doi.org/10.1007/978-3-642-21308-3_2
- Feather, N. T. (1973). The measurement of values: effects of different assessment procedures. *Australian Journal of Psychology*, 25(3) 221-31.
<https://doi.org/10.1080/00049537308255849>
- Green, P. E. & Rao, V. R. (1972). *Applied multidimensional scaling*. Holt, Rinehart and Winston.
- Groves, R. M. & Kahn, R. L. (1979). *Surveys by telephone: A national comparison with personal interviews*. Academic Press.
- Hino, A. & Imai, R. (2019). Ranking and rating_Neglected biases in factor analysis of postmaterialist values. *International Journal of Public Opinion Research*, 31(2), 368-381.
<https://doi.org/10.1093/ijpor/edy007>
- Jaworska, N. & Chupetlovska-Anastasova, A. (2009). A review of multidimensional scaling (MDS) and its utility in various psychological domains. *Tutorials in quantitative methods for psychology*, 5(1), 1-10.
<https://doi.org/10.20982/tqmp.05.1.p001>
- Jeon, M., Jin, I. H., Schweinberger, M., & Baugh, S. (2021). Mapping unobserved item - respondent interactions: a latent space item response model with interaction map. *Psychometrika*, 86(2), 378-403.
<https://doi.org/10.1007/s11336-021-09762-5>
- Jin, W. J., Jung, J., & Han, Y. (2023). An exploration of the psychological dimension for the evaluation of victim-likeness: Application of INDSCAL and multi-dimensional unfolding. *Korean Journal of Social and Personality Psychology*, 37(2), 159-187.
<https://doi.org/10.21193/kjspp.2023.37.2.003>

- Krosnick, J. A. (1999). Maximizing questionnaire quality. In Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (Eds.), *Measures of political attitudes: Volume 2 of measures of social psychological attitudes* (pp. 37-57). Academic Press.
- Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1), 1-27.
https://doi.org/10.1007/BF02289565
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22(140), 1-55.
- Mair, P., Borg, I., & Rusch, T. (2016). Goodness-of-fit assessment in multidimensional scaling and unfolding. *Multivariate behavioral research*, 51(6), 772-789.
- Mair, P., Groenen, P. J., & de Leeuw, J. (2022). More on multidimensional scaling and unfolding in R: smacof version 2. *Journal of Statistical Software*, 102, 1-47.
https://doi.org/10.18637/jss.v102.i10
- Munson, J. M. & McIntyre, S. H. (1979). Developing practical procedures for the measurement of personal values in cross-cultural marketing. *Journal of Marketing Research*, 16(1), 48-52.
https://doi.org/10.2307/3150873
- Phillips, D. L. (1973). *Abandoning Method*. Jossey-Bass Publishers.
- Rankin, W. L. & Grube, J. W. (1980). A comparison of ranking and rating procedures for value system measurement. *European Journal of Social Psychology*, 10(3), 233-246.
https://doi.org/10.1002/ejsp.2420100303
- Reynolds, T. J. & Jolly, J. P. (1980). Measuring personal values: An evaluation of alternative methods. *Journal of Marketing Research*, 17(4), 531-36. https://doi.org/10.2307/3150506
- Rokeach, M. (1973). *Rokeach Value Survey (RVS)* [Database record]. APA PsycTests.
https://doi.org/10.1037/t01381-000
- Russell, P. A. & Gray, C. D. (1994). Ranking or rating? Some data and their implications for the measurement of evaluative response. *British Journal of Psychology*, 85(1), 79-92.
https://doi.org/10.1111/j.2044-8295.1994.tb02509.x
- Spence, I. & Young, F. W. (1978). Monte Carlo studies in nonmetric scaling. *Psychometrika*, 43(1), 115-117.
https://doi.org/10.1007/BF02294095
- Torgerson, W. S. (1952). Multidimensional scaling: I. Theory and method. *Psychometrika*, 17(4), 401-419. https://doi.org/10.1007/BF02288916
- Torgerson, W. S. (1958). *Theory and methods of scaling*. Wiley.
- van Herk, H. & van de Velden, M. (2007). Insight into the relative merits of rating and ranking in a cross-national context using three-way correspondence analysis. *Food Quality and Preference*, 18(8), 1096-1105.
https://doi.org/10.1016/j.foodqual.2007.05.006
- Vanleeuwen, D. M. & Mandabach, K. H. (2002). A note on the reliability of ranked items. *Sociological Methods Research*, 31(1), 87-105.
https://doi.org/10.1177/0049124102031001004

1차원고접수 : 2023. 07. 13.

2차원고접수 : 2023. 08. 24.

최종게재결정 : 2023. 09. 25.

Multidimensional unfolding and characteristics of data: Comparison of ranking and rating data

Yuhwa Han

Department of Psychology, Chungbuk National University /
Institute for Legal Studies, Yonsei University

This study compared the results of unfolding analyses, a technique to visualize judges and judged objects in a single figure, using ranking and rating data. Multidimensional scaling models, such as the unfolding model, are relatively free from statistical assumptions and can be advantageous tools for researchers interested in small groups or individual differences. The data used in the analysis were from 37 participants who were asked to measure 33 behaviors on the blameworthiness dimension, and the data were collected in two ways for each of the 33 behaviors. The data from paired comparison method was converted into ranked data, and the rated data was used as raw data. To understand the relationship between the ranking and rating data, we calculated the rank correlation coefficient (ρ) between the aggregated statistics of the two data, and compared the correlation matrices of the 33 behaviors from the two data. The results showed that the two data were not related to each other, and the correlation matrices were not similar. As a result of the unfolding analysis, the placement of the behaviors was more distributed when using the ranking data than the rating data. Therefore, the author concluded that ranking data would be more useful to visually understand the relationship between the judgment objects and the judges. This study showed that the results of the unfolding analysis of ranking and rating data from the same judges are different. The current study suggested that, when applying the unfolding model, researchers should consider which type of data can be useful for the individual difference study.

Key words : multidimensional scaling, unfolding model, ranking data, rating data, individual difference study